Lung Cancer Detection Using Deep Learning Algorithms

Ritesh Jaiswal[1], Pooja Shukla[2], Mohini Varpe[3], Pranay Krishna More[4]

[1,2,3,4] *Organization Name:- VPPCOE & VA*
Department :- IT
University Name:- University Of Mumbai

Abstract: One of the leading causes of death worldwide, in both men and women, is lung cancer. According to WHO, the estimated number of lung cancer cases per year is two million. The overall 5-year survival rate for lung cancer patients increases from 16 to 56% if the disease is detected in time Computed Tomography (CT) scan can provide valuable information in the diagnosis of lung diseases This work's primary goals are to identify cancerous lung nodules from the provided input lung image and to categorise lung cancer according to its severity. This study employs cutting-edge Deep learning techniques to locate the malignant lung nodules. This study employs cutting-edge Deep learning techniques to locate the malignant lung nodules. Cancer patients' CT scanned lung images are obtained from various facilities. using image processing techniques like pre-processing, segmentation techniques such as watershed algorithm and feature extraction, area of interest is separated. Features such as texture, geometric, volumetric and intensity features are extracted. Finally, these features are classified using CNN.

***Keywords**- Lung Cancer, CT, Deep Learning, Watershed, CNN. (Key words)*

I. INTRODUCTION

Lung cancer disease is the second largest death threat to the world after heart attack, as this cancer is responsible for the largest number of deaths, compared to the number of deaths caused by any other cancer type. [1]. Lung cancer is characterised by unchecked cell proliferation that results in the development of lung nodules. It is reported that lung cancer is responsible for around 19% deaths globally mostly due to alcohol and tobacco consumption. The rate of survival is assured by only 15% survival chances, for a survival period of 5 years. [2]. The main reason for such a high fatality rate is because therapy is delayed due to discovery occurring at a later stage. Chances of survival can rise by 50–70% if lung cancer is discovered sooner. Non-small cell lung cancer and small cell lung cancer are the two major groups into which the lung cancer can be classified based on the cell characteristics. [7] non-small cell lung cancer is the most common type of lung cancer contributing to about 85-90% of total lung cancer cases, while the other 10-15% of the cases is diagnosed with small cell lung cancer. The leading cause of cancer-related deaths worldwide is lung cancer. The ultimate stage of lung cancer is when the symptoms become apparent so it is very tough to identify in its beginning stage. Because of this, compared to all other cancer forms, lung cancer has a particularly high mortality rate. The two kind of lung disease which develop and spread in an unexpected way, are small cell lung malignancies (SCLC) and non-little cell lung tumors (NSCLC) [1]. The phase of lung disease alludes to the degree to which the growth has spread in the lung. The World Health Organization reported that more than 7.6 million people worldwide lost their lives to lung cancer each year. Moreover, the death rates of lung cancer are expected upon to keep rising, to wind up around 17 million worldwide in 2030[2]. Despite being the best imaging tool in the medical sector, clinicians find it challenging to interpret and detect cancer from CT scan data. In year of 2005, around 1,362,825 new cancer cases are expected and around 571,590 deaths are expected to happen due to cancer in the United States. It was evaluated that there will be 162,921 deaths from lung cancer, which occurs 30% of all cancer deaths. [3] The extent of the spread of cancer is the basis for the division of lung cancer into stages. It comprises of four stages namely stage I-The cancer is confined to the lung, stages II and III-the cancer is confined to the chest (with larger and more invasive tumor classified as stage III) and Stage IV-Cancer has spread from the chest to other parts of the body.

There are many techniques to diagnose the lung cancer such as X-rays, Computed Tomography (CT), Magnetic Resonance Imaging (MRI scan), and Sputum Cytology. The problem with these techniques is that it can be time consuming and makes detection possible only at later stages. Despite being the best imaging tool in the medical sector, clinicians find it challenging to interpret and detect cancer from CT scan data. Hence, computer assisted diagnosis might be useful for clinicians to precisely identify the malignant cells. Computer aided techniques such as Deep learning and image processing have been implemented. In our proposed algorithm we have tried to solve these problems. Our developed algorithm can detect cancer affected cell and the corresponding stage such as initial, middle, or final stage. If no cancer affected cell is found in the input image, then it checks the probability of lung cancer.

II. PROJECT STATEMENT

For many diagnostic and therapeutic purposes, automatic fault detection in CT scans is crucial Tumor segmentation and classification are exceedingly difficult because of the large amount of data present in CT scans and the hazy boundaries. To improve accuracy, yield, and speed up diagnosis, one automatic lung cancer detection method has been introduced in this work.

III. LITERATURE SURVEY

- 2019-Arnaud A. A. Setio, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J. van Riel, Mathi-Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks.
- January 2018-Anum Masood,Bin Sheng-Computer Assisted Decision Support System in Stage Classification on CT images
- January 2019-Moffy Vas1, Amita Dessai2-Lung cancer detection system using lung CT image processing
- October 2020 -Bijaya Kumar Hatuwal1, Himal Chand Thapa2 -Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images

PROPOSED ARCHITECTURE

1) Data collection

Images are collected from the hospitals or from google. The CT images of lungs acquired from the hospital database. We will analyze how CNN algorithm helps us to distinguish between cancerous and non-cancerous images.

2) Preprocessing

Cropping of the image in first step is done to eliminate the unwanted portions from the image Next median filters are applied to the images, which are basically used to get rid of the salt and pepper noise present in the images. The usage of a 3*3 median filter and its contribution to the improvement of the photographs.

3) Segmentation

It is necessary to convert the images to binary in order to morphologically segment the lungs, which also lessens computational complexity and storage problems. The opening operation using the periodic line structuring element tends to remove some of the foreground pixels from the edges of the region of foreground pixels.

4) Feature extraction

Feature extraction helps in extracting out significant items of data which serve as an input to the classifier. The first step is to resize the image into three different resolutions.

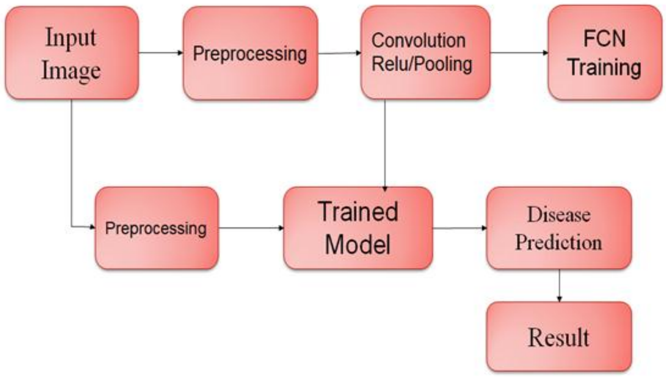


Fig 1.1 Architecture Of System

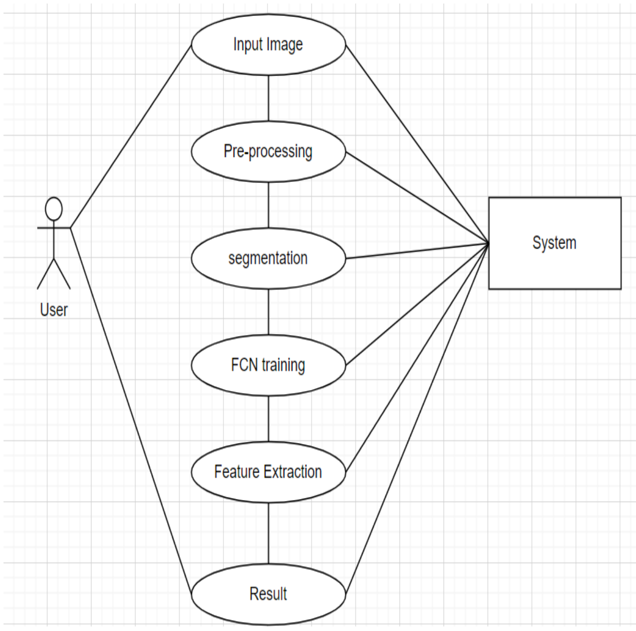


Fig 2.1 Use Case Diagram

Methodology

For lung cancer Detection we use CNN methodology to find accurate result. The methodology adopted in this project was carried out in five steps which are shown with the help of a flow chart in Fig. shows Each step of the flow chart is explained.

Step 1: Convolution Operation

The three components that go into the convolution are as follows:

1. Input image
2. Feature detector
3. Feature map

Two CNN architectures were tested for patch classification into normal and tumor class: VGG, the winner of ImageNet competition 2014, and Res Net which won ImageNet challenge in 2015. More precisely, here we used VGG16 (with 16 layers) and ResNet50 (with 50 layers). In VGG architecture all convolutional layers use filters with small receptive field of size 3x3. Part of convolutional layers are followed by max spatial pooling over 2x2 window and stride 2 what effectively down samples resolution by factor 2. 2 fully connected layers, each having 4096 channels, and layer with 2 sigmoid activated outputs corresponding to normal and tumor class are connected to the last pooling layer. Overview of VGG16 CNN is shown in figure.

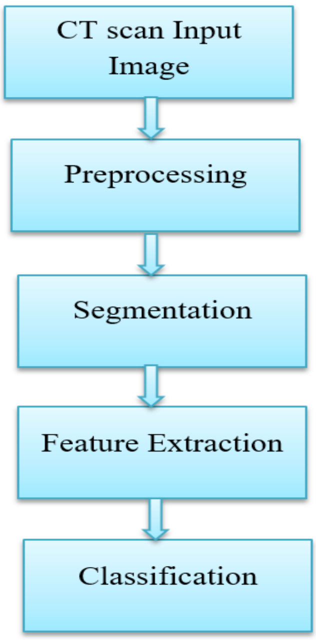


Fig 3.1 Methodology Block Diagram

Res Net architecture is built with Res Net blocks to solve the accuracy degradation problem when creating deeper architectures with more layers. Instead of learning mapping function, the residual block fits residual mapping. Main hypothesis here is that it is easier to find residual mapping that original one. If H(x) denotes original mapping (figure 3), stacked layers models mapping F(x) = H(x) ⍰ x. At the top of ResNet50 network we added average pooling layer followed by fully connected layer with 2 sigmoid activated outputs. For both tested architectures weights pre-trained on ImageNet are used for initialization to speed-up convergence. Using outputs from CNN, 256-times down sampled tumor heat map is created by assigning each patch probability of being tumor.

Step 1(b): RELU Layer

We want to accomplish that since photos are inherently non-linear. Any image you look at will show you that it has a lot of non-linear elements (example, the transition between pixels, the borders, the colors, etc.). In order to compensate for whatever linearity we might impose on an image when we run it through the convolution procedure, the rectifier works to further break up the linearity.

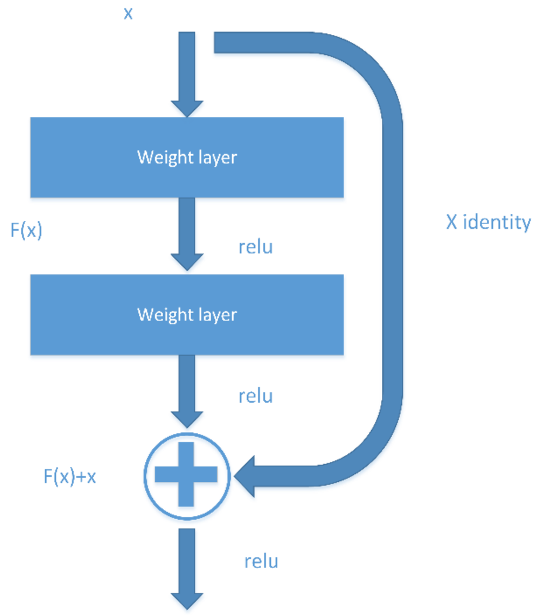


Fig 3.2 ResNet Building Block

Step 2: Pooling

Again, max pooling is concerned with teaching your convolution neural network to recognize that despite all of these differences that we mentioned, they are all images are same. In order to do that, the network needs to acquire a property that is known as "spatial variance." This property makes the network capable of detecting the object in without being misled by variations in the textures, distances from the locations where they were taken, angles, or anything else.

Step 3: Flattening

This will be a brief breakdown of the flattening process and how data move from pooled to flattened layers when working with Convolutional Neural Networks.

Step 4: Pooling

You end up with a long vector of input data after the flattening stage, which you then run through the artificial neural network to create it processed further which is called pooling. Types of pooling: Mean, Max, Sum

CT scan

Computer technology is primarily used in CT scans to create or display digital images of the human body's internal organs, which aids medical professionals in visualising the body's interior. The advanced x-ray and computer technology used in a CT scan. A combination of soft tissue, bone, and blood arteries may be visible on a CT scan. Certain malignancies can be identified with CT scans, which can also aid in the detection of edoema, bleeding, and bone and tissue calcification. The contrast material used during a CT scan is often iodine. The CT scan image of lung cancer is an input for this proposed algorithm. The image from the CT scan is hazy. The noise is present in this image. Noise disturbances may cause because of electronic imaging sensors, sensor temperature, insufficient Light levels, film granularity, and channel noise. So preprocessing is essential for such images to remove blurriness from it and make it sharper.

Algorithm

We propose CT Scan image quality enhancement and its application using CNN Algorithm. For developing dependable and ordinary techniques to identify the brain tumor, extract the quality of it for medicinal determination, visualization, and the presence forecast. IT is Robust and scalable CNN based image segmentation and features extraction by considering different types of the dataset with minimum computation efforts. The use of appropriate feature extraction and reduction models may help to reduce the detection time and improving the accuracy.

Steps of CNN algorithm:

- Arrange material in folders with labels, such as photographs of lung cancer.
- Read dataset
- Read the attributes of each image and label it with the dataset folder's name below.
- Store it in model file
- Get input image
- Read features of input image
- Compare features of stored features
- Display the label as a near-matching feature prediction.

Lung cancer Detection Comprises of the following steps:-

- 2-D ConvNets.

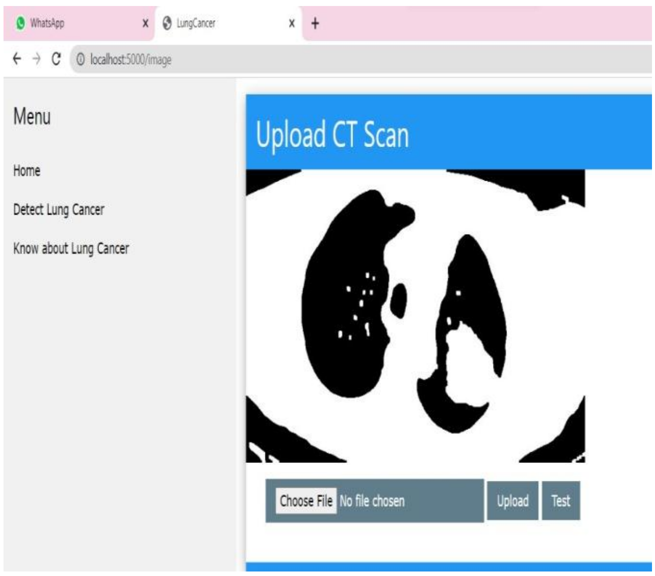
  Pulmonary detection using 2-D ConvNets.
- Deep Embedded Clustering (DEC)

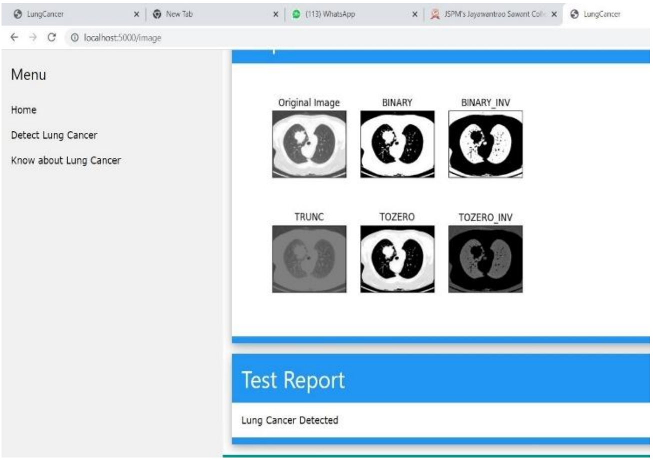
  Clustering the Pulmonary and non-pulmonary area using Deep Embedded Clustering (DEC)
- Spherical Harmonics Function

  Detection of mesh malignant and benign using Spherical Harmonics Module Gradient
- Segmenting the background image using a gradient back-propagation throughout the deep networks.

Results

This work's primary goals are to identify cancerous lung nodules from the provided input lung image and to categorise lung cancer according to its severity. This study employs cutting-edge Deep learning techniques to locate the malignant lung nodules. Cancer patients' CT scanned lung images are obtained from various facilities. Using image processing techniques like pre-processing, segmentation techniques such as watershed algorithm and feature extraction, area of interest is separated. Features such as texture, geometric, volumetric and intensity features are extracted. Finally, these features are classified using CNN.





Conclusions

Automated system for lung cancer detection. Application of median Filter to eliminate impulse noise in the images proved to be a success. The morphological operations also contributed towards satisfactory results in the process of segmentation. Artificial neural networks proved to be a good classifier with acceptable accuracy. The methodology adopted in this project resulted in an accuracy of 92% for the hospital database. This system aims at increasing the accuracy and speed of the lung cancer detection system. It also helps in detecting the cancer at earlier stages.

REFERENCES

[1] Bijaya Kumar Hatuwal1, Himal Chand Thapa2, "Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images", "International Journal of Computer Trends and Technology" October (2020).

[2] Vaishnavi. D1, Arya. K. S2, Devi Abirami. T3, M. N. Kavitha4, "Lung Cancer Detection using Machine Learning", "International Journal of Engineering Research & Technology (IJERT)", Special Issue – (2019).

[3] Alan L. Yuille4 "Deep Supervision for Pancreatic Cyst Segmentation in Abdominal CT Scans" (2018).

[4] Moffy Vas1, Amita Dessai 2, "Lung cancer detection system using lung CT image processing", IEEE, (2017).

[5] Mario Buty1, Ziyue Xu1, Mingchen Gao" Characterization of Lung Nodule Malignancy using Hybrid Shape and Appearance Features" (2017).

[6] Buty, M., Xu, Z., Gao, M., Bagci, U., Wu, A., Mollura, D.J.: Characterization of Lung Nodule Malignancy Using Hybrid Shape and Appearance Features. In: MICCAI. pp. 662–670. Springer (2016).

[7] Mohsen Keshani, Zohreh Azimifar and Reza Boostani, "Lung nodule segmentation using activecontour modeling" MVIP, IEEE, (2016).

[8] Gawade Prathamesh Pratap, R.P Chauhan, "Detection of lung cancer cells using image processing techniques ", 1st IEEE International Conference on Power Electronics, Intelligent control and energy systems, IEEE, (2016)

[9] Arnaud A. A. Setio, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J. van Riel, Mathi"Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks"(2016).

[10] Junyuan Xie, Ross Girshick "Unsupervised Deep Embedding for Clustering Analysis" (2016).

[11] Badrul Alam Mia, Mohammad Abu Yusuf, "Detection of lung cancer from CT image using image processing and neural network", International conference on Electrical Engineering and Information Communication Technology (ICEEICT), IEEE, May, (2015).

[12] K Punithavathy, M M Ramya, Sumathi Poobal, "Analysis of statistical texture features for automatic lung cancer detection in PET/CT images", International Conference on Robotics, Automation, Control and Embedded systems (RACE), IEEE ,18-20 February (2015).

[13] Muhammed Anshad, S.S Kumar, Recent methods for the detection of tumor using computer aided diagnosis", International Conference on Control, Instrumentation, Communication and Computational technologies (ICCICCT), IEEE transactions, (2014).

[14] Nooshin Hadavi, Md Jan Nordin, Ali Shojaeipour, "Lung cancer diagnosis using CT-scan images based on cellular learning automata", International conference on Computer and Information Sciences (ICCOINS), IEEE, (2014).

[15] Anita Chaudhary, Sonit Sukhraj Singh "Lung cancer detection on CT images using image processing", computing sciences 2012 international conference, IEEE, (2012).

[16] Robert M Haralick, K.Shanmugam, Itshak Dinstein, "Textural Features for_ Image Classification" IEEE Transactions on Systems, Man and Cybernetics, 3(6), pp. 610- 621, (1973).