Early Detection of lung Cancer Using Machine Learning Technique

B.Devananda Rao[1]
Department of Computer Science and Engineering
MLR Institute of Technology, Hyderabad
dev.bolleddu@gmail.com

Dr.Mahammad Arshad[2]
Department of Information Technology
Deccan College of Engineering & Technology, Hyderabad
arshad@deccancollege.ac.in

2023 International Conference on Computer Communication and Informatics (ICCCI) | 979-8-3503-4821-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCCI56745.2023.10128389

Abstract— The study and diagnosis of the lung diseases is the investigating point of interest for the medical experts existing from the aforetime to the current day. A diagnostic system is required to overcome this approach to reduce the time in diagnosis of lung cancers that is putting human life at risk. Few projects are proposed for this purpose and many other are under trails. To identify the cancer cells in the CT images a neural network model is proposed. It is also used to recognize issues in therapeutic imaging applications. A framework is originated for the identification of lung cancer that includes application of deep neural network system and AI. For staging of lung cancer classification of CNN is required. Deep learning methods are used on supervised learning to obtain better results. The proposed framework includes neural network that comprises of many steps like collection of images, pre-preparation, pixel enhancement, segmentation of images, and feature extraction. To outline the present research, it explains that to attain accurate results in detection and treatment of lung cancer cases at a low cost machine learning plays an unpredictable role.

Keywords- *lung cancer, CT images, convolutional neural network (CNN), framework, supervised learning.*

I. INTRODUCTION

Due to the widespread use of smoking and ccontamination of air in the world lung cancer cases are increasing day by day and is the leading cause of death worldwide. Lung Cancer constitutes about 41% of all the cancer in India. When the cells grow uncontrollably and abnormally in number and size they develop into cancer.

Diagnosis of malignant lung cancer in its initial stages is very important for the treatment and also to increase the survival period. Diagnosing lung cancer is time-consuming and costly as it requires human efforts make decisions. With the application of machine learning methods in image detection and deep learning programs the detection of lung cancer is made easy and accurate.

In the present study to improve the accuracy of the identification of tumor we use profound neural network system. In this study non-small cell lung cancer Radio genomic datasets of lung malignancy CT images are dealt with learning methods [7, 8]. Various steps like pre-processing, image extraction, binarization, segmentation and thresholding are included into a framework. In framework inputs are taken from the CT chest images and deep learning methods are applied to predict malignancy at early stages. To improve the exactness of the predicted results the present study applied Convolutional neural network system with DICOM dataset.

A. AIMS AND OBJECTIVES

1. The aim of the study is to provide a framework to confirm the prediction of lung cancer made by the specialists.
2. Deep learning techniques are utilized to detect growth of tumor and staging of the tumor in a lesser time, and cost with few efforts.
3. Image processing and AI are the automated programs that lessen human efforts in detection of the lung cancer in the pictures.

B. PROBLEM STATEMENT

• The lesion of lung cancer is detected based on the data obtained from images and not with symptoms.
• applying deep learning techniques the probability of the lung cancer disease and its stages are predicted with the input from CT images.
• Various programs are used to increase the accuracy of the detection better than in previous research.
• To identify the lung cancer at an early stage by the specialists they are provided exceptional tools.

II. REVIEW OF LITERATURE

According to YutongXie, [1] a multi-view knowledge-based collective (MV-KBC) deep learning model is applied to differentiate benign lung nodules from the malignant nodes with the information of chest CT images. To train the model they utilized 9 KBC sub models [1]. This model is tested on the LIDC-IDRI data set and compared with the five modern classification approaches.

Qing Wu and Wenbing Zhao [2] showed that early stage of small cell lung cancer can be distinguished from

979-8-3503-4821-7/23/$31.00 ©2023 IEEE

the metastatic malignancy with vectored histogram of EDM AI calculation.

As per Lilik Anifah et al. [3] Gray-Level Co-occurrence Matrices (GLCM) is a feature of Artificial neural network back propagation used for the diagnosis of lung cancer. The database of the cancer imaging archive contains of 50 CT images we obtain the information of lung. The Neural Network Backpropagation method has three layers is a process that consists of the steps like pre-processing of image, extraction of features, segmentation of images and recognition of growth of tumor. The normal lung can be distinguished from the malignant lung with 80% of accuracy by applying frame-work [3].

A combination process of image processing, SVM, and watershed segmentation is applied in the detection of the lung cancer by Prof. Anuradha Deshpande and Dhanesh Lokhande[4]. This procedure helps in attaining required data from various interpretative properties of the images in a Fused Image format. Denser tissues are examined by the CT films whereas delicate tissue is filtered by MRI. By melding the pertinent data of the two images, we attain the accurate data. The quality of this merged image is obtained by the combination procedure.

As per Abbas Khosravi and Amin Khatami [6] deep learning mechanism and auto encoders help in sorting datasets with high dimensions that made the classification of datasets difficult [6].

III. PROPOSED SYSTEM APPROACH

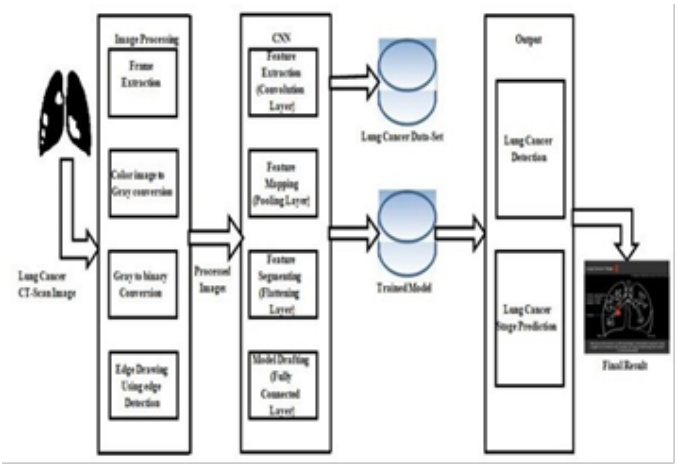


Fig. 1. System Architecture

A. Phases of the present model:

1) *Image Processing: -*
Image processing is a program used to withdraw valuable data from the images and magnify it. An image composed of hues over RGB colors. The input is usually an image. The output obtained on applying image processing on input may be an image or particular characteristic of the image.

2) *Image Filtering: -*
Image filtering is a process that removes certain features like noise, improves edges, encompasses smoothing, honing, and improves the images. The pixel value of the output image is detected by applying algorithms to the input images by filtering image frequency and its spatial domain.

3) *Feature Extraction: -*
Feature extraction is an important step in processing because it transforms pictorial data to alphanumerical data. It extracts the features of importance in an image and presents them for future processing.

In our framework, CNN is incorporated.

4) *Segmentation: -*
To detect the objects and image boundaries like curves, lines a digital image is dissociated into multiple segments depending on the set of pixels. A common property is noticed among the pixels present in a segment. This is known as segmentation of images. The common method used is thresholding.

5) *Edge Detection: -*
The spots in the digital picture where brightness of images has discontinuities and changes distinctly Edge Detection is applied in image processing and computer vision. Edge detection is most familiar program that sharply detects the discontinuities in intensity values of images. The image information is depended on its edges that bounds the two regions. To extract this information, we applied canny edge detector in the present research.

6) *Feature Recognition: -*
In computer aided design a region of a part with few engaging properties of topography is a feature. The main objective of the feature recognition is produce higher-level features of data manufacturing from the entities of lower level like surfaces, curves, and edges of the images. Feature recognition technology has the ability to identify various features and group geometric structures like faces from the solid models into functional features like holes, ribs and slots [7]. CNN is learnt through knowing and it identifies the relation and connection within information.

IV. MATHEMATICAL MODEL

System Description: S= {I, F, O} S indicates a system for detection and stage prediction of lung cancer.

INPUT:

F=F1,F2,F3...FN Functions to execute modules

I=C1, C2, C3... input of systems lung cancer images

O=R1, R2, Rn prediction results in the form of stages I=result access by User

F:

F1=Image processing applied on Lung Cancer images

F2=feature extraction from images

F3= results of stage detection

O:

R1= model creation from training.

R2= model-based image testing

Success:

1. By working on the datasets of CT images we can increase the accuracy.
2. By applying this model, we can fasten the results.

Failures:

1. The vast database consumes more time to attain the desired information of targeted images.
2. Hardware failure.
3. Software failure.

Mathematical Model in Equation on format Notation

Where,

M= Set of all entities.

LCIT1= Lung cancer images type 1

LCIT2= Lung cancer images type 2

LCITN= Lung cancer images type N

TLCI=Total Lung cancer images

Total number of images are calculated by the Equation:

Total number of images of lung cancer = number of lung cancer images type 1+ =number of lung cancer images type 2+…..+ = number of lung cancer images type N.

$$\sum TLCI = \sum LCIT1 + \sum LCIT2 + \ldots + \sum LCITN \quad \ldots \text{Equation (1)}$$

A. ALGORITHM

1) Convolutional Neural Network (CNN)

Steps to be followed for image processing by convolution neural network: -

• Image handling utilizing Convolutional neural systems (CNN) has been used in different fields, for example, facial recognition, analyzing documents, historical and environmental collections, understanding climate, drug discovery, video analysis, advertising, etc.

• The lung cancer image presentations are highly complex that makes the prediction of lung cancer patient outcome difficult and questionable. To make it better large number of data required for the development of models. In the present study CT scan images of the lung cancer patients are considered and their features extracted by applying deep learning programs to improve the prediction of patient outcome models.
• Ultimately, various data formats, like genomic features and imaging features are processed easily by machine learning and deep learning techniques.
• Various procedures in the processing of images with CNN are:

1. Resizing the image, noise removal, image reading and removal of some morphological operations are included in image pre-processing. The CNN pre-dealing with pictures is done by removing unnecessary distortions, enhancing and improving the images, and redoing or restoring the image.

2. Feature extraction segregates facial features, corners, edges and joints to condense the images. It combines new set of features by choosing the more appropriate one and puts an end to the different entities.

3. Image division into locales done by Segmentation.

4. To identify and discover any feature or object in images Recognition is used.

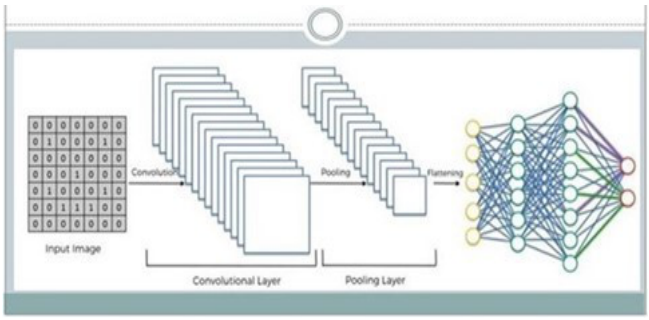


Fig. 2. CNN- schematic diagram

Four different layers of CNN are-

- Convolutional Layer
- Pooling
- Flattening
- Fully Connected layer

1. Convolutional Layer

Convolutional layer is used to sharpen and enhance the images. The set of filters or kernels present in the convolutional layer extract various features based on pixel

size. Using different features in various operations it performs convolutions on input resulting in definite feature map. To get the final output or feature map in the convolutional layer all these attained images are assembled in one layer.

2. Pooling

To reduce the training time of the networks and also reduce the problem of overfitting pooling is used. Max Pooling is used to reduce the number of pixels in the image output of the extracted layer.

3. Flattening

To convert the pooled features of 2D data into a single continuous column or vector flattening is used. The flattening is utilized as an input to the next connected layer.

4. Fully Connected Layer

Alike multi-layer perception (MLP) all neurons in one layer is associated with all neurons of the previous layer in fully connected layers. Fully connected layer adds weight matrix to the previously analyzed feature to anticipate the current layer resulting in final classification.

B. Convolutional Neural Network Approach (CNN)

Step 1- Image input
Step 2- Processing of image by open-cv.
Step 3- Extraction of image features
Step 4- Generation of Machine model
Step 5- Classification of lung cancer
Step 6- Detection of lung cancer stage**.**

Tensor Flow

Tensor flow is a library with open source utilized for numerical computation drawn by graphs of data flow. For image recognition, prognosis of the disease with the image prediction, and classification deep neural networks are used that are trained and run by Tensor Flow [7]. In this present program Tensor Flow set-up is employed to execute CNN.

Python

To achieve various tasks in machine learning Python provides multiple routs with its predetermined set of libraries. In our program python is utilized because of it is packed with numerous functions. Python is adaptable, high-level programming language, and is object oriented.

Open-CV

Image related components are dealt with Open-cv library that performs image processing. Using this library, the reading of images and extraction of pixels from the images of the lung cancer patients are done. In the present project reading, transforming images and displaying images are managed in Open-CV by applying NumPy library, which is a python library.

Image Processing

Detection of edges, image smoothening, image banarization, Gaussian filtering included under image processing.

V. OBSERVATION AND RESULTS

Table 1. staging of lung cancer

Primary Tumor (T)	**Size of tumor**
T1	< 3cm in diameter; T 1a <= 2cm; T1b 2cm to 3cm
T2	3cm to 5cm; T 2a 3cm to 4cm; T 2b 4cm to 5cm
T3	>5cm<=7cm
T4	Size greater than 7cm

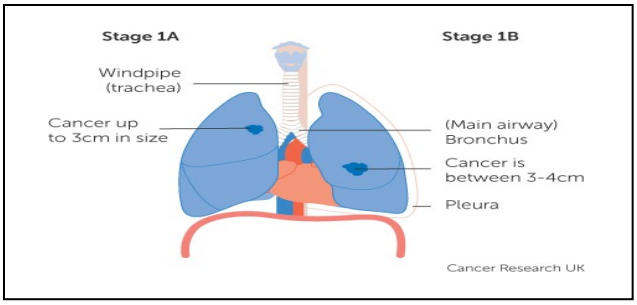


Fig. 3. Stages of Lung Cancer nodule

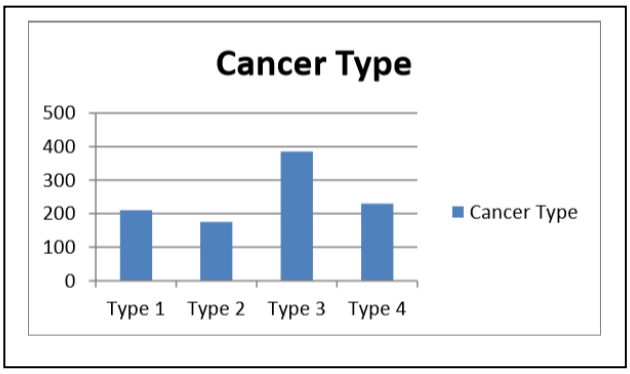
In the study experimental setup over 1000 images of lung cancer done is shown in Table no.2. these images were again classified into type1, type2, type3, and type4.
Out of 1000 images in the study 210 are in type1 dataset, 175 are under type2set, 385 are in type3 set and 230 images are included under type4 dataset of lung cancer images.

Table 2. Testing dataset of lung cancer images

Sr No.	**Types**	**Number of images**
1	Type 1 Dataset images	**210**
2	Type 2 dataset images	**175**

3	Type 3 dataset images	**385**
4	Type 4 dataset images	**230**

The graphical representation of the above data is as follows. Number of images of type 1 cancer is 210, type 2 lung cancer is 175, type 3 lung cancer is 385, and type 4 lung cancer is 230.



Graph 1. Number of images of lung cancer type

VI. CONCLUSION

In the present study deep learning mechanisms, image processing methods, and extraction of features done to attain better appropriate results in the diagnosis of disease.

The study focused on the approaches that increased the detection of lung cancers and its stages like neural network system. To differentiate and group the similarities under one roof in the diagnosis of the lung cancer disease artificial intelligence has its elemental role.

VII. FUTURE WORK

To make the prognosis accurate, the practitioners are using various imaging techniques like X-rays, computed tomography, magnetic resonance imaging, positron tomography. Lung cancer is a disease that effects socially, economically and, emotionally. This can be controlled and further damage can be arrested by early diagnosis. Development of vaccines, drugs help in controlling measures of cancer.

REFERENCES

[1]. YutongXie, "Knowledge-based Collaborative Deep Learning for Benign Malignant Lung Nodule Classification on Chest CT," 2018, IEEE.

[2]. Vijayakumar, T. "Classification of Brain Cancer Type using Machine Learning." *Journal of Artificial Intelligence* 1, no. 02 (2019): 105-113.

[3]. LilikAnifah, Haryanto, RinaHarimurti, "Cancer lung detection on CT Scan image using ANN backpropagation based gray level co-occurrence matrix feature." 978-1-5386-3172-0/17/ 2017 IEEE.

[4]. Prof. Anuradha Deshpande, Dhanesh Lokhande, "Lung cancer detection with a fusion of CT and MRI image using image processing." (IJARCET) Volume 4 Issue 3, March 2015.

[5]. RachidSammouda, "Segmentation and analysis of CT chest images for early lung cancer detection." Global Summit on Computer & Information Technology 978-1-5090-2659-3/17 2017 IEEE.

[6]. Abbas Khosravi, Amin Khatami, "Lung cancer classification using deep learned features on low population dataset." Canadian Conference on the Electrical and Computer Engineering (CCECE) 978 -1-5090-5538- 8/17 2017 IEEE.

[7]. Vijeeta Patil , Manohar Madgi and Ajmeera Kiran ,"Early prediction of Alzheimer's disease using convolutional neural network: a review,", The Egyptian Journal of Neurology, Psychiatry and Neurosurgery, (2022) 58:130, https://doi.org/10.1186/s41983-022-00571-w

[8]. P. Chinnasamy, S. Arun Kumar, V. Navya, K. Lakshmi Priya, Siva Sruthi Boddu,Machine learning based cardiovascular disease prediction, Materials Today: Proceedings, Volume 64, Part 1, 2022, Pages 459-463,ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2022.04.907