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.2022.1005234 |

Lung Cancer Detection System using Convolutional Neural Network

Aditya Kshirsagar, Akash Vyavahare, Rohit Zarkar, Aniket Pawar, Prof. Smita Gumaste

Dept. of Computer Engineering, JSCOE, Savitribai Phule Pune University, Pune, Maharashtra, India

ABSTRACT: Lung cancer is one of the most dangerous cancers. Thousands of people are infected with it, and if they do not detect it in the early stages of the disease, the patient's chances of survival are slim to none. For the reasons stated above, and to aid in the fight against this dreadful disease, early detection via artificial intelligence processes is necessary. Using a convolutional neural network technique with proposed architecture, a computer-aided Lung Cancer Detection system (LCDS) system is proposed for detecting lung cancer in human lungs by using computed tomography (CT scan) images for assisting patient with the diagnosis of the lung cancer in three stages normal, beginning or malignant. The proposed have accuracy about 89.45 percent.

KEYWORDS: Dataset, CNN (Convolutional Neural Network), ReLU, Pooling, Fully-Connected Layer.

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 the uncontrolled growth of the cells, thus leading to the formation 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 cause of such high death rate is the detection in later stages, thus leading to delayed treatment. If lung cancer is detected at an earlier stage, chances of survival can increase up to 50-70%. 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. Lung cancer is the major cause of cancer death in the world. The symptoms of lung cancer come into light at the final stage. So, it is very tough to identify in its beginning stage. For this reason, the death percentage is very high for lung cancer in comparison with all other types of cancer. 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. According to a statistic conducted by world health organization that every year more than 7.6 million people died of lung cancer. Moreover, the death rates of lung cancer are expected upon to keep rising, to wind up around 17 million worldwide in 2030[2]. We found that lung cancers deaths in Bangladesh reached 9,660 or 1.33% of total deaths, according to the latest WHO data published. 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. Although, CT scan imaging is best imaging technique in medical field, it is difficult for doctors to interpret and identify the cancer from CT scan images. Therefore, computer aided diagnosis can be helpful for doctors to identify the cancerous cells accurately. 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. RELATED WORK

In the last few years Convolutional neural networks have become dominant approach in medical image analysis that outperforms algorithms based on handcrafted features. Esteva et al. [6] proposed the CNN-based system for skin lesion classification that achieves performance comparable with dermatologists Assessment. Regarding

| 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.2022.1005234 |

lungcancer diagnosis, methods proposed so far have dealt mostly with radiology. In [7] image-based radionics features strongly related to survival are extracted from positron emission tomography-computed tomography (PET/CT) scans. In [8] CNN is employed for classification of lung nodule images yielding accuracy of 86.4%. In digital pathology tasks CNNs have been used on cell level for mitosis detection [9] and cell nuclei detection [10]. CAMELYON16 was the first challenge dealing with WSIs to detect breast cancer metastases in lymph nodes. Thanks to availability of large annotated training set in this challenge, it was possible to train deeper and more powerful CNN architectures like Google Net, VGG-Net and ResNet . Method that gives best result in this challenge is described in [14].

It performs patch-based classification to discriminate tumor patches from normal patches using combination of 2 GoogLeNet architectures where one of them is trained with and another without hard-negative mining. Aim of TUPAC challenge was WSI based mitosis detection in breast cancer tissue and tumor grading prediction. In best performing method [15] ROI regions are firstly extracted from WSI based on cell density. This is followed by mitosis detection using ResNet CNN architecture. Finally, each WSI is represented by feature vector including the number of mitoses and cells in each patch as well as other features derived from statistics. This feature vector is fed to SVM classifier to predict tumor proliferation.Arnaud A. A. Setio, Explained about the proposed architecture which comprises multiple streams of 2-D ConvNets, for which the outputs are combined using a dedicated fusion method to get the final classification but the morphological variation of nodules is often greater than what a single candidate detection algorithm.K. Narmada, G. Prabakaran, S. Mohan proposed a novel framework to classify both small cell and large cell lung cancer and predict its type and treatment using CNN. The paper also concentrates on the preprocessing and segmentation processes to accomplish the accuracy in prediction. The experiment results in Python - Tensor Flow with Kaggle image dataset show that compared to state of the art of classification and prediction method.

III. SYSTEM ARCHITECTURE

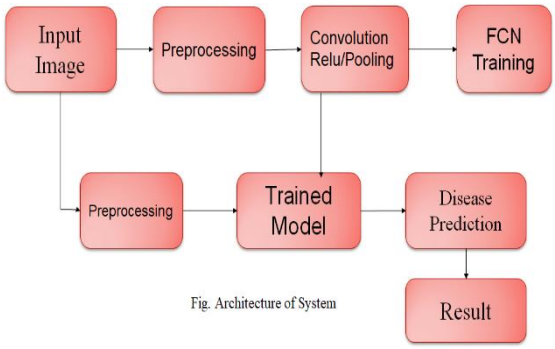


Fig. Architecture of System

1) Data collection

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

2) Pre-processing

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.

3) Segmentation

Converting the images to binary reduces computational complexity and storage issues and also is a pre-requisite for morphological segmentation of lungs. 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.

| 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.2022.1005234 |

MATHEMATICAL MODEL

$$\text{Chi-Square} : D = \sum_{i=1}^{n} \frac{(hist1_i - hist2_i)^2}{hist1_i}$$

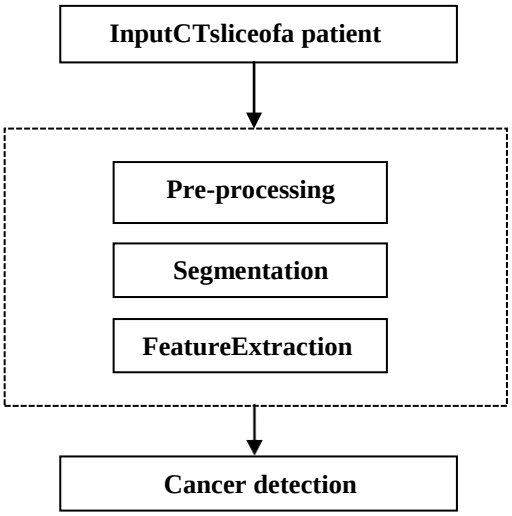
$$\text{Euclidean Distance} : D = \sqrt{\sum_{i=1}^{n} (hist1_i - hist2_i)^2}$$

$$\text{Normalized Euclidean Distance} : D = \sqrt{\sum_{i=1}^{n} \frac{(hist1_i - hist2_i)^2}{n}}$$

$$\text{Absolute Value} : D = \sum_{i=1}^{n} |hist1_i - hist2_i|$$

IV. PROJECT IMPLEMENTATION

The proposed system includes five main processing steps;theyarepre-processing,segmentation,featureextraction,tumor detection and tumor stage identification. The fig below shows the procedure to detect lung cancer using CNN.



Block diagram of the proposed lung cancer detection system

The fundamental steps within the proposed gadget are pre- processing and segmentation. The segmented image can be given for feature extraction, for this reason cancer element in the lung canbe identified. depending on the feature of the lung, the level of the lung most cancers may be diagnosed and the result may be given to the sufferers or medical doctors.The CT picture slice is pre-processed for enhancing the image. This procedure sharpens the picture for the processing easiness. The pre- processed picture is segmented using the Marker controlled Watershed segmentation. The output of the segmentation is the binary image displaying the tumor quantities. The segmented binary photograph is used for the extraction of capabilities of the tumor portion. The functions extracting for the identity of the degree of cancer are area, Perimeter, Eccentricity, Convex region and suggest intensity. based totally on the edge value and the extracted features the tumor is staged. the edge values were acquired after experimental analysis on the dataset with the help of specialists

The different steps involved in the proposed method are thefollowing:

Step 1: Read in the CT image slice of a particular patient..

Step 2: Pre-process the CT image slice.

| 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.2022.1005234 |

Step3: Segment the pre-processed image using Marker Controlled Watershed segmentation.

Step 4: Convert the segmented image into a binary image.

Step 5: Extract the features from the binary image.

Step 6: With the extracted features, identify the cancer..

Pre-processing

Pre-processing is **a** vital step for most of the image processing programs. The pre-processing method used right here is sprucing of the CT picture. **accordingly,** the finer details in the photo are greater. The image received after the pre-processing process is suitable for in addition processing. The sharpening technique makes use of a convolution approach for contrast enhancement.

$$\frac{1}{(\alpha+1)}\begin{bmatrix} -\alpha & \alpha-1 & -\alpha \\ \alpha-1 & \alpha+5 & \alpha-1 \\ -\alpha & \alpha-1 & -\alpha \end{bmatrix}$$

Where value of α can vary from 0.0 to 1.0. The default value of α is 0.2. In this work the value of α is 0.2.
Image segmentation is an important step in image analysis, object representation, visualization and many other image processing tasks. The image segmentation process partitions the image into multiple segments. It is used to locate objects and boundaries within the image. Image segmentation helps medical field for detecting the death threatening diseases such as cancer.

There are different segmentation methods available. In this work Marker Controlled Watershed segmentation is used. The main advantage of using Marker controlled watershed segmentation is it removes the over segmentation within the image. This method solves the over segmentation problem by using the markers.

The different steps involved in Marker Controlled Segmentation are the following:
Step 1: Read in the color image and convert it to gray scale image.
Step 2: Compute the Gradient Magnitude as the segmentation function.
Step 3: Mark the foreground objects within the image.
Step 4: Find out the background marker points within the image.
Step 5: Find out the watershed transform of the segmented function of the image.
Step 6: Resultant segmented binary image is obtained.

Feature Extraction

Feature extraction is an essential step for the analysis of image. This step determines the relevant information for the processing. Thus, the normality or abnormality within the lung can be identified. The extracted features are used for detection and staging of the lung cancer. The different features that are extracting here are Area, Perimeter, Eccentricity, Convex Area and Mean Intensity.

Cancer Detection

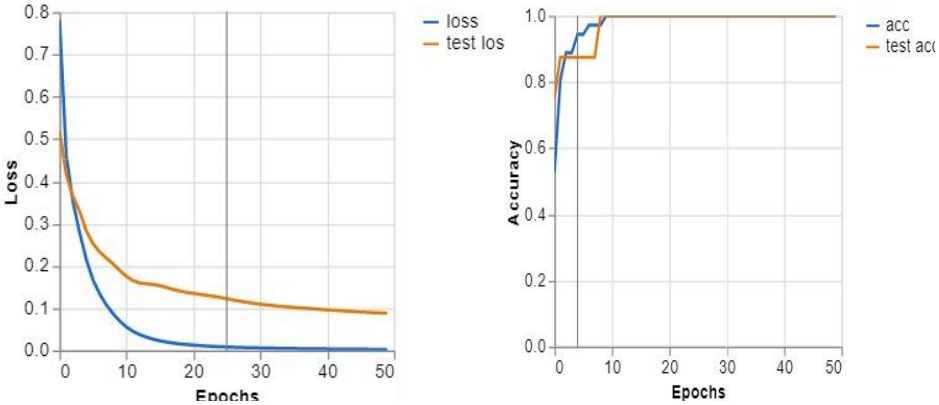
The features extracted in the previous step used for the detection of the cancer. Based on the threshold value for the feature ‘area’ obtained from experimental analysis, the probable tumorous regions within the lung are detected. From the probable tumorous regions, the individual tumor portions are selected by the experts. Depending on the features of the selected tumor portion the stage of the tumor is detected accurately by the classifier. The classifiers used for the study are Support Vector Machine (SVM), Naive Bayes Multinominal classifier (NBM), Naive Bayes Tree (NB tree) and Random tree

| 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.2022.1005234 |

V. SIMULATION RESULTS



Above shown are Accuracy and loss graphs of the projects. First figure shows epochs vs Loss graph and second graph shows epochs vs accuracy graph.
Classifier Accuracy

Classifier	Accuracy
SVM	61.11%
NBM	75%
NBTree	91.67%
RandomTree	94.4%

Among the above classifiers random tree classifier is found to be more accurate. So, for the lung cancer detection Random tree is used.

VI. CONCLUSION AND FUTURE WORK

The proposed project shows accuracy about 89.45%.The goal of this planned study is to overcome the challenges of early diagnosis of lung cancer nodules before they worsen. This work develops a computer-aided diagnosis technique for early detection of this fatal cancer for this goal. As a data source for the suggested model, chest tomography scans were used. The purpose of this work was to improve a CNN deep learning model that could successfully detect and categories lung cancer nodules.

REFERENCES

1. Welch HG, Schwartz LM, Woloshin S. Are increasing 5-year survival rates evidence of success against cancer? Jama. 2000;283(22):2975-8.
2. Asuntha A, Srinivasan A. Deep learning for lung Cancer detection and classification. Multimedia Tools Applications. 2020:1-32.
3. Nie L, Zhang L, Yang Y, Wang M, Hong R, Chua T-S, editors. Beyond doctors: Future health prediction from multimedia and multimodal observations. Proceedings of the 23rd ACM international conference on Multimedia; 2015.
4. American Lung Association, URL www.lung.org.
5. Siegel, Rebecca, Naishadham, Deepa, Jemal, Ahmedin. Cancer statistics, 2013. a cancer journal for clinicians. 2013;63(1):11-30.
6. W. H. Organization, "Cancer," 2018. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs297/en/
7. Indian J Med Res. 2015 Jan; 141(1): 5–7.Lung cancer: Prevalent trends & emerging concepts
8. ReLU activation Functions https://www.quora.com/What-is-the-ReLU-layer-in-CNN
9. IQ-OTH/NCCD - Lung Cancer Dataset. It Includes CT scans of patients diagnosed with Lung Cancer.
Available: - https://www.kaggle.com/adityamahimkar/iqothnccd-lung-cancer-dataset
10. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights into imaging. 2018;9(4):611-29.

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