



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 6, June 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Implementation on Automatic Detection & Classification of Lung Cancer Using CNN

Dhanshri Dilip Sarwade, Prof. S.G. Shinde

Department of Electronics and Telecommunication, T.P.C.T's College of Engineering, Dharashiv, Maharashtra, India

ABSTRACT: Lung cancer is influenced by a variety of medical conditions that differ based on location. The early detection of lung cancer is critical if the goal is to reduce the high death rate. The global lung screening programme focuses on imaging PET/CT exams among the most mature populations at risk in order to improve early detection rates. Despite the use of invasive techniques, adverse effects seldom appear until infection has progressed, making it difficult for radiologists to identify sores. Every year, the American Cancer Society assesses the number of new cancer cases and deaths that will occur worldwide in the current year and compiles the most recent data on tumour frequency, mortality, and survival. Genuine and exact information is the foundation of disease control efforts. Tobacco use has been linked to more than 3/4 of the disease cases. Furthermore, inherited factors, exposure to environmental toxins, and secondhand smoke all contribute to the rapid spread of sickness. Cures such as chemotherapy, radiation, surgery, and epidermal open medicines improve survival rates and personal happiness. This technique focuses on diagnosing at important stages and ahead of schedule using sophisticated computational procedures that eliminate noise using segmentation algorithms and computations, which is the fundamental concept of digital image processing. The location of CT images obtained from cancer research organisations is studied using MATLAB.

KEYWORDS: Lung cancer, CNN, MATLAB

I. INTRODUCTION

Tumour is a significant global health concern and the second leading cause of mortality in all age groups. Cancer may be defined as uncontrolled cell growth with the ability to spread throughout the body. Our bodies contain both red blood cells (RBCs) and white blood cells. The primary role is to transport new oxygen (O₂) throughout the body via blood flow, which causes blood to appear red. [1] In the lungs, tissue only absorbs oxygen (O₂) from RBCs. The hereditary composition of erythrocytes contains a high concentration of haemoglobin. The cell film is made up of proteins and lipids, and it serves as the foundation for physiological cell capability. They lack an essential component of the cell, haemoglobin. Each second, around 20 lakh new RBCs are generated. [2] The cells are delivered in the bone marrow and circulate throughout the body for approximately 4 months, alternating between arteries and veins.

Image processing is used to minimise the size of images while maintaining quality. These actions do not increase the likelihood of data content; rather, they decrease it if entropy is a data metric. The primary goal of preparation is to increase pixel power by transitioning from discrete to computerised images, sectioning to pixels, performing numerical operations on pixels, and reproducing images with higher quality. [11] Pre-processing of CT images is the first step in image evaluation, followed by division handling and certain morphological procedures to identify tumour spots/cells. It can also be used to determine the extent of malignancy spread, i.e. how much of the lung is affected by the illness. The morphological processes are fundamentally related by comparing the size and status of the malignant cell to that of the normal cell, and then the contaminated cell images are shown on a dim scale image with the maximum intensity.

II. PROBLEM STATEMENT

In recent years, this instrument has received a lot of attention, particularly in the health industry. With efforts like as the LUNA Challenge in 2016 and Kaggle's Data Science Bowl 2017, people all across the world are working to make radiologists' and oncologists' jobs simpler, as well as healthcare services more accessible and affordable. Hospitals in Kosovo are also overcrowded, making it more difficult for physicians and patients to work. Furthermore, patients must wait for extended periods of time to receive a good diagnosis, forcing many to travel overseas in search of alternative diagnoses.

In desperate need of an effective solution that reduces the radiologist's time in detecting, classifying cancer, and writing reports for patients, an application that could reduce the time the day-to-day time-consuming activities doctors have to go through in diagnosing a malignant and benign cancer is not only helpful but would also save lives.

III. ALGORITHM DETAILS

CNN

Neural networks are a set of algorithms, modeled loosely after the human brain, that are designed to recognize patterns. They interpret sensory data through a kind of machine perception, labeling or clustering raw input. The patterns they recognize are numerical, contained in vectors, into which all real-world data, be it images, sound, text or time series, must be translated.

Neural networks help us cluster and classify. You can think of them as a clustering and classification layer on top of the data you store and manage. They help to group unlabeled data according to similarities among the example inputs, and they classify data when they have a labelled dataset to train on. (Neural networks can also extract features that are fed to other algorithms for clustering and classification; so you can think of deep neural networks as components of larger machine-learning applications involving algorithms for reinforcement learning, classification and regression.)

Deep learning maps inputs to outputs. It finds correlations. It is known as a “universal approximator”, because it can learn to approximate an unknown function $f(x) = y$ between any input x and any output y , assuming they are related at all (by correlation or causation).

Classification

All classification tasks depend upon labeled datasets; that is, humans must transfer their knowledge to the dataset in order for a neural network to learn the correlation between labels and data. This is known as supervised learning.

- Detect faces, identify people in images, recognize facial expressions (angry, joyful)
- Identify objects in images (stop signs, pedestrians, lane markers...)
- Recognize gestures in video
- Detect voices, identify speakers, transcribe speech to text, recognize sentiment in voices
- Classify text as spam (in emails), or fraudulent (in insurance claims); recognize sentiment in text (customer feedback)

Any labels that humans can generate, any outcomes that you care about and which correlate to data, can be used to train a neural network.

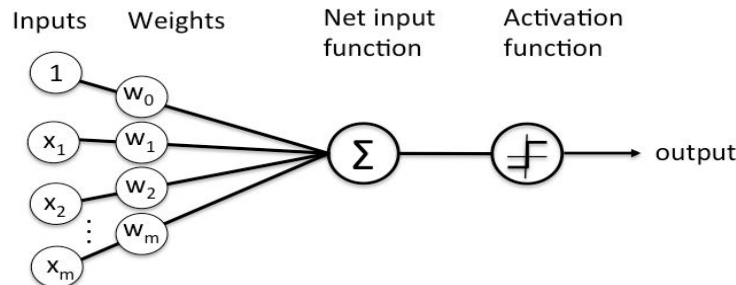
Clustering

Clustering or grouping is the detection of similarities. Deep learning does not require labels to detect similarities. Learning without labels is called unsupervised learning. Unlabeled data is the majority of data in the world. One law of machine learning is: the more data an algorithm can train on, the more accurate it will be. Therefore, unsupervised learning has the potential to produce highly accurate models.

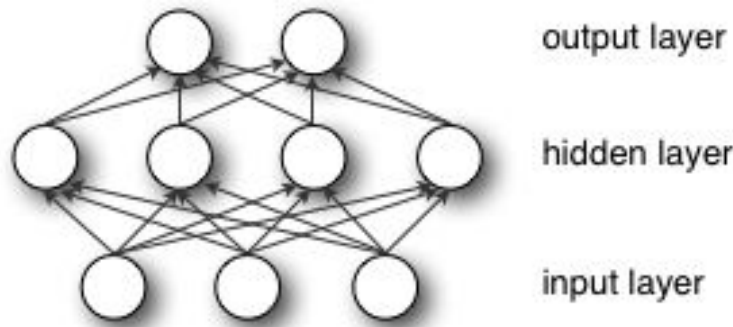
- Search: Comparing documents, images or sounds to surface similar items.
- Anomaly detection: The flipside of detecting similarities is detecting anomalies, or unusual behavior. In many cases, unusual behavior correlates highly with things you want to detect and prevent, such as fraud.

Neural Network Elements

Deep learning is the name we use for “stacked neural networks”; that is, networks composed of several layers. The layers are made of nodes. A node is just a place where computation happens, loosely patterned on a neuron in the human brain, which fires when it encounters sufficient stimuli. A node combines input from the data with a set of coefficients, or weights, that either amplify or dampen that input, thereby assigning significance to inputs with regard to the task the algorithm is trying to learn; e.g. which input is most helpful is classifying data without error? These input-weight products are summed and then the sum is passed through a node’s so-called activation function, to determine whether and to what extent that signal should progress further through the network to affect the ultimate outcome, say, an act of classification. If the signals pass through, the neuron has been “activated.”



A node layer is a row of those neuron-like switches that turn on or off as the input is fed through the net. Each layer's output is simultaneously the subsequent layer's input, starting from an initial input layer receiving your data.



Pairing the model's adjustable weights with input features is how we assign significance to those features with regard to how the neural network classifies and clusters input.

Deep-learning networks are distinguished from the more commonplace single-hidden-layer neural networks by their depth; that is, the number of node layers through which data must pass in a multistep process of pattern recognition. Deep-learning networks perform automatic feature extraction without human intervention, unlike most traditional machine-learning algorithms. Given that feature extraction is a task that can take teams of data scientists years to accomplish, deep learning is a way to circumvent the chokepoint of limited experts. It augments the powers of small data science teams, which by their nature do not scale.

When training on unlabeled data, each node layer in a deep network learns features automatically by repeatedly trying to reconstruct the input from which it draws its samples, attempting to minimize the difference between the network's guesses and the probability distribution of the input data itself. Restricted Boltzmann machines, for examples, create so-called reconstructions in this manner.

Deep-learning networks end in an output layer: a logistic, or softmax, classifier that assigns a likelihood to a particular outcome or label. We call that predictive, but it is predictive in a broad sense. Given raw data in the form of an image, a deep-learning network may decide, for example, that the input data is 90 percent likely to represent a person.

Artificial Intelligence has been witnessing a monumental growth in bridging the gap between the capabilities of humans and machines. Researchers and enthusiasts alike, work on numerous aspects of the field to make amazing things happen. One of many such areas is the domain of Computer Vision. The agenda for this field is to enable machines to view the world as humans do, perceive it in a similar manner and even use the knowledge for a multitude of tasks such as Image & Video recognition, Image Analysis & Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. The advancements in Computer Vision with Deep Learning has been constructed and perfected with time, primarily over one particular algorithm — a Convolutional Neural Network.

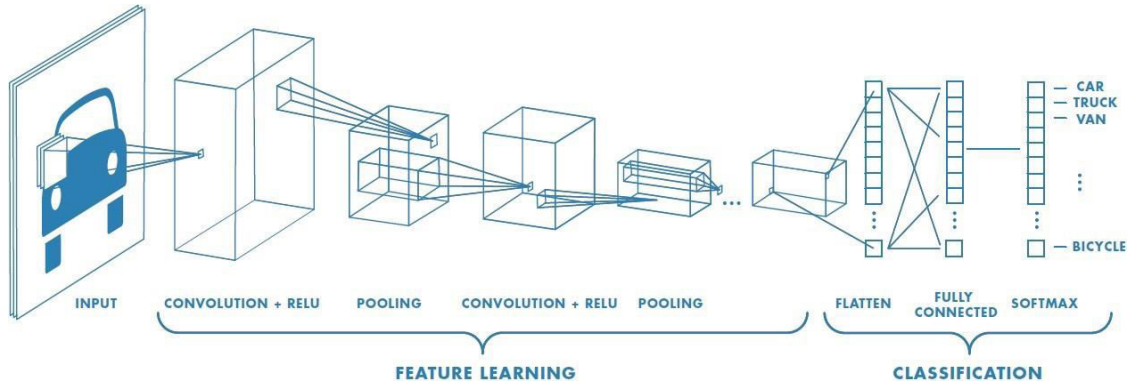


Fig 1 architect CNN

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.

The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

IV. METHODOLOGY

The methodology adopted in this project was carried out in five steps which are shown with the help of a flow chart in Fig.2. Each step of the flow chart is explained below.

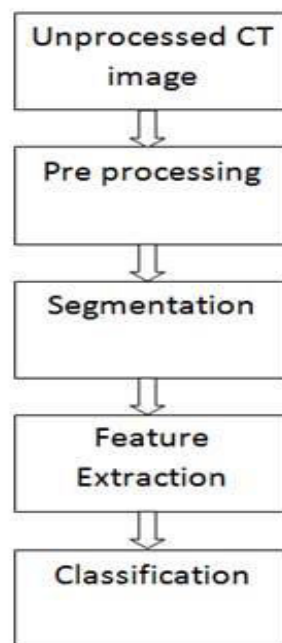


Fig.2 Methodology Block Diagram

1) Data collection

The CT images of lungs acquired from the hospital database are shown in Fig.3. We will analyze how this algorithm helps us to distinguish between cancerous and non-cancerous images.

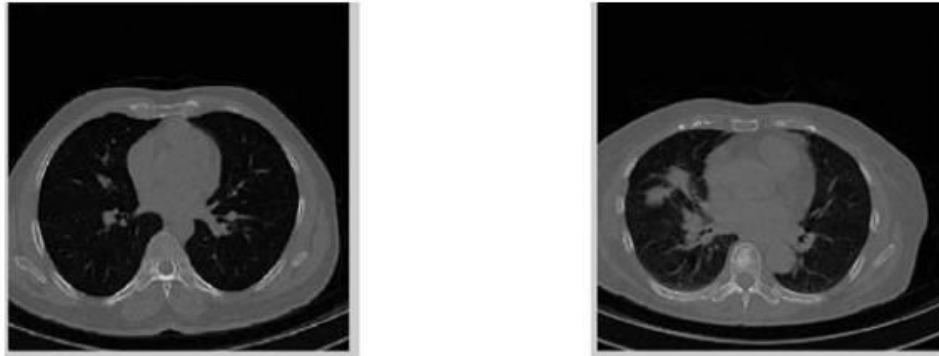


Fig.3 .a. Non- cancerous image

b. Cancerous image

2) Pre-processing

Preprocessing involved the steps shown in Fig.4.

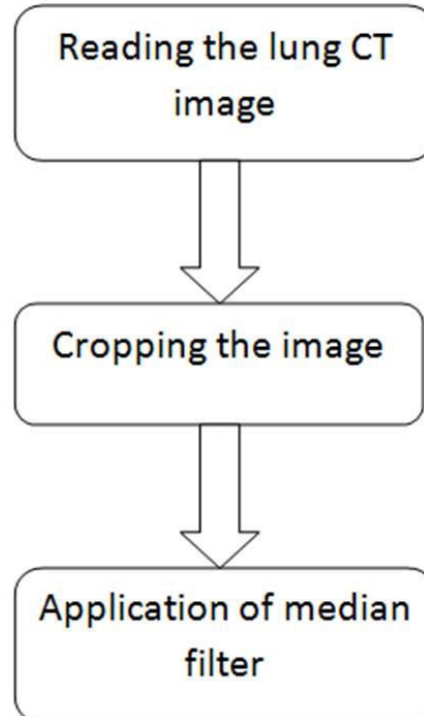


Fig.4 Pre-processing flow diagram

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. A median filter of size 3*3 was used and its contribution towards enhancement of the images is shown in Fig.5.

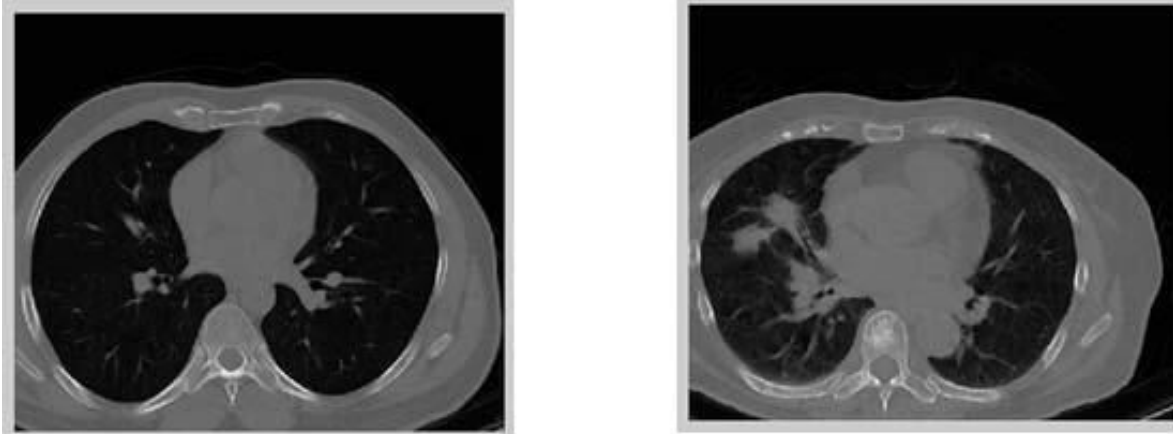


Fig. 5 Median filtered images

3) Segmentation

Segmentation steps are depicted in flow diagram, shown in Fig.6. and each step is discussed in detail. Converting the images to binary reduces computational complexity and storage issues and also is a pre-requisite for morphological segmentation of lungs.

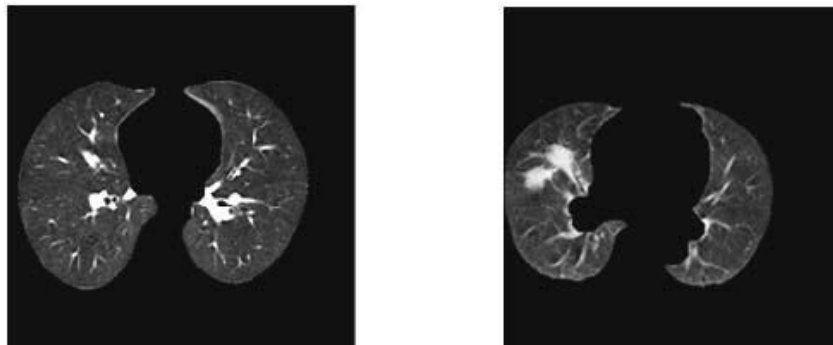


Fig. 6 Segmented Lungs

The opening operation using the periodicline structuring element tends to remove some of the foreground pixels from the edges of the region of foreground pixels.

V. RESULTS

Lung Cancer Detection malignant cases detection is performed using convolutional neural networks. In particular, we used well known deep CNN models for classification of images which have are pre-trained from large image databases and retrained them to learn Lung Cancer Detection malignant vs negative cases.

Chest scan images which have been clinically diagnosed as lung cancer malignant are preprocessed and afterwards are used to retrain existing deep CNN models for image classification. CT images preprocessing consists of image resizing and pixel values normalization to meet the input specifications of each pretrained deep CNN model. The CNN models are retrained as binary classifiers to identify malignant Lung Cancer Detection against non-Lung Cancer Detection chest CT images.

The retrained deep CNN models and used for testing, receiving as input new chest X-rays with unknown clinical diagnosis in order to automatically label them as malignant Lung Cancer Detection cases or not, i.e. providing a binary decision per chest scan. The block diagram of the evaluated architecture for detection of Lung Cancer Detection malignant cases from chest CT images.

Input Image Selection

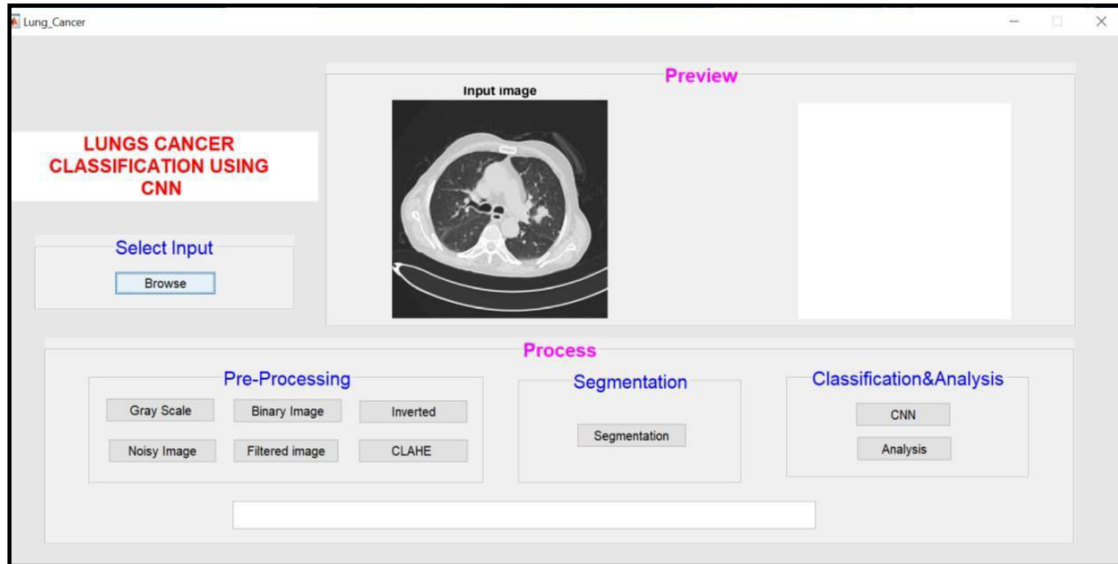


Fig 7 Input Image Selection

Preprocessing: Grey Scaled Image:

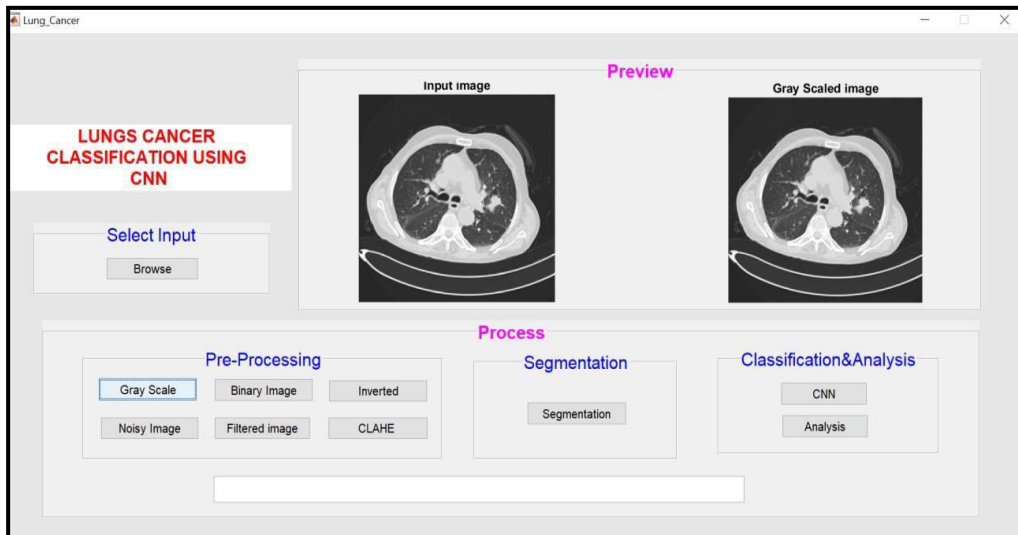


Fig 8 Grey Scaled Image

Binary Image:

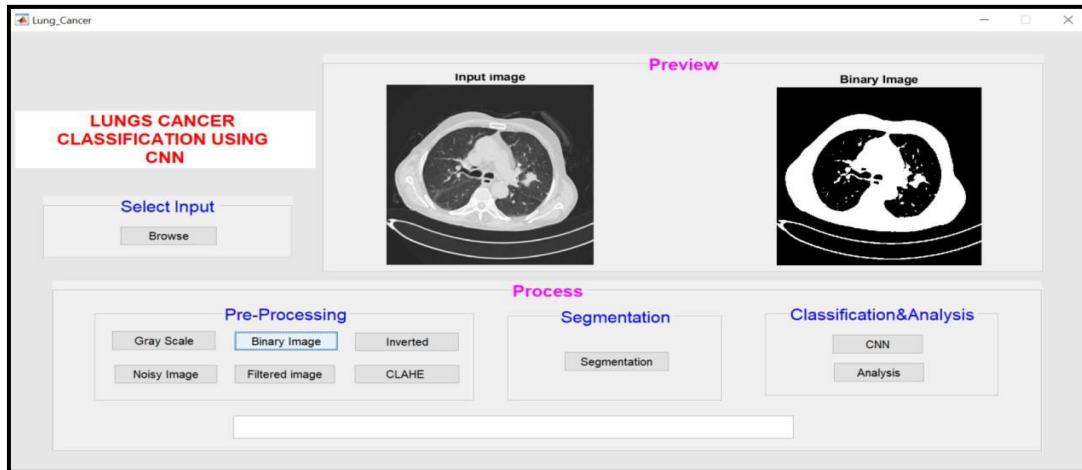


Fig 9 Binary Image



Fig 10 Inverted Image

Noisy Image:



Fig 11 Noisy Image

Filtered Image:



Fig 12 Filtered Image

Histo Equalization Image

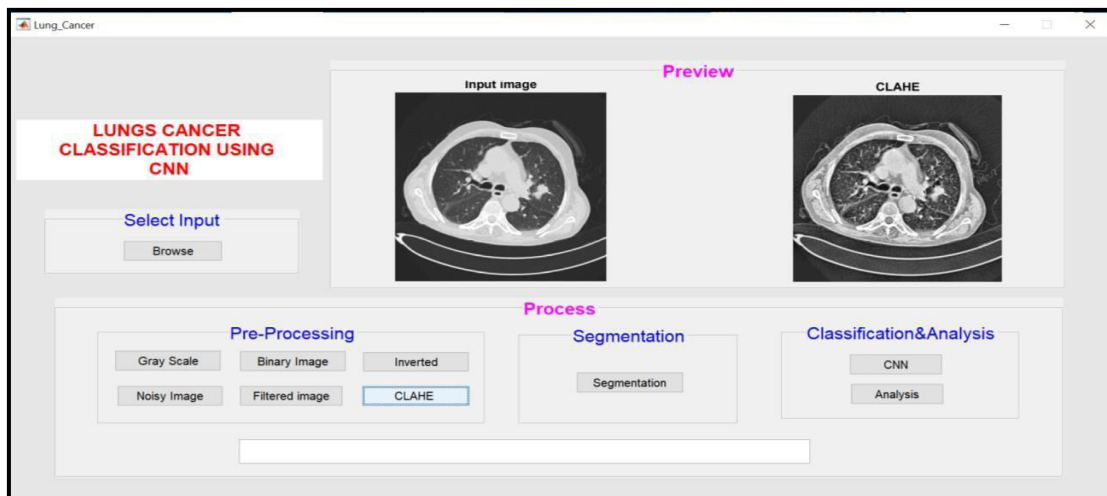


Fig 13. Histo Equalization Image

Segmentation:

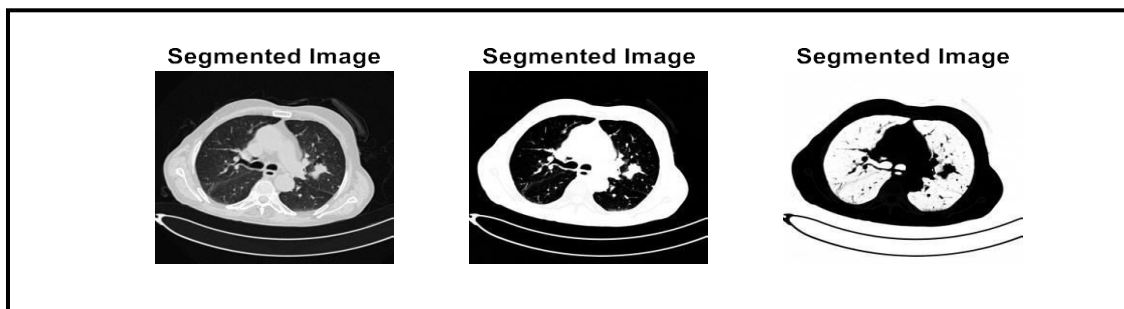


Fig 14 Segmentation

Classification of Lung Cancer Detection:

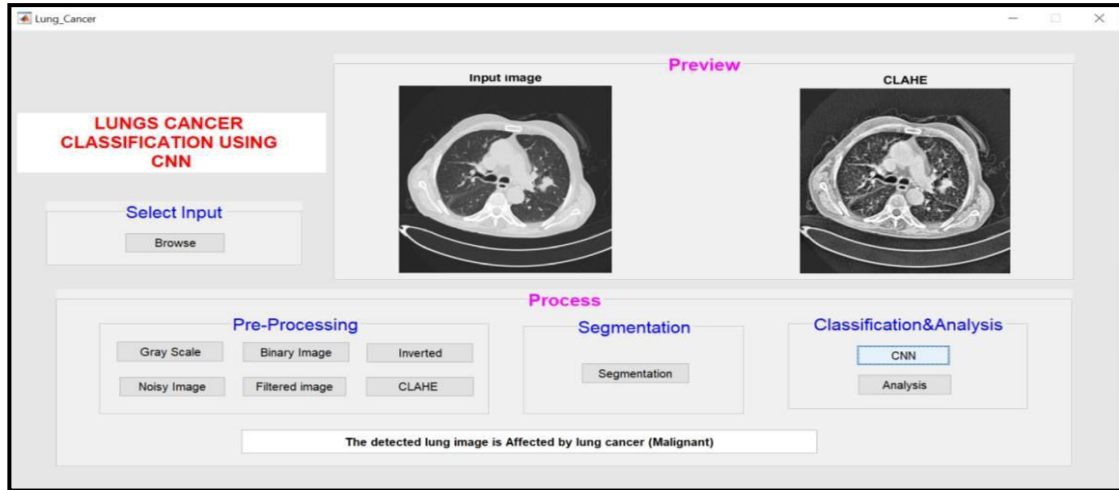


Fig 15 Classification of Lung Cancer Detection

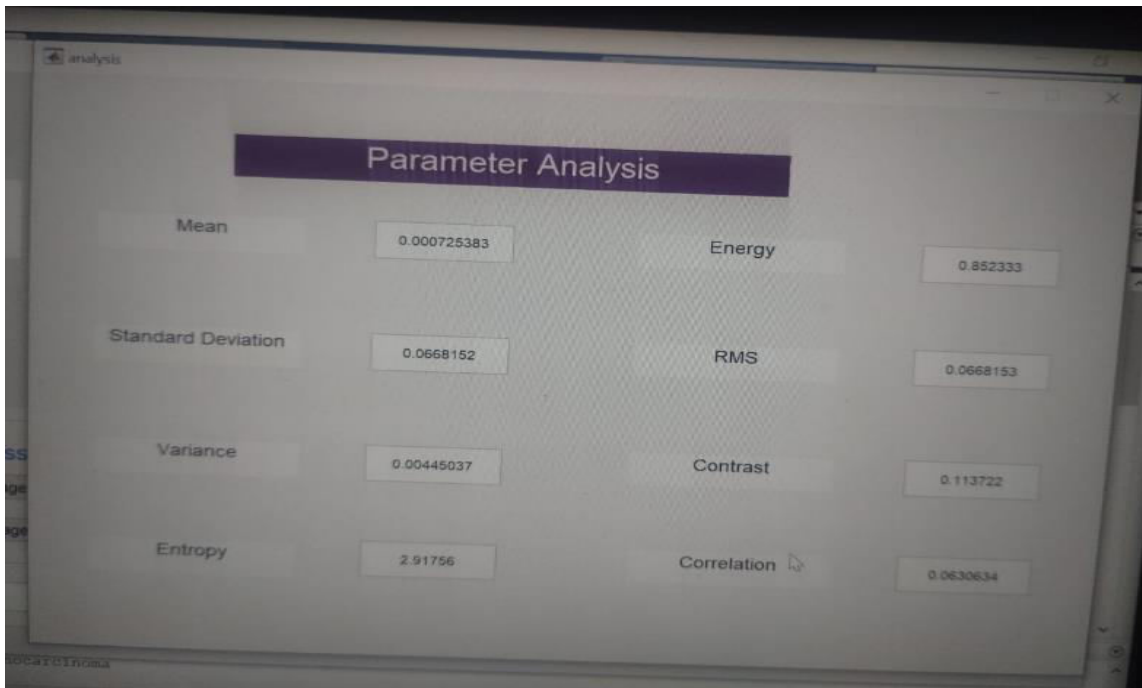


Fig. 16 Parameter analysis

Severity Status :

The performance evaluation of this approach can be estimated based on sensitivity, specificity and accuracy

Sensitivity = $TP / (TP + FN)$ Specificity = $TN / (TN + FP)$

Accuracy = $(TP + TN) / (TP + FP + TN + FN)$

The true maligents (TPs) identified by this approach was 94 and hence the sensitivity was reported as 76.42% (94/123).

The number of false maligents (FPs) was 38 and hence the specificity was reported as 78.53 % (139/177). Accuracy was estimated as 77.67% (233/300)

Final result which shown in the above fig, convolution neural network classify the image and the detected lung image is

affected by lung cancer (malignant).

Step 1: Normal : In almost all cases, benign lung tumors require no treatment, but your doctor will probably monitor your tumor for changes.

Step 2: Adenocarcinoma : Lung adenocarcinoma (a type of non-small cell lung cancer) is highly invasive. Even with early diagnosis, the chance of survival after 5 years is only 61%. If the cancer has spread to distant organs at the time of diagnosis, this survival rate drops to only 6%.

Step 3 : Malignant: A term for a disease in which abnormal cells can divide uncontrollably and invade adjacent tissues. Malignant cells can also spread to other parts of the body through the blood and lymphatic system. There are several main types of malignant tumors. Carcinoma is a malignant disease that begins in the skin or tissue that lines or covers the internal organs. Sarcomas are malignant diseases that occur in bone, cartilage, fat, muscle, blood vessels, or other connective or supporting tissue. Leukemia is a malignant disease that begins in hematopoietic tissues such as the bone marrow and produces an excess of abnormal blood cells. Lymphoma and multiple myeloma are malignant tumors that arise from cells of the immune system. Cancer of the central nervous system is a malignant tumor that begins in the tissues of the brain and spinal cord. Also called cancer.

VI. CONCLUSION

Several Computer Aided Diagnosis (CAD) techniques have been developed to diagnose early-stage lung cancer using CT scans. The CAD systems primarily focus on recognising and detecting lung nodules. There are no CAD techniques to determine the stage of lung cancer. Staging lung cancer at the time of diagnosis is critical since therapy is determined by the stage of the cancer. The main disadvantages of present CAD systems are their accuracy in segmenting nodules and staging lung cancer. This study focused on establishing CAD techniques for segmenting nodules and identifying different stages of lung cancer, which would help radiologists analyse the illness. The staging of cancer throughout its examination is the most important predictor of survival and decides therapy.

REFERENCES

- [1] Priyanka Kamra, Rashmi Vishraj, Kanica and Savita Gupta, "Performance Comparison of Image Segmentation Techniques for Lung Nodule Detection in CT Images", IEEE International Conference on Signal Processing, Computing and Control, 24-26, Sept 2015.
- [2] Negar Mirderikvand, Marjan Naderan and Amir Jamshidnezhad, "Accurate Automatic Localisation of Lung Nodules using Graph Cut and Snake Algorithms", IEEE 6th International Conference on Computer and Knowledge Engineering, Oct 2016.
- [3] Qiu Shi, Wen Desheng, Cui Ying, and Feng Jun, "Lung Nodules Detection in CT Images Using Gestalt-Based algorithm", Chinese Journal of Electronics, Vol. 25, No. 4, July 2016.
- [4] Sara Soltaninejad, Irene Cheng and Anup Basu, "Robust Lung Segmentation Combining Adaptive Concave Hulls with Active Contours", IEEE International Conference on Systems, Man and Cybernetics, Oct 2016.
- [5] Lung Imaging Database Consortium www.wiki.cancerimagingarchive.net
- [6] Sayali Satish Kankar, N.D. Thomare and S.S. Lokhande, "Detection of Lung Cancer Using Marker-Controlled Watershed Transform" IEEE International Conference on Pervasive Computing, 8-10 Jan 2015.
- [7] Tong Jia, Hao Zhang and Haixiu Meng, "A novel lung nodules detection scheme based on vessel segmentation on CT images", Bio-Medical Materials and Engineering 24 (2014) 3179 3186.
- [8] Nisar Ahmed Memon, Anwar Majid Mirza, and S.A.M.Gilani, "Segmentation of Lungs from CT Scan Images for Early Diagnosis of Lung Cancer", World Academy of Science, Engineering, and Technology 20, 2006
- [9] R. Gonzalez and R. Woods, Digital Imaging Processing (NJ, Prentice Hall, 2002).
- [10] Ivanov, A., Zhilenkov, A., The prospects of use of deep learning neural networks in problems of dynamic images recognition, Russian Young Researchers in Electrical and Electronic Engineering (EIconRus), Russia, 2018
- [11] Huang, T., Gao, F., Wang, J., Combining Deep Convolutional Neural Network and SVM to SAR Image Target Recognition, International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), United Kingdom, 2017
- [12] Stanitsas, P., Cherian, A., Truskinovsky, A., Active convolutional neural networks for cancerous tissue recognition, International Conference on Image Processing (ICIP), 2017, China.

- [13] Cruz-Roa AA., Arevalo Ovalle JE., Madabhushi A., González Osorio FA., A deep learning architecture for image representation, visual interpretability and automated basal- cell carcinoma cancer detection, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, Japan
- [14] Sarraf, S., Tofinghi, G., Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data, Future Technologies Conference, USA, 2016
- [15] Knight, Sean Blandin, et al. "Progress and Prospects of Early Detection in Lung Cancer." Open Biology, vol. 7, no. 9, 2017, p. 170070., doi:10.1098/rsob.170070.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details