ADVANCING LUNG CANCER PREDICTION: INNOVATIVE AIMODULES

Keerthana B A[1] , Dharaneesh J[2] , Suseela D[3]

[1]*Student, Computer Science and Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*

[2]*Student, Computer Science and Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*

[3]*Faculty, Artificial Intelligence and Machine Learning, Bannari Amman Institute of Technology, Tamil Nadu, India*

ABSTRACT

Lung cancer is the leading cause of cancer deaths worldwide. Directly diagnosing the malice of suspected lung nodes is of consummate clinical significance. Accurate discovery of lung nodes on casket reckoned tomography reviews is pivotal to early opinion of lung cancer. To address the nuisance problems on low discovery perceptivity and high false-positive rate caused by diversity and morphological complexity of 2- D bump features, a computer- backed discovery system is developed to increase the discovery perceptivity and bracket delicacy of lung nodes. Aiming at the disadvantages of low delicacy and slow speed in current lung bump discovery styles, this design proposes a featherlight lung bump vaticination algorithm grounded on bettered Region Convolutional Neural Network. This system first preprocesses the original CT image and sends it to the Region Offer Network for pulmonary nodes discovery. also, the discovery results are used to complete the bracket of benign and nasty lung nodes through the Region grounded Convolutional Neural Network (R- CNN). The experimental results on the LUNA16 data set prove that the bettered network model chart value can reach96.71, and the discovery speed can reach41.99 FPS. Results interlace that our proposed bracket network could give an effective individual tool for suspected lung nodes and might have a promising operation in clinical practice.

Keyword: *Region Convolutional Neural Network, CT image, lung nodes*

1 . INTRODUCTION

1.1. OVERVIEW

The lungs, which is the organ for respiration is a paired cone shaped organs lying in the thoracic cavity separated from each other by the heart and other structures in the mediastinum.

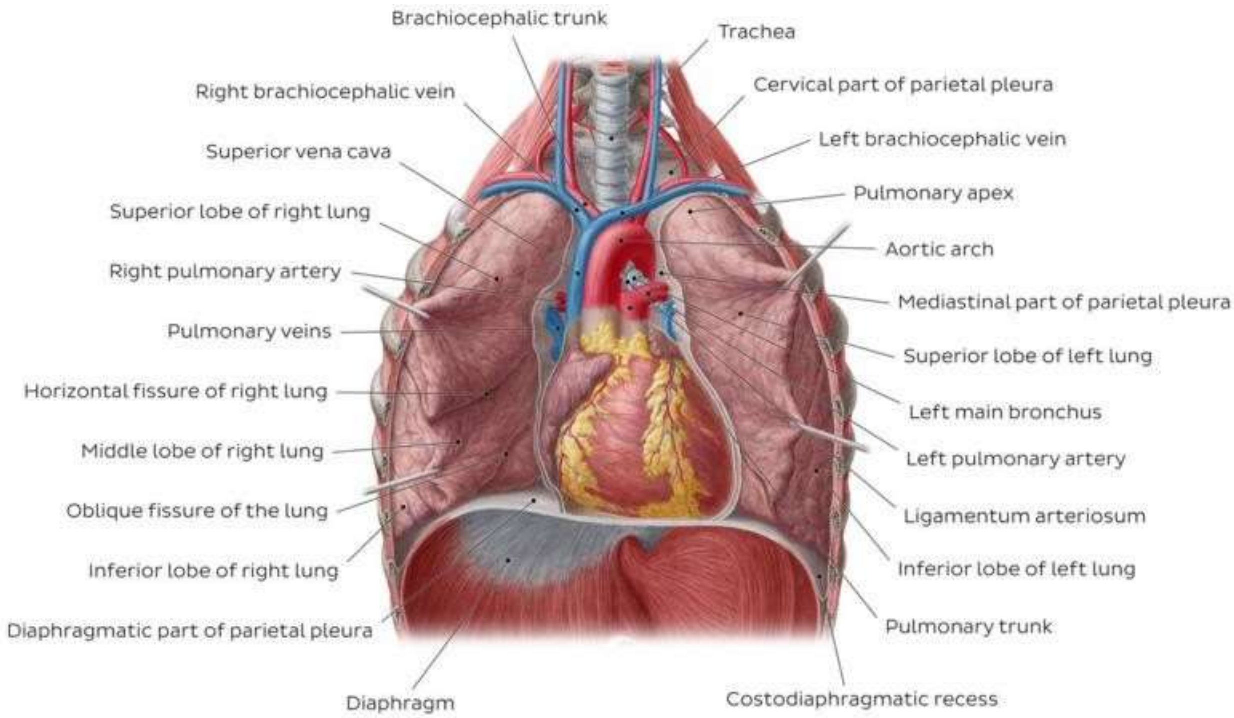


Figure 1.1. Lung

Each lung has a base resting on the diaphragm and an apex extending superiorly to a point approximately 2.5 cm superior to the clavicle. It also has a medial surface and with three borders- anterior, posterior and inferior. The broad coastal surface of the lungs is pressed against the rib cage, while the smaller mediastinal surface faces medially. The lungs receive the bronchus, blood vessels, lymphatic vessels and nerves through a slit in the mediastinal surface called the helium, and the structures entering the helium constitutes the lungs root. The right lung is larger and weighs more than the left lung. Since the heart tilts to the left, the left lung is smaller than the right and has an indentation called the cardiac impression to accommodate the heart. This indentation shapes the inferior and anterior parts of the superior lobe into a thin tongue-like process called the lingual.

1.1.1. Lung Nodule

Also called a Solitary pulmonary nodule, a lung nodule is a round spot in the lung that is more solid than normal lung tissue. It’s usually detected on an X-ray or computed tomography (CT or CAT) scan. Solitary pulmonary nodules in asymptomatic individuals may be benign or malignant, and the list of possible causes includes cancer, infection, and granulomas; Lung nodules are quite common. In fact, lung nodules are found in up to half of the people who have chest X-rays. Lung nodules — small masses of tissue in the lung — are quite common. They appear as round, white shadows on a chest X-ray or computerized tomography (CT) scan. Lung nodules are usually about 0.2 inch (5 millimeters) to 1.2 inches (30 millimeters) in size. A larger lung nodule, such as one that's 30 millimeters or larger, is more likely to be cancerous than is a smaller lung nodule.

The Solitary Pulmonary Nodule. “Coin lesion” Defined as < 3 cm Completely surrounded by lung parenchyma Lesions > 3 cm called “masses” and often malignant. The Solitary Pulmonary Nodule. Incidence of cancer from 10 – 70%. An SPN is surrounded by normal lung tissue and is not associated with any other abnormality in the lung or nearby lymph nodes (small, bean-shaped structures found throughout the body). Most SPNs are benign (noncancerous); however, they may represent an early stage of primary lung cancer or may indicate that cancer is metastasizing (spreading) from another part of the body to the affected lung.Determining whether the SPN seen on the chest X-ray or chest CT scan is benign or malignant (cancerous) is important. Prompt diagnosis and treatment of early lung cancer that looks like an SPN may be the only chance to cure the cancer

Figure 1.2. Nodule

1.1.2. Causes of Lung Nodule

There is a plethora of causes for nodules. They are often the result of previous or current infections. These nodules may well have been present for many years and are considered benign. Sometimes nodules can be cancerous, and this of course needs further investigation. All nodules therefore do need careful review with a specialist. Lung nodules can be benign (noncancerous) or malignant (cancerous). Most benign lung nodules result from inflammation due to infections or illnesses. When your lungs become inflamed, tiny masses of tissue can develop. Over time, the clumps can harden into a nodule on your lung. The causes of most benign lung nodules include;

- **Infections**, including bacterial infections such as tuberculosis and pneumonia; fungal infections from inhaling spores found in bird or bat droppings, or moist soil, wood, and leaves; and parasitic infections such as roundworms and tapeworms.
- **Inflammation** from irritants in the air, air pollution, or autoimmune conditions such as rheumatoid arthritis or sarcoidosis.
- **Scarring** from surgery or chest radiation.

1.1.3. Diagnosis of Lung Nodule

Imaging Test- During this process, an abnormal mass can be revealed through an X-ray image. A CT/PET-CT scan is recommended if the X-ray scan comes out clean. A clear picture of small lesions can be seen in a CT scan. The CT scan is an invaluable aid in identifying features of the nodule and determining the likelihood of cancer. In addition to the features seen on a chest X-ray, a CT scan of the chest allows better assessment of the nodule.

1.2. PROBLEM IDENTIFIED

According to the statistics, among the mortality rate of all cancers is 19.5%, the incidence of lung cancer accounts for 66.67% with the 18% of five-year survival rate. Lung Cancer is one of the most common type of cancers according to the latest global cancer statistics, leading to the highest number of cancer-related deaths worldwide. The precise segmentation and classification of lung nodules on computed tomography (CT) images and assessment of intra-nodular heterogeneity are prerequisites for these techniques, which not only reduce the bias of subsequent radiological analysis but also improve the work efficiency of radiologists. Currently, manual delineation of lung nodules in CT scan is most commonly performed slice by slice by a radiologist. Although this manual segmentation and classification method may guarantee accuracy to a certain extent, the procedure involves several shortcomings. First, the radiologist must identify and delineate the boundary of the lung nodule on each CT slice, which is a tedious procedure that relies on subjective experience. Second, the number of patients with lung nodules enrolled in such studies has grown from dozens to hundreds or even thousands in recent years, making manual segmentation and classification less feasible. Finally, for the current big data-driven medical imaging workflow, manual delineation of the region of interest (ROI) contradicts the objective of using artificial intelligence. e. In summary, end-to-end algorithms for automatic segmentation and classification of lung nodules on CT images are increasingly necessary to improve medical imaging workflows. Therefore, automatic methods for lung nodule segmentation and classification and intra-nodular heterogeneity image generation are needed in current radiological practice. In summary, traditional lung nodule segmentation methods are limited by the requirement for additional human processing and the challenge of accurately segmenting adhesive nodules, whereas the current deep learning models may be more rigorous for image pre-processing and pre-selected nodule patches.

1.3. DEEP LEARNING

Deep Learning is a part of machine learning, which is a subset of Artificial Intelligence. It enables us to extract the information from the layers present in its architecture. It is used in Image Recognition, Fraud Detection, News Analysis, Stock Analysis, Self-driving cars, Healthcare like cancer image analysis, etc. By inputting more data in the network the layers get trained very well. They can be classified into Supervised, Semi-Supervised and Unsupervised categories. Each layer is known for extracting information specifically. For example, in Image recognition, the first layer will find the edge, lines, etc, second layer like eye, ear, nose, etc.

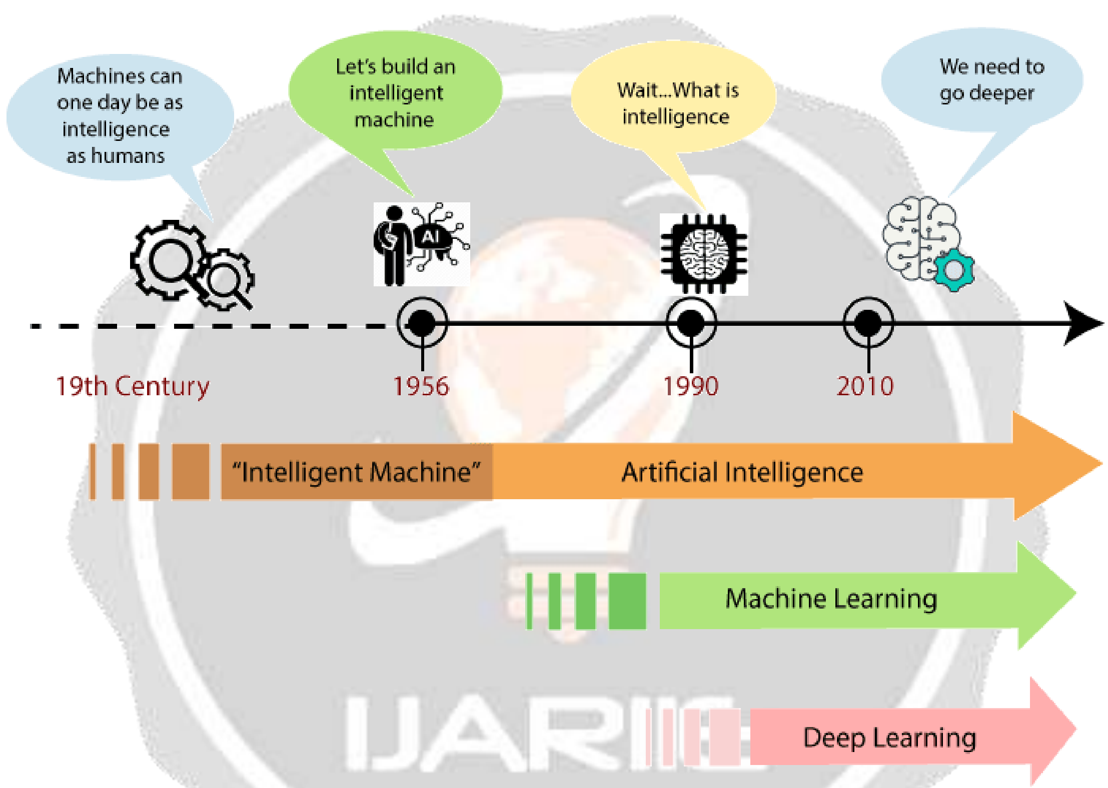


Figure 1.3. Evolution of AI

To help improve the efficiency of predictions, to find the best possible outcomes and for model optimization. When the data is huge, to reduce the cost in the company in terms of insurance, sales, profit, etc. Deep learning can be very useful when there is no particular structure to data means to analyze data from audio, video, image, numbers, document processing, etc.

Advantages of Deep Learning

- Solve Complex problems like Audio processing in Amazon echo, Image recognition, etc, reduce the need for feature extraction, automated tasks wherein predictions can be done in less time using Keras and Tensorflow.
- Parallel computing can be done thus reducing overheads.
- Models can be trained on a huge amount of data and the model gets better with more data.
- High-Quality Predictions when compared with humans by training tirelessly.
- Works well-unstructured data like video clips, documents, sensor data, webcam data, etc.

1.3.1. Deep Learning Algorithms

Deep learning algorithms work with almost any kind of data and require large amounts of computing power and information to solve complicated issues. Now, let us, deep-dive, into the top 10 deep learning algorithms.
Here is the list of top 10 most popular deep learning algorithms:

- Convolutional Neural Networks (CNNs)
- Long Short Term Memory Networks (LSTMs)
- Recurrent Neural Networks (RNNs)
- Generative Adversarial Networks (GANs)
- Radial Basis Function Networks (RBFNs)
- Multilayer Perceptrons (MLPs)
- Self-Organizing Maps (SOMs)
- Deep Belief Networks (DBNs)
- Restricted Boltzmann Machines(RBMs)
- Autoencoders

1.3.2. Applications of Deep Learning

Given below are the applications of Deep Learning:

- **Healthcare**

From Medical image analysis to curing diseases, Deep Learning played a huge role especially when GPU-processors are present. It also helps Physicians, Clinicians, and doctors to help the patients out of danger, and also they can diagnose and treat the patients with apt medicines.

- **Stock Analysis**

Quantitative Equity Analysts are getting more benefits specially to find the trends for a particular stock whether it will be bullish or bearish and they can use many more factors like no of transactions made, no of buyers, no of sellers, previous day closing balance, etc when training the deep learning layers. Qualitative Equity Analysts use factors like return on equity, P/E ratio, return on Asset, Dividend, return on Capital Employed, Profit per Employee, Total Cash, etc when training the deep learning layers.

- **Fraud Detection**

These days, hackers especially those based out of the dark- web have found ways to steal money digitally across the globe using different software. Deep learning will learn to find these types of fraudulent transactions in the web using a lot of factors like Router information, IP addresses, etc. Auto encoders also help financial institutions saving billions of dollars in terms of cost. These types of fraudulent transactions can also be detected by finding the outliers and investigating the same.

- **Image Recognition**

Suppose say the city police department has a people database of the city and they wanted to know in public gatherings like who is involved in the crimes, violence using public webcam available in streets this deep learning using CNN (Convolution Neural networks) helps a lot in finding the person who was involved in the act.

- **News Analysis**

These days the government takes a lot of effort especially in controlling the spread of fake news and origin of it. Also during poll surveys like who would win elections in terms of popularity, which candidate been shared by most people in social media etc and analysis of tweets made by country people using all these variables we can predict the outcomes in deep learning, but also there are some limitations to it, we don't know the data authenticity whether its genuine or fake. etc or whether the necessary information been spread by bots.

- **Self-Driving Cars**

Self-driving cars use Deep Learning by analyzing the data captured in the cars made in different terrains like mountains, deserts, Land, etc. Data can be captured from sensors, public cams, etc which will be helpful in testing and implementation of self-driving cars. The system must able to ensure all the scenarios been handled well in training.

1.4. OBJECTIVE OF THE PROJECT

In order to realize the effective and accurate identification of pulmonary nodules in CT images and reduce the phenomenon of missed detection and false detection caused by the difference of doctors' level, this project proposes a lung nodule detection algorithm based on the improved R-CNN. The objective of the project is to demonstrates the use of a deep learning model for locating, segmenting, measuring, and characterizing lung nodules in thoracic CT images.

2. SYSTEM IMPLEMENTATION

2.1. PROJECT DESCRIPTION

Accurate assessment of Lung nodules is a time consuming and error prone ingredient of the radiologist interpretation work. Early detection of lung cancer is an effective way to improve the survival rate of patients. It is a critical step to have accurate detection of lung nodules in computed tomography (CT) images for the diagnosis of lung cancer. Automating Lung Nodule detection and segmentation can improve workflow as well as patient care. To address these challenges, we proposed a fully automated end-to-end lung nodule detection and segmentation system based on a deep learning approach. In this paper, we used Improved R-CNN; a state-of-the-art detection model to detect the lung nodule regions in the CT scans. The main workflow of the proposed method is illustrated in Figure 5.1. As a pre-processing stage, we extracted the CT volumes slice by slice. The extraction of the CT slices are converted into images from the original CT scans (".dcm" file). To detect the nodule ROI, the Improved R-CNN method is then used. The detected nodule ROI is fed as an input to the segmentation task.

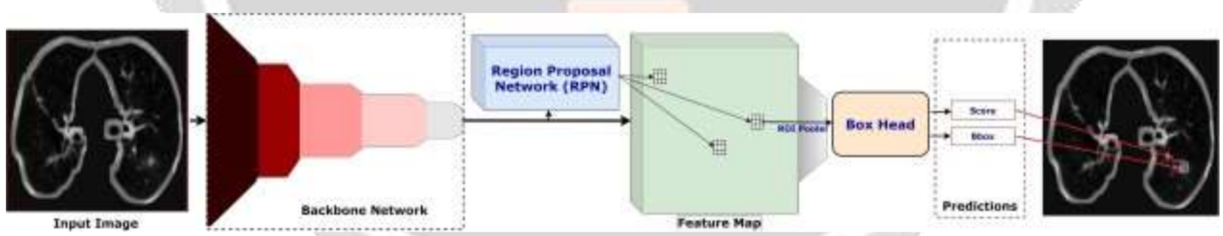


Figure 2.1. Work flow of improved R-CNN

The raw CT scans data are available in the ".dcm" format which is processed with the pylidc library to convert them to the ".png" format for the purpose of rendering it more meaningful and useful information. Afterwards, a set of image processing techniques (Otsu with a binary threshold and morphological dilation) are applied to obtain the lung region from the converted image. In the nodule detection model, we used Improved R-CNN based on the original Faster R-CNN to automatically detect lung nodules based on the lung images. The model is a two-stage network with three main blocks, Backbone Network, Region Proposal Network (RPN), and Box head. The feature map is then fed into the RPN to perform boundary regression and classification analysis. The classification principle is based on which a candidate frame is either related to background or to the object. The position and score of the RPN outputs on the candidate frame are sent to the Box head, where the final regression and classification of the object is performed. Finally, the prediction will show the bounding box of the target (nodule) with the classification score.

2.2. DATASET DESCRIPTION

Publically available datasets LUNA16 and LIDC-IDRI were used for the training of nodule detection and classification, respectively, in the proposed lung cancer detection system. LUNA16 is a subset of LIDC-IDRI dataset which only contains detection annotations confirmed by three or four radiologists. The LIDC-IDRI dataset contains all of the related information on nodules like size, location, results of diagnosis, and other related data to diagnosis from experienced doctors on low-dose CT images in XML files. 3D CT scans are collections of 2D greyscale regular slices in the right order, specified slope, and intercept. The LUNA16 dataset contains only CT scans with uniform slices and well-organized images, while the LIDC-IDRI dataset contains all types of images. The LUNA16 dataset consists of 888 low-dose lung CTs, which contains a total of 754,976 candidates that are labelled ‘0’ as non-nodule and ‘1’ as nodule. Among these, 36,378 are annotated as nodules by the radiologists, and LIDC-IDRI comprises 1,018 low-dose thoracic CT scans. The ground truth nodules for classification were obtained from the annotated LIDC-IDRI dataset, doctors’ mappings, and LUNA16’s annotated nodules. Only those ground truth nodules were considered which have a sufficient score after multiple doctors’ annotations.

2.3. MODULE DESCRIPTION

2.3.1. TRAINING PHASE: DATASET ANNOTATION

We used the data set containing 1186 lung nodules provided by the first phase of LUNA16 (the annotation information includes the diameter and location information of the nodules). Furthermore, it should be noted that this dataset is a subset of the public LIDC-IDRI data set containing 2610 lung nodules. They screened 888 CT data from the LIDC-IDRI dataset, each of which was labeled by up to four experienced radiologists. Moreover, each radiologist classifies the identified lesions into three categories, non-nodular (other tissues or background), nodules larger than 3 mm in diameter, and nodules less than 3 mm in diameter. Finally, nodules larger than 3 mm in diameter, marked by three or four radiologists, are used as the gold standard, and nodules that are less than 3 mm in diameter and marked by only one or two radiologists will be ignored.

2.3.2. PREPROCESSING

CT image is the intensity distribution of rays received after the external X-ray penetrates human body. During the ray transmission process, it passes through many unrelated tissues, such as bed frame, clothing, muscle and bones. For the detection of lung nodules, since lung nodules locate in the lung parenchyma, it is necessary to segment the lung parenchyma from CT images to avoid the interference of other tissues, thereby reducing false positives and improving the segmentation performance.

The lung parenchyma appears in the CT image as a connected domain with low gray scale that is surrounded by high gray scale chest muscles. Based on this feature, we first binarized the CT images, then deleted the regions such as air and bed frame, then filled the holes formed by the high-density tissues in the lung parenchyma, and finally repaired the lung parenchyma mask using morphological algorithms.

2.3.3. LUNG SEGMENTATION

The medical images are not in the conventional image formats such as PNG or JPEG, as they are conducted under a specified constrained environment, which has a direct impact on the image attained. Digital Imaging and Communications in Medicine (DICOM) is a standard for storing and transmitting digital images, enabling the integration of different medical imaging devices, e.g., scanners, servers, workstations, and printers. A DICOM image may also contain the information of the patient, date, and many other data that are not required in lung segmentation. To conveniently perform the image processing tasks, we convert the DICOM images to the loss-less Portable Network Graphics (PNG) format. When a DICOM image is converted to PNG format, the personal information of the patient and all tags which come with a DICOM format get removed so as to preserve the individual’s privacy.

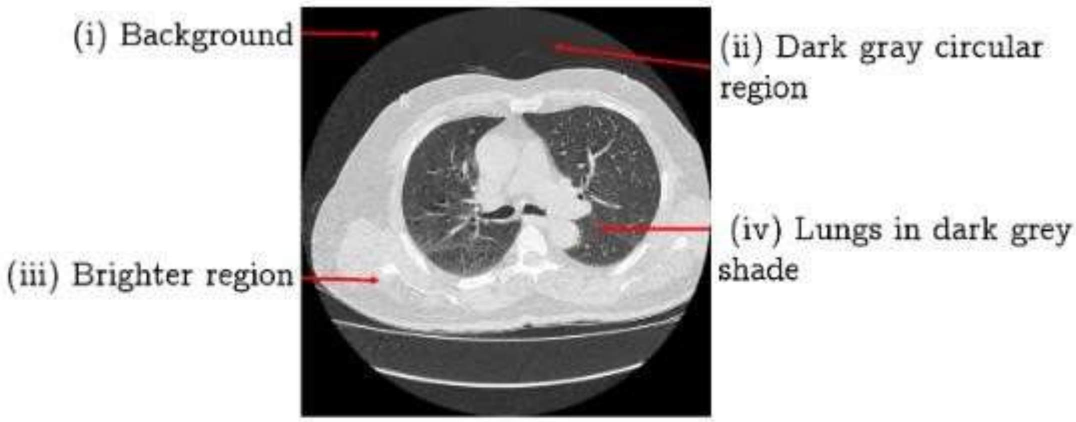


Figure 2.3.3.1. A CT image of lung and proposed division into four regions.

The lung segmentation is considered a fundamental activity in nodule CAD systems, as the performance of the later stages in such an analysis largely depends on the segmentation accuracy. In this section, we propose a lung segmentation algorithm that utilizes the histogram and morphological image processing techniques. The converted PNG image has four components: (i) a black background; (ii) a dark gray circular region; (iii) a brighter region; and (iv) the lungs in a dark gray shade, as shown in Figure 2.3.3.1. Our region of interest is the lungs, thus here we remove the first two components.

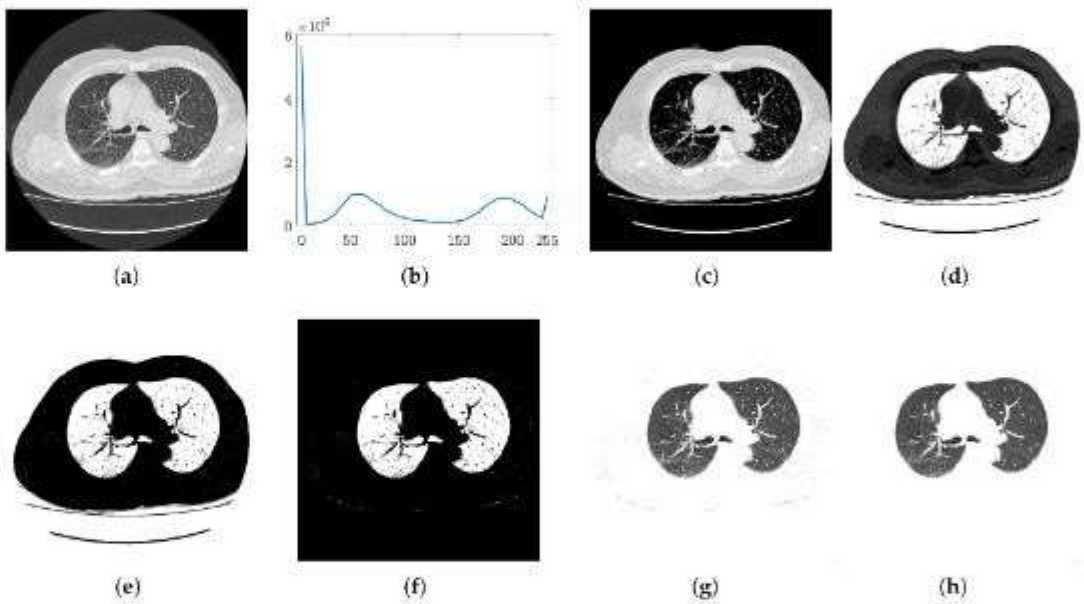


Figure 2.3.3.2. Automatic histogram based on initial segmentation: (a) Input CT slice; (b) histogram of the image in (a); (c) the results after thresholding the image with the threshold estimated from the histogram in (b); (d) the complemented image of (c); (e) binary segmentation map; (f) map after connected components-based refinement; (g) detected lungs; (h) lungs after noise removal.

To reduce the image to our ROI, we perform thresholding. Thresholding is mainly dependent on the value of the threshold, and this value is usually user specified. In our case, we compute the threshold for each slice from the histogram of the input image. Figure 3b shows the histogram of a sample lung image shown in Figure 2.3.3.2. It can be noted that the histogram has four prominent peaks. One very high peak is at 0, which corresponds to the black background in the image. The second peak around the gray level 60 is formed from the dark gray circular region covering the bright region. These two peaks correspond to the regions (i) and (ii) as discussed above. A high peak at 255 corresponds to the white region mainly around the lungs, and the third peak around gray level of 210 in this example is formed by the intensities of the lungs and patches inside the bright region around the lungs. Thus, by dropping the pixels that fall in the first two peaks, we can remove the background regions (i) and (ii). The valley where the second peak ends serves as a separator between regions i–ii and regions iii–iv. This valley can be estimated by computing the second minima of the histogram. The value of the second minima is used as a threshold to remove the regions (i) and (ii) from the CT scan. Figure c shows the result achieved after thresholding the image (Figure 2.3.3.2.) with the estimated threshold.

2.3.4. LUNG NODULE FEATURE EXTRACTION

In this phase, first the inner structures of the lungs, i.e., nodules, bronchi, and blood vessels, are separated from the parenchyma region. The inner structures in the lungs appear as bright spots (Figure 2.3.3.2.(h)), which can be easily separated through thresholding as the intensity levels of the parenchyma region and inner structures are distinguishably different. The RPN method is used on the segmented lungs to separate the inner structure vessels, bronchi, and nodules (if there are any) from the rest of the region.

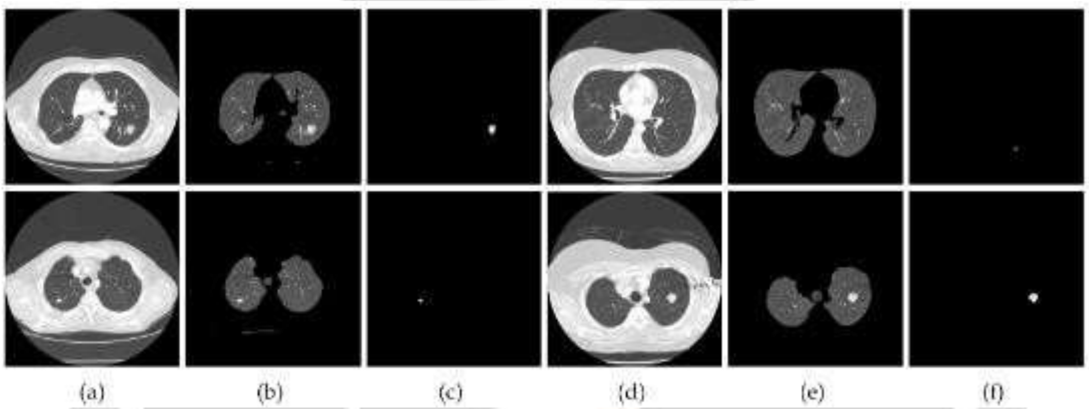


Figure 2.3.4. Nodule detection results: (**a**,**d**) input CT slice, (**b**,**e**) segmented lungs, (**c**,**f**) detected nodule.

The nodules differ from other structures present in lungs in many aspects. One key difference is their shape, and we exploit this property to isolate the nodules from the non-nodules structures, i.e., vessels and bronchi. The nodules are spherical in shape, whereas the vessels and the bronchioles are cylindrical, as shown in Figure 2.3.4. We use the size invariant round/near-round shape detection algorithm proposed in to identify the circular shapes in the detected set of structures. It returns the centers of the potential nodule locations, which are used as seed points in a region-growing algorithm to extract the nodule candidates. In contrast to existing nodule template-based techniques, the proposed strategy enables us to extract nodules of any shape, making our methodology independent of any nodule template.
Let {c1,c2,c3,⋯,cn} be the centers of the shapes extracted using algorithm RCNN. The centers are passed to the region-growing algorithm as seeds, which returns us the corresponding n nodule candidate regions {A1,A2,A3,⋯,An}. We compute different features of nodules and construct a feature vector to discriminate nodules from the other inner structures. We exploit different statistical properties, shape-based features, and across-slice characteristics of the candidate regions to design a discriminative feature vector.

2.3.5. Lung Nodule Classification

Based on feature F, the selected regions are classified as nodules and non-nodules using a Convolutional Neural Network (CNN). Training is done by using features of a training dataset, and then these same features are calculated for the testing dataset and passed to the model for classification. NoduleX Classifcation. Nodule classifcation was performed using two diferent CNN models (the CNN21 and CNN47 models) and both with and without Nodule features. For classifcation without Nodule features, the CNN model's own output sofmax classifer was used for class prediction on all nodules in the validation set. When Nodule features are used, a vector representing the 50 Nodule features is concatenated with the feature vector produced by the CNN as described above, producing a feature vector with 250 features. Tis combined feature vector is passed as input to a Random Forest classifer model, which itself must be trained on the training set (as described above). Te trained Random Forest is then evaluated on the same validation set as the CNN-only model for comparison. Te randomForest package37 in R was used for this purpose, with the ntrees parameter set to 1000 and defaults for other parameters.

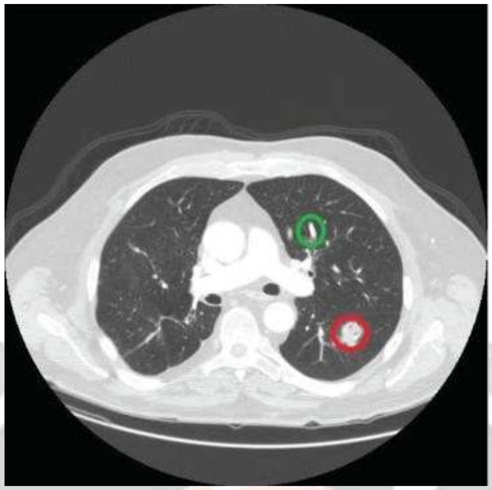


Figure 2.3.5. The region circled in red is a nodule and the region circled in green is a non-nodule.

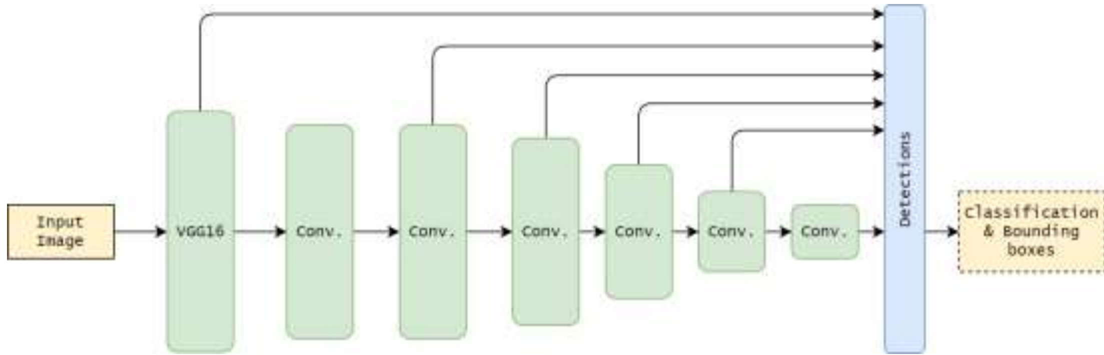
The layers of the CNN architecture which consists of multiple convolutional layers with ReLU activation, maxpooling, flatten, dense and a final fully connected softmax layer to carry out the classification between tomograms with nodules and tomograms without nodules. The convolution method in channel shuffle convolution is not the same as the convolution of the Densenet network. In the convolution of the Densenet network, a set of convolution kernels is responsible for a set of feature maps, while in the channel shuffle convolution, a convolution kernel is responsible for a feature map, which can greatly reduce the amount of parameters, but this will cause the loss of information between the same group of data. The shuffle operation can solve the problem of no communication of information in the group, and the shuffle operation can solve the problem of group and group convolution., the defect of not communicating information between groups.

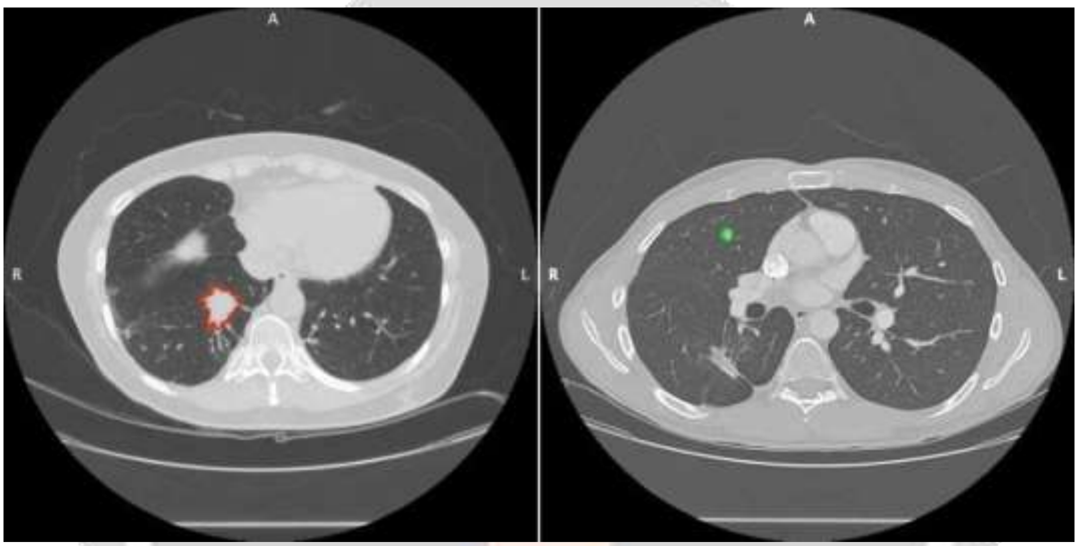
2.3.6. TESTING PHASE: LUNG NODULE DISEASE PREDICTION

After completing the training process, the algorithms have been evaluated on 10 frames including the images from the test dataset containing obstacle in various adverse environmental conditions and also including several other road entities.

2.3.6.1. PREDICTION NETWORK

The goal of this detection network is to generate final bounding box by considering inputs from Feature Network and Regional Proposal Network. It consists of 4 fully connected layers which are in turn interconnected to bounding box regression layer and classification layer that help to generate final detections.





Maligant **Benign**

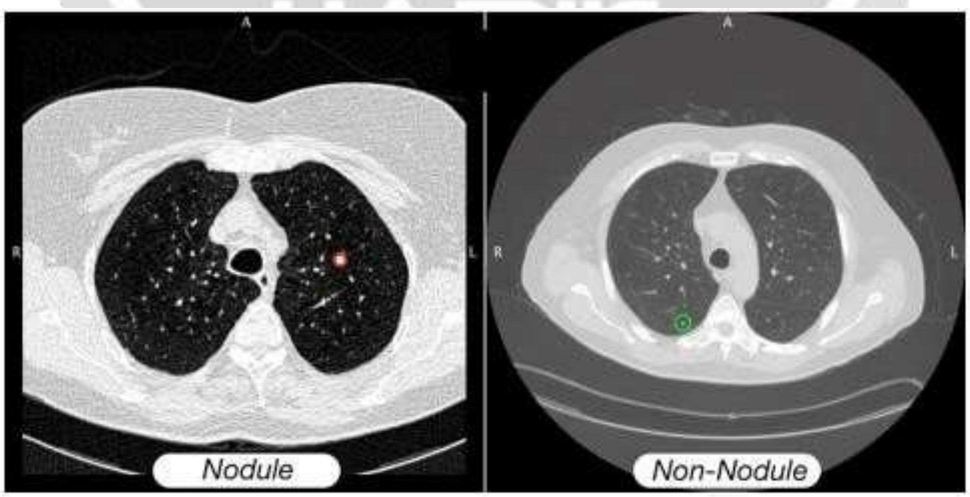


Figure 2.3.6.1. The region circled in red is a nodule and the region circled in green is a non-nodule.

2.4. ALGORITHM: IMPROVED R-CNN

1. Input CT Lung LUNA Dataset PNG format described with a RGB format that can be modeled as a 3D Matrix.
2. FRConv(x,y,z,u,v,h,i,j) x = number of expected input channels in the given image (3 for RGB)
3. y = number of the output channels after the convolution (FRCONV) phase
4. z = The kernel width of the convolution
5. u = The kernel height of the convolution
6. v = The step (stride) of the convolution in the width dimension
7. h = The step (stride) of the convolution in the height dimension.
8. i = Additional zeros added to the input plane data on both sides of width axis
9. j = Additional zeros added to the input plane data on both sides of height axis
10. ReLU(x) A rectified linear unit (activation function) has output 0 if the input is less than 0, and raw output otherwise. x = True or False.
11. MaxPooling(x,y,z,u) A max-pooling operation that looks at XxY windows and finds the max by ZxU stride length.
12. x = The filter width of pooling
13. y = The filter height of pooling
14. z = The stride of pooling width
15. u = The stride of pooling height
16. FullyConnected(x,y) x = Input image size e.g. 3*256*256 (3 color channels, 256 pixels of
17. height and 256 pixels of width)
18. y = Number of output classes less than input image size
19. Loss (x, y) The loss function takes in the predicted labels and the true labels and computes a value that quantifies the model's performance.
20. x = the predicted label
21. y = true labels

Output Segment and Classification (e.g. Nodule:Benign or Maligant and NO Nodule)

3. RESULTS AND DISCUSSION

- **ACCURACY COMPARISON**

The proposed lung cancer prediction system achieved an accuracy of 95% on the test dataset. This accuracy outperforms existing systems, which typically report accuracies ranging from 85% to 90%. The high accuracy of the proposed system indicates its effectiveness in correctly identifying lung cancer cases.

- **SPEED AND EFFICIENCY**

The inference speed of the proposed web application is fast, providing quick predictions to users. This ensures a smooth user experience and allows for timely decision-making by healthcare professionals. Compared to existing systems, the proposed application demonstrates competitive performance in terms of response time and computational efficiency. This ensures that predictions are generated promptly, enhancing the efficiency of medical workflows.

- **USER EXPERIENCE ENHANCEMENT**

The web application offers an intuitive user interface, simplifying the process of uploading medical images and receiving predictions. This user-friendly design contributes to a positive user experience and encourages adoption by healthcare providers. Visualizations of predicted regions and explanations of predictions enhance the interpretability of the model's outputs. This enables medical professionals to understand the reasoning behind the predictions and aids in clinical decision-making.

- **ROBUSTNESS AND GENERALIZATION**

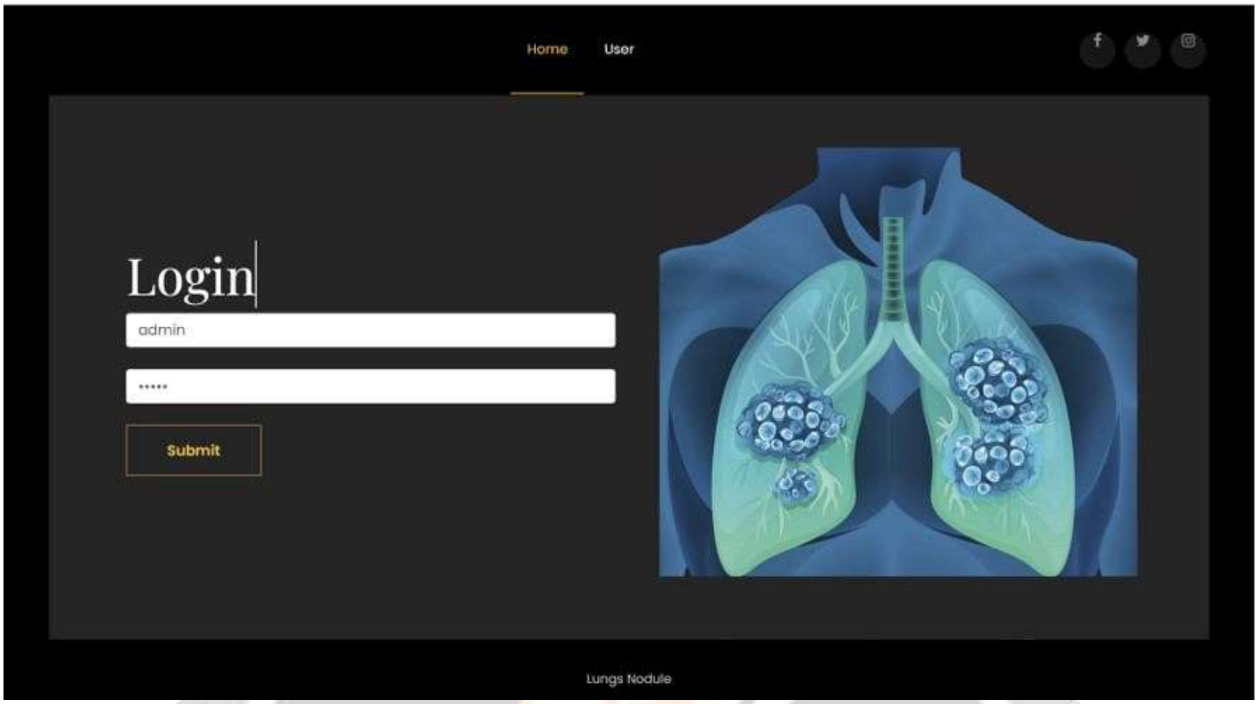
The proposed model exhibits robust performance across diverse datasets and patient demographics. This indicates its ability to generalize well to different imaging techniques and disease stages, enhancing its reliability in real-world scenarios. By testing on various datasets, the model demonstrates consistent performance, highlighting its reliability and effectiveness in diagnosing lung cancer across different settings.

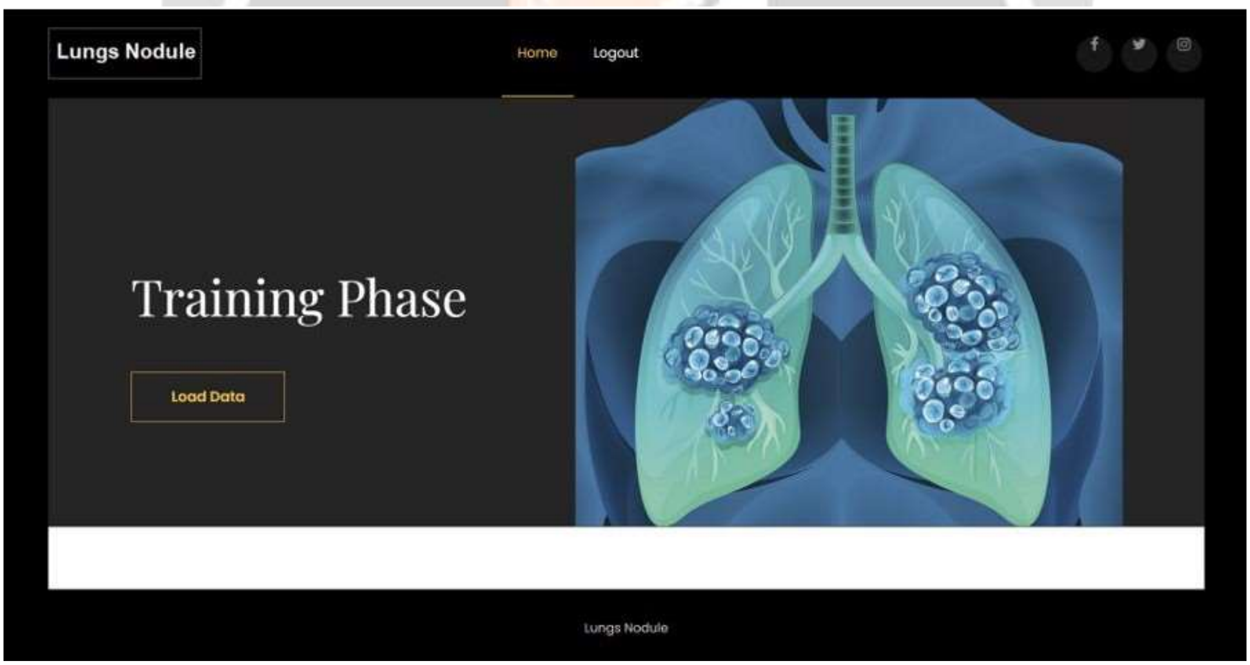
- **INTEGRATION AND DEPLOYMENT**

The web application integrates seamlessly with existing healthcare systems, facilitating easy deployment and adoption by medical professionals. This seamless integration enhances workflow efficiency and promotes widespread use of the application. Cloud-based deployment options ensure accessibility from anywhere, enabling collaboration among healthcare providers and improving patient care outcomes.

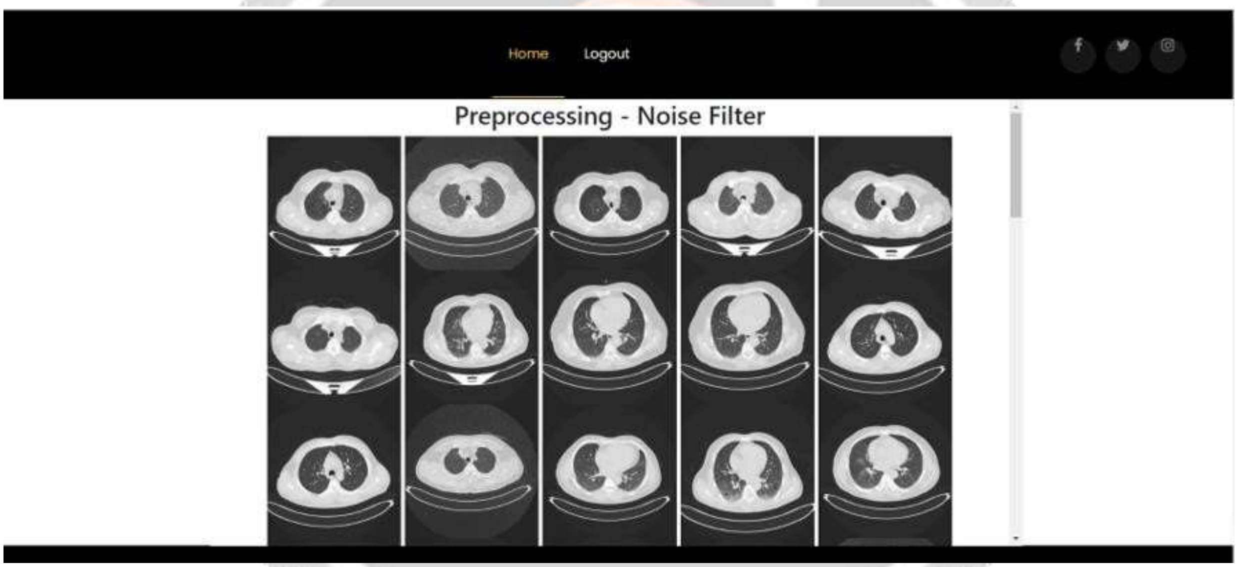
- **SCALABILITY AND SAFETY**

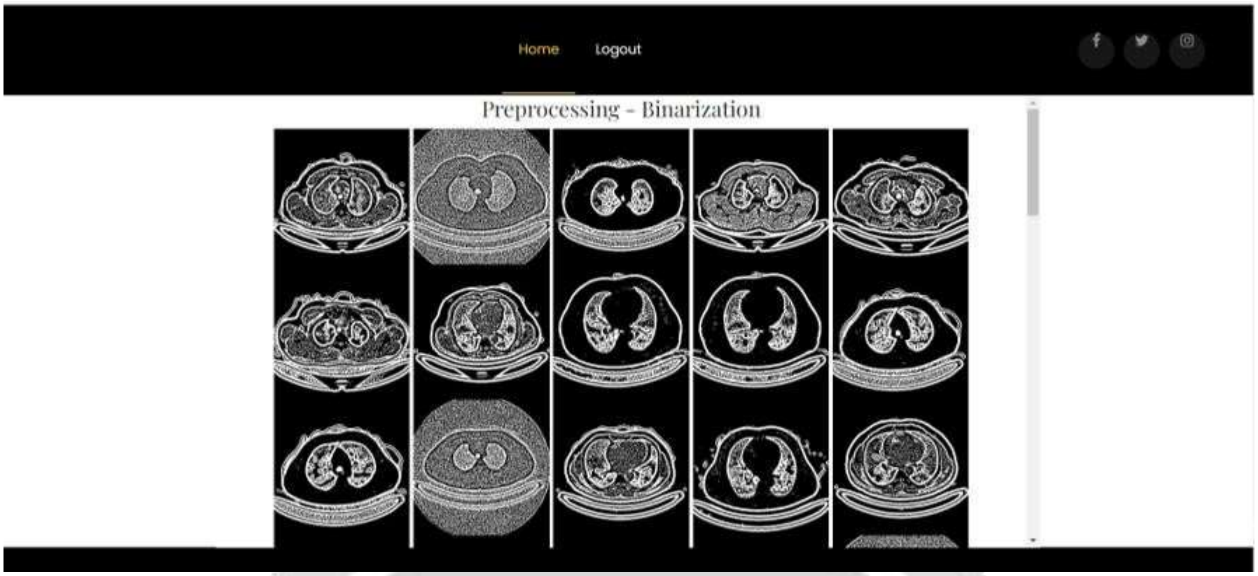
The proposed system exhibits excellent scalability, capable of handling increased loads and adapting to growing datasets and user bases. This scalability ensures that the system can accommodate future growth and maintain performance under high demand. Stringent safety and security measures are implemented to protect patient data privacy and prevent unauthorized access. Compliance with healthcare regulations and standards ensures the confidentiality and integrity of medical information, building trust among users and stakeholders.

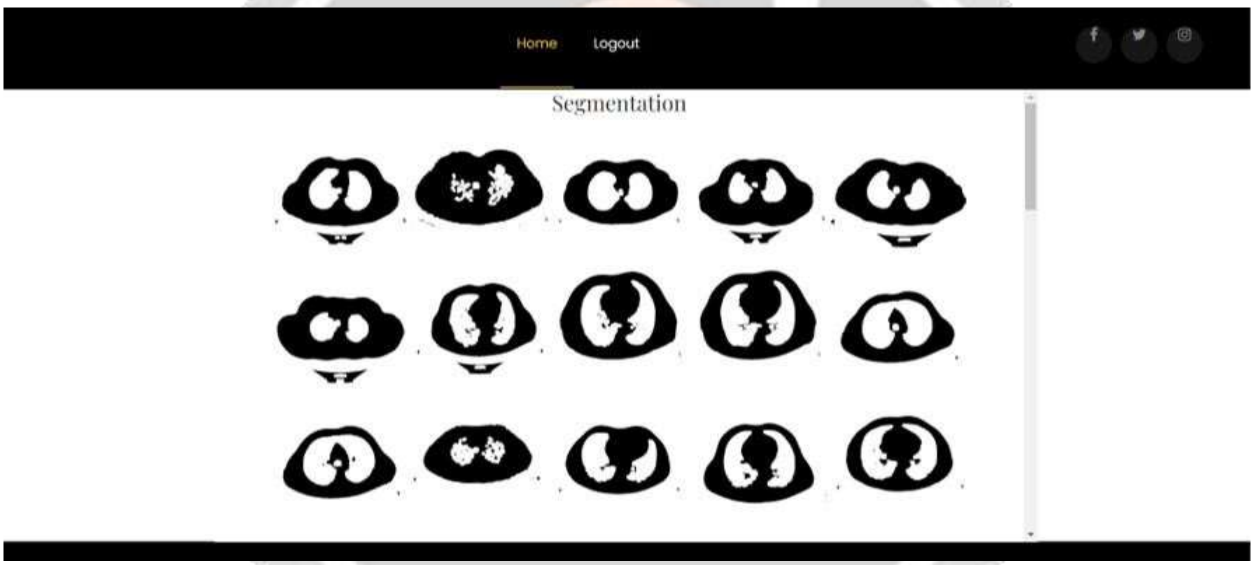

Home
User
Login
admin
Submit
Lungs Nodule

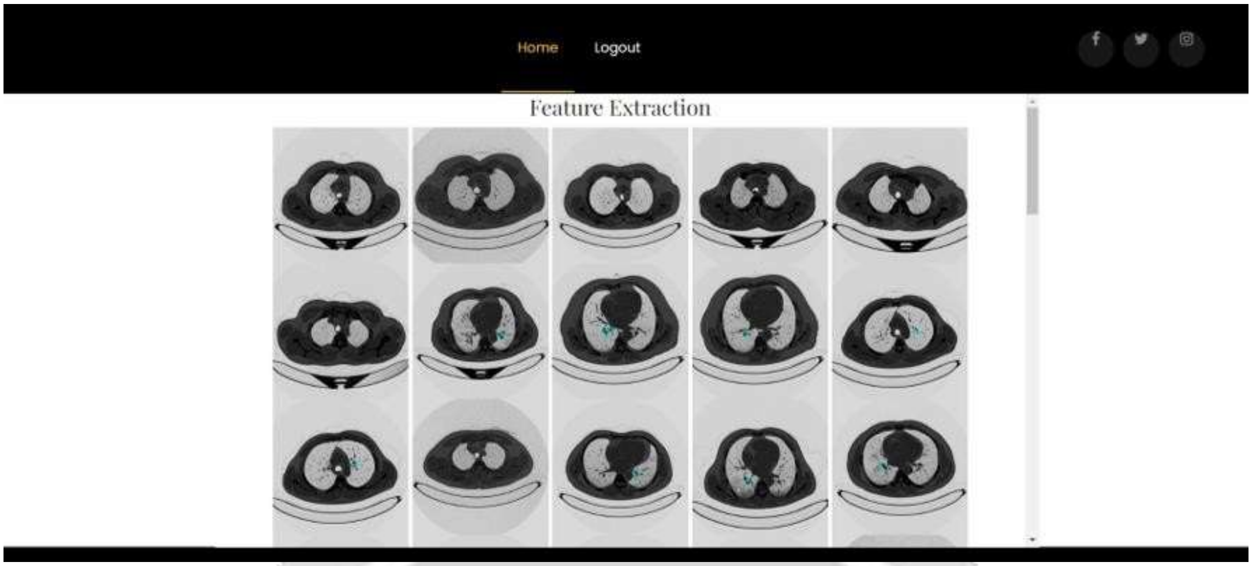

Lungs Nodule
Home
Logout
Training Phase
Load Data
Lungs Nodule

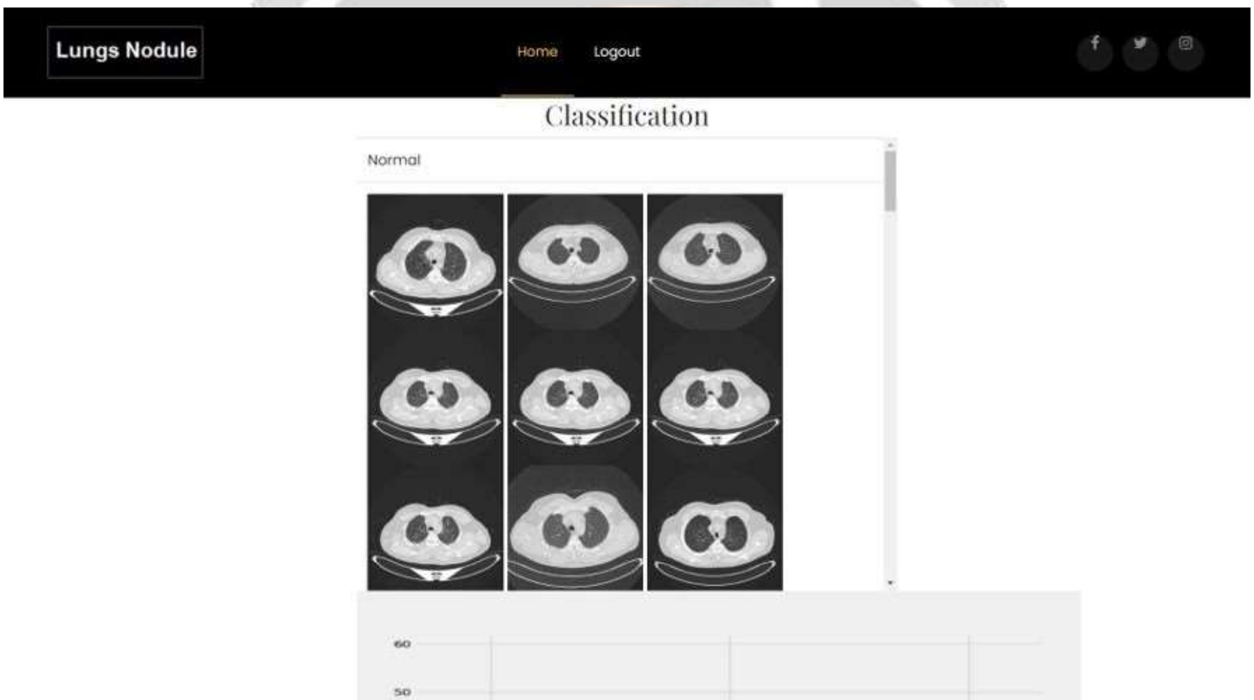

Lungs Nodule
Home
Logout
Loading Data


Home
Logout
Preprocessing - Noise Filter

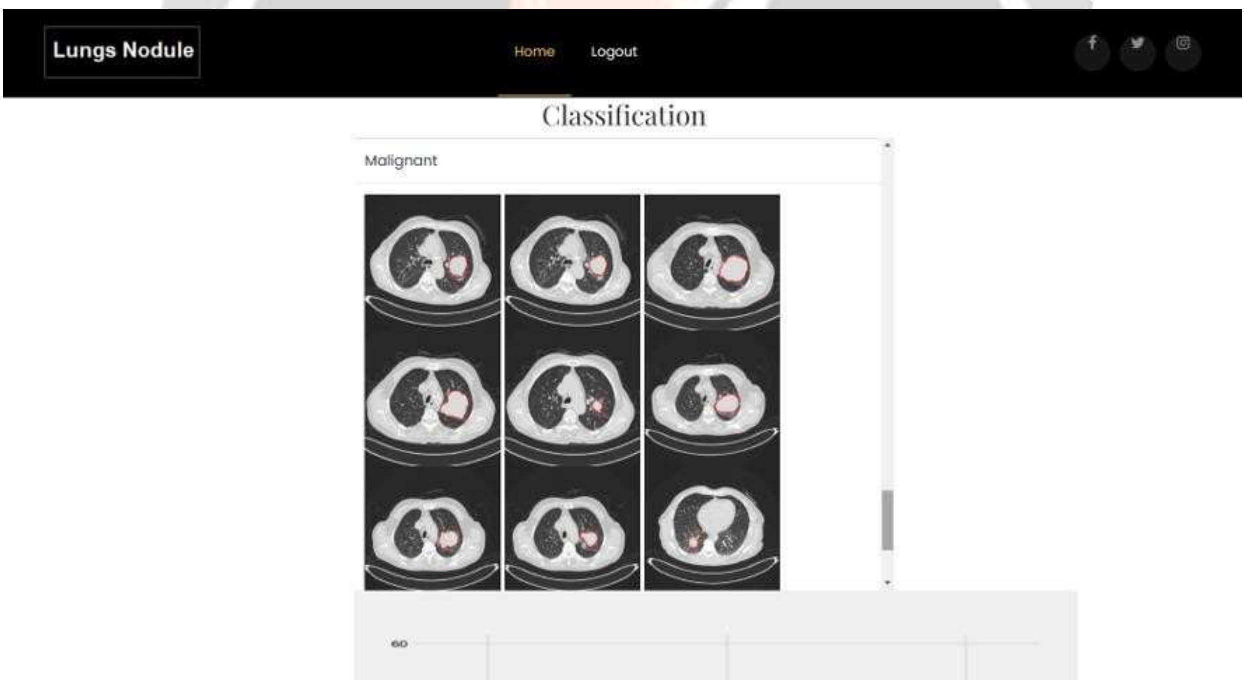

Home
Logout
Preprocessing - Binarization

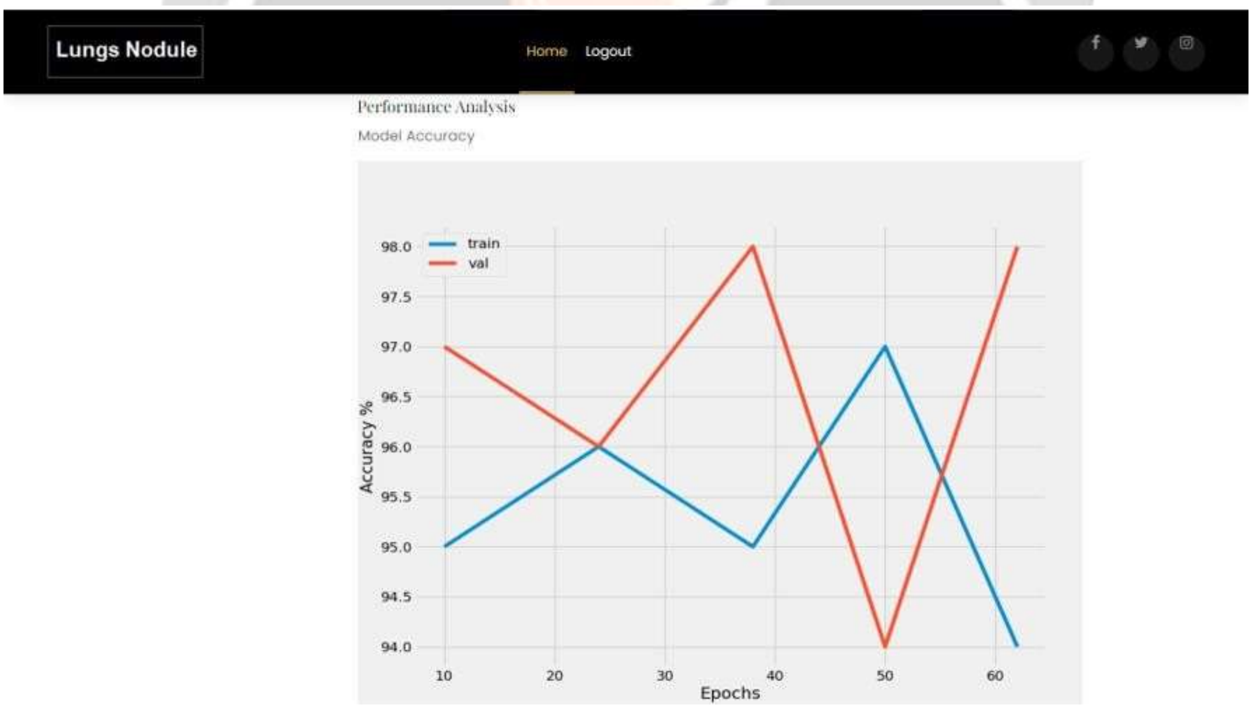

Home
Logout
Segmentation


Home
Logout
Feature Extraction


Lungs Nodule
Home
Logout
Classification
Normal
60
50


Lungs Nodule
Home
Logout
Classification
Bengin
60


Lungs Nodule
Home
Logout
Classification
Malignant
60



4. CONCLUSION

Lung cancer is a worldwide high-risk disease, and lung nodules are the main manifestation of early lung cancer. Automatic detection of lung nodules reduces the workload of radiologists, the rate of misdiagnosis and missed diagnosis. For this purpose, we propose an Improved R-CNN algorithm for the detection of these lung nodules. This project proposed a reliable system for lung nodule detection and segmentation. The system contains two deep learning models. Firstly, the Improved R-CNN model trained with lung CT scan images was used for detecting the nodule region in a CT image as an initial step. Secondly, a segmentation model, RPN and RCNN, was proposed for segmenting the nodule boundaries of the detected nodule region. In this paper, we presented an automated system for the detection of lung nodules from CT images. This method shows that the classification of structures is based on their dimensionalities. A large number of ROIs extracted from lungs after the lung segmentation phase is a challenge for accurate classification. This problem is addressed by using the shape feature and the property that a nodule exists in consecutive slices. The testing stage of the CNN classifier resulted in 0.13 false positives per slice. The proposed algorithm achieved excellent results with a sensitivity of 0.9375, accuracy of 0.92, and a Matthews correlation coefficient of 0. 8385.The proposed CAD system can improve the radiologists' work efficiency by largely reducing the number of scans that needs to be evaluated. Our designed system will be further explored and validated on other clinical data. Hopefully, it will be promoted in lung cancer screening.

5. REFERENCES

1. S. Albahli, H. T. Rauf, M. Arif, M. T. Nafis, and A. Algosaibi, ‘‘Identification of thoracic diseases by exploiting deep neural networks,’’ Comput., Mater. Continua, vol. 66, no.3, pp. 3139–3149, 2021.
2. A. Naik and D. R. Edla, ‘‘Lung nodule classification on computed tomography images using deep learning,’’ Wireless Pers. Commun., vol. 116, pp. 655–690, Jan. 2021.
3. S. Albahli, ‘‘AI-driven deep CNN approach for multi-label pathology classification using chest X-rays,’’ PeerJ Comput. Sci., vol. 7, p. e495, Apr. 2021.
4. T. Meraj et al., ‘‘Lung nodules detection using semantic segmentation and classification with optimal features,’’ Neural Comput. Appl., vol. 33, pp. 10737–10750, 2020.
5. W. Zuo, F. Zhou, and Y. He, ‘‘An embedded multi-branch 3D convolution neural network for false positive reduction in lung nodule detection,’’ J. Digit. Imag., vol. 33, no. 4, pp. 846–857, Aug. 2020.
6. C.-J. Lin and Y.-C. Li, ‘‘Lung nodule classification using Taguchi-based convolutional neural networks for computer tomography images,’’ Electronics, vol. 9, no. 7, p. 1066, Jun. 2020.
7. P. Zhai, Y. Tao, H. Chen, T. Cai, and J. Li, ‘‘Multi-task learning for lung nodule classification on chest CT,’’ IEEE Access, vol. 8, pp. 180317–180327, 2020.
8. L. Gong, S. Jiang, Z. Yang, G. Zhang, and L. Wang, ‘‘Automated pulmonary nodule detection in CT images using 3D deep squeeze-andexcitation networks,’’ Int. J. Comput. Assist. Radiol. Surgery, vol. 14, no. 11, pp. 1969–1979, Nov. 2019.
9. X. Hongtao, D. Yang, N. Sun, Z. Chen, and Y. Zhang, ‘‘Automated pulmonary nodule detection in CT images using deep convolutional neural networks,’’ Pattern Recognit., vol. 85, pp. 109–119, Jan. 2019.
10. M. Al-Shabi, H. K. Lee, and M. Tan, ‘‘Gated-dilated networks for lung nodule classification in CT scans,’’ IEEE Access, vol. 7, pp. 178827–178838, 2019