IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005324 |

Lung Cancer Detection and Classification

N.Ashok Kumar[1], Seelam Mounica[2], Thota Thriveni[3], Vineela Sheba Kondru[4]

Asst. Professor, Department of IT, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India[1]

B. Tech Students, Department of IT, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India[2,3,4]

ABSTRACT: In the field of medicine, identification and treatment of cancer is considered as one of the biggest challenge in treatment of chronic illness. The survival of patients depends on timely detection and cure. Experts use the CT scan or Computed Tomography Scan images of patients to detect and classify nodules, before proceeding with advanced treatment procedures. The present day advances in artificial intelligence, machine learning based on deep learning model can be used to develop sophisticated Computer Aided Diagnosis systems to detect cancerous nodules. The proposed system is based on Convolutional Neural Networks to categorize nodules detected in CT scan images as malignant or benign. Image processing and Neural Networks have been extensively used in the detection and classification of cancerous nodules. Hence CNNs are more appropriate, for the task of nodule detection and classification. CNN's have more properties like multiple feature extraction. When convolution layer, subsampling or pooling layer, fully connected layer such layers are combined, leading to Deep CNNs, it helps in increasing the accuracy of classification. The proposed CNN model will be suitable for the early detection and classification of CT scans images containing nodules with accuracy of 93.52%, using the domain knowledge of the CT scan images of lung in the field of medicine and Neural Network.

KEYWORDS: Computed tomography, Convolutional Neural Network (CNN), Computer-aided diagnosis.

I. INTRODUCTION

Now-a- day's lung cancer has become one of the leading causes of cancer related deaths of humans. So it is necessary for the trained radiologists to identify the cancer accurately in the lung in the early stage as possible to reduce the deaths of humans [2] .So this is very complicated job to detect the lung nodules which is affected by cancer and non-cancer in its early stages.

In the past days, radiologist were manually analysing the CT scans images of lung, looking for the potential nodules and identify the cancerous and non- cancerous in those nodules. This process needs a high knowledge about the lung nodules which is very tedious and time consuming. So this can resolved by introducing the computer aided diagnosis (CAD) system to detect the lung nodules and classification of nodules as either cancerous or non- cancerous. This system will be as second opinion to the radiologist to detect and analyse the nodules of the lung.

In US, lung cancer has become the cause for the cancer related deaths. Approximately 229,447 new cases of lung cancer were there and in that 159,124 related deaths. So early diagnosis of lung cancer can improve the effectiveness of treatment and increase the chance of the survival of patient's. The lung cancer can be identified by the non-invasive imaging modalities those are computed tomography(CT), contrast-enhanced computed tomography(CE-CT), low-dose computed tomography(LDCT) and positron emission tomography (PET)[3].

In the past years, CAD system has been developed for both the nodule segmentation and also classification of lung nodules as cancerous or non-cancerous. But this system for segmentation will generate a very good detection of lung nodules but however in the process they lead to many false positive, when actual positive compared to this false positive segmentation the ratio can be in hundreds[3]. So this system can lead many false positive, when dealing with lung cancer. It's better to label anything that looks like nodule in the image, and the radiologists need to fallow the manual methods of detecting the lung cancer.

Traditional algorithms of image processing used to detect unique features of images. So this requires the hand crafted features has to be created which learns the features at manual process. So in this method it is very complex to differentiate the features of cancerous and non-cancerous nodules.so the deep learning can avoid all these problems and are capable of tackling with the problems like image recognition, video recognition, speech recognition and natural language processing etc [7]. Manual feature extraction requires expert knowledge of the lung cancer to the designer.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005324 |

Deep learning will have the properties to learn all the features in the images.

In particular, Convolutional Neural Network (CNN) extracts the features of input images by using one or more layers of convolution and subsampling or max-pooling layer that are in the hierarchical manner. The general CNN consists of three layers – Convolution, Max-polling and Fully connected. The classification of images is done by CNN by extracting features in each layer and producing a final model. When other test image is given to model compares the features of both and classify the images to different classes by the accuracy. In this paper, we have used the CNN for the classification of CT scan images of lung cancer as cancerous or non-cancerous.

II. EXISTING SYSTEM

The existing CAD system used for early detection of lung cancer with the help of CT images has been unsatisfactory because of its low sensitivity and high False Positive Rates (FPR). *Positron emission tomography* (PET), *computed tomography* (CT), *low-dose computed tomography* (LDCT), and *contrast-enhanced computed tomography* (CE-CT) are the most common noninvasive imaging modalities for detecting and diagnosing lung nodules. PET scans are used to discriminate between malignant and benign lung nodules. Early detection of the nodules can be based on CT and LDCT scans that allow for reconstructing the anatomy of and detecting the anatomic changes in the chest. The CE-CT allows for reconstructing the anatomy of the chest and assessing the detected nodule's characteristics.

The limitations of existing system are:

- No preprocessing feat smoothing has been applied.
- There has been no proper classification current best solution, and proposed new method.
- It has not enforced its classification as normal or irregular.

III. PROPOSED SYSTEM

Convolutional neural networks are designed for minimizing the number of parameters and adjusting the architecture of the network for image classification. Convolutional neural networks are made up of a series of layers organized according to their features and functionality. A ConvNet 's architecture is close to that of the human brain connectivity pattern of neurons inspired by the Visual Cortex structure. Data augmentation the process by which data quantity and complexity increase. We will collect fresh data rather than converting the data already available. Data augmentation is an important phase in deep learning, since we require vast quantities of data in deep learning and in certain instances it's not really possible to capture thousands or even millions of images, so this data augmentation comes to the scene. It allows us to maximize the dataset size and add uncertainty within the datasets.

The benefits of the proposed system are:

- The results obtained from training and testing part are fed into in CNN layers where the images are classified and the output is obtained.
- The algorithm used for classification is CNN with Adaboost where accuracy calculation is of the images is done based on the sample weights of images.

III. MODULES

Module-1 : Lung cancer detection and nodule classification using CNN
The dataset of the Lung Image Database Consortium image collection (LIDC-IDRI) is an international image resource for evaluating and identifying lung cancer. It consists of CT images in DICOM format. The size of the original images are 512 x 512 but it is difficult to train large size images in DCNN so preprocessed the images to reduce size suitable for the network.

Module-2 : Lung cancer classification using CNN
The input images are applied to the Deep CNN model for training the model using 90% training images. After training, the model is evaluated by 10 % of testing image dataset collected from the same dataset. Here the images samples are

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005324 |

used to feed the network model which classifies into cancerous or Non Cancerous Images.

IV. METHODOLOGY

A. Nodule Extraction and Pre-Processing

The dataset include the information of the nodules in the candidate.csv file which consists of serial nodules of lung image, co-ordinates of the nodules present in the lung image. The LUNA16 dataset • Load LUNA16 candidates.csv • Load lines with class = 1 (Positives) • Load lines with class = 0(Negatives) • Load x,y,z co-ordinates and label • Split training, validation and testing dataset

• For record in dataset,

a. Extract nodule based on x,y,z co ordinate

b. Normalize the extracted image to filter unwanted voxels.

c. Save nodule in jpeg format

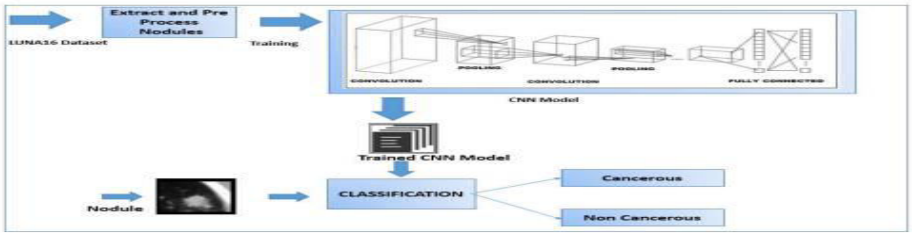

Fig 1: System Architecture

B. Training and Testing in Convolutional Neural Network

The nodules are extracted from the annotations given in the dataset by the radiologists and normalize those nodules and save in the jpeg format. These images are given as input to CNN for classification of image as cancerous and noncancerous. The pre-processing of images is not explicitly done in CNN because CNN classifier implicitly does all preprocessing of images.

i. Training/Testing/Validation dataset creation :

• Augment the training images by applying rotation to increase dataset variations and improve overall module accuracy.

• Package training, validation, test images into HDF5 files

ii. Training/Validation :

• Load training and validation dataset from h5files

• Define Tensorflow CNN model

• Define and initialize Tensorflow deep neural network model

• Invoke model.fit with training and validation dataset for specific number of epochs

• Save the trained model.

iii. Testing :

• Select a Patient ID

• Retrieve the nodules for selected Patient ID

• Load the tensor flow trained CNN model

• Invoke model.predict to classify the test nodules

• Invoke model.eval to determine the accuracy

C. Concepts of Convolutional Neural Network

A typical Convolutional Neural Network (CNN) is the type of neural network used as classifier to make image classification. All the set of dataset images are loaded on to CNN where each layer reads the features of images and forward to next layers. In the general CNN model of image recognition, a hand developed feature extractor collect the data or information from the input image and eliminate the irrelevant variables. This extractor is followed by classifier which is to train the feature vectors into classes.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005324 |

V. RESULTS & DISCUSSION

The input given will be a CT scan image of the patient

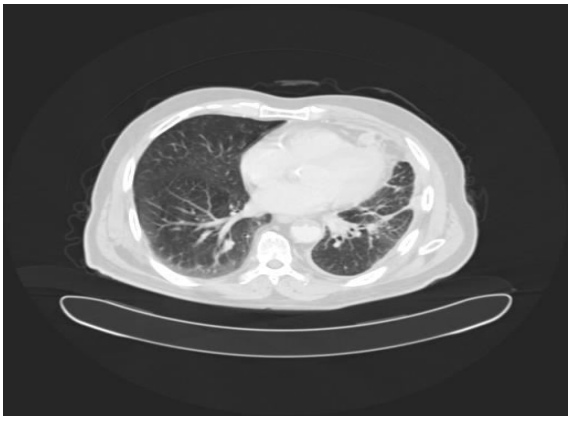

Fig 2: Input CT scan image

The outputs contains the detection of cancer tumor cells in the lungs and also classifying the cells either into benign or malignant.

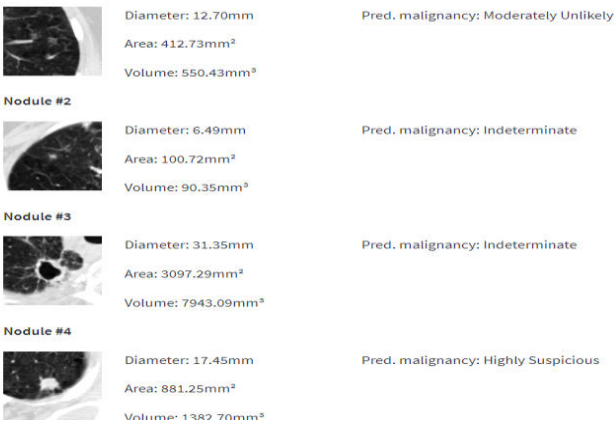

Fig 3: Lung cancer detection

Fig 4: Lung cancer classification

VI. CONCLUSION

In our research work, we have used deep Convolutional neural networks for classifying the ct images of lung nodules into cancerous (malignant) and non cancerous (benign). Thus preprocessing has been done before applying input ct images to network model to make equal sizes and format of the images. The dataset used in our research work belongs to LIDC dataset. Hence we achieved an accuracy of 100% which is the better results comparable to previous research

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005324 |

papers as mentioned. As a future work, the experiments could be performed by using Deep CNN architecture for other types of cancer.

VI. FUTURE SCOPE

The future enhancement of the proposed work can be done as below:

- The system could be enhanced to classify the CT scans by increasing layers in network to increase the accuracy.
- The system can be improved by including different types of dataset(UCI,TCIA etc.)
- The system can be improved by incorporating 3D lung cancer images for classification.

REFERENCES

1. Jiang, H., Qian, W., Gao, M., Li, Y. " An automatic detection system of lung nodule based on multigroup patch-based deep learning network" IEEE Journal of Biomedical and Health Informatics,22(4):1227-1237.
2. Disha Sharma, Gagandeep Jindal, "Identifying Lung Cancer Using Image Processing Techniques", International Conference on Computational Techniques and Artificial Intelligence, pp. 116-120, 2011.
3. Farzad Vasheghani Farahani "Lung Nodule diagnosis from CT images Based on Ensemble Learning." IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology , 2015.
4. Atsushi Teramoto at el. "Automated classification of benign and malignant cells from lung cytologicalimages using deep convolutional neural network" Informatics in medicine unlocked 16,2019,100205 elsvier.
5. Xin-Yu Jin, Yu-Chen Zhang, Qi-Liang Jin "Pulmonary Nodule Detection Based on CT Images Using Convolution Neural Network." 9Th International Symposium on Computational Intelligence and Design. 2016.
6. Ryota Shimizu, Shusuke Yanagawa, Yasutaka Monde, Hiroki Yamagishi,Mototsugu Hamada, Toru Shimizu, and Tadahiro Kuroda "Deep Learning Application Trial to Lung Cancer Diagnosis for Medical Sensor Systems" International Symposium on Computers and Communications, 2016.
7. Po-Whei Huang, Phen-Lan Lin, Cheng-Hsiung Lee, C. H. Kuo, "A Classification System of Lung Nodules in CT Images Based on Fractional Brownian Motion Model", IEEE International Conference on System Science and Engineering, July 2016.
8. Vaishali C. Patil, Shrinivas R. Dhotre, "Lung Cancer Detection from Images of Computer Tomography Scan", International Journal of Advanced Research in Computer and Communication Engineering , Vol. 5, Issue 7, July 2016.
9. Ailton Felix, Marcelo Oliveira, Aydano Machado, Jose Raniery, "Using 3D Texture and Margin Sharpness Features on Classification of Small Pulmonary Nodules" ,29th SIBGRAPI Conference on Graphics, Patterns and Images, 2016.
10. Rotem Golan, Christian Jacob, Jorg Denzinger, "Lung nodule detection in CT images using deep convolutional neural networks" International Joint Conference on Neural Networks , 2017.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details