WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Lung Cancer Detection Using Deep Learning

Siddhi Rajendra Dunke[1], Swapna Suresh Tarade [1], Pratiksha Balaji Waghule[2], Prof. S.R.Kolase[3]

[1]UG Student, Department of Computer Engineering, Jaihind College Of Engineering, Kuran , Pune, Maharashtra, India
[2]Professor, Department of Computer Engineering, Jaihind College Of Engineering, Kuran , Pune, Maharashtra, India

ABSTRACT

In recent year, so many computer aided diagnosis systems are designed for diagnosis of several diseases.Lung cancer detection at early stage has become very important and also very easy with image processing and deep learning techniques. In this study lung patient computed tomography scan images are used to detect and classify the lung nodules and to detect the malignancy level of that nodules. The CT scan images are segmented using U-Net architecture . This paper proposes 3D multi path VGG- like network, which is evaluated on 3D cubes, extracted from lung image database consortium and image database resource initiative (LIDC-IDRI).Lung Nodule Analysis are classified and malignancy level is detected using this CNN architecture with 94% of accuracy and 0.387732 of logloss.

Keywords-CNN,image processing,Deep learning,LUNA16,Data Science Bowl 2017.

1.INTRODUCTION

According to the survey of world Health Organization (WHO), Lung cancer was the second most leading cause of death in 2015 and it is on fifth rank 2017. It is most common in smokers accounting 85% of cases among all. So many CAD systems are developed in recent years. Detection of lung cancer at early stage is necessary to prevent the disease .Deep learning is proven to be the best method for medical image processing lung nodule detection and classification, feature extraction and lung cancer stage prediction. In the first stage of this system used image processing techniques to extract lung regions. The segmentation is done using K means. CT images are one of the filtering mechanism that use attractive fields to capture images in films. In this work, we use deep learning to identify lung cancer type from CT images of patients It has a two-fold challenge:artificial intelligent models trained by public datasets cannot meet such practical requires, and the amount of collected patients data is quite few. To slove the problem, we use densely connected convolutional networks to classify cancer. We can cure lung cancer, only if you identifying the yearly stage.So here, we use deep learning algorithms to detect the lung cancer. This can be made faster and more accurate. In this study we propose deep learning strategies to improve cancer characterization. Inspired by learning from CNN approaches, we propose new algorithm, proportion -PNN, to characterize cancer types. In the structure of cancer cell, where most of the cells are overlapped with each other. Hence early detection of lung cancer is more challenging task.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 rate.Deep learning not only accelerates the critical task but also improves the precision of the computer and the performance of CT images detection and classification.

Literature survey

In paper Pankaj Nanglia, Sumit Kumar focuses their study mainly on the classification of lung images as normal and adnormal . In proposed method median filter was used to eliminate impulse noise from the images. Mathematical morphological operation enables accurate lung segmentation and detect tumour region. Three gemotrical features i.e Area perimeter,eccentricity was extracted from segmented region and fed to the SVM classifier for classification.[1]

In paper Ruchita Tekade, Prof.Dr.K.Rajeshwari studied the concept of lung nodule detection and malignancy level prediction using lung CT images. This experiment has conducted using LIDC_IDRI,LUNA16 and Data science bowl 2017 dataset on CUDA enabled GPU Tesla K20. They used U-NET for segmentation of lung nodule from lung CT images and 3D multigraph VGG like architecture for classifying lung nodule and predict malignancy level. Combing these two approaches have given the better results. This approach given the accuracy as 95.66% and loss 0.09 and dice coefficient of 90% and for predicting log loss is 38%[2]

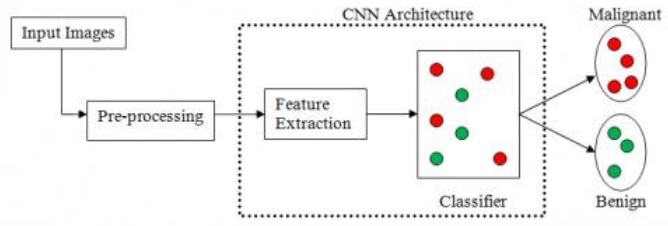
In paper Moffy Vas, Amita Desai, studied mainly on the classification of lung images cancerous and non-cancerous. In their proposed method pre-processing was done, in which unwanted portion of the lung CT scan was removed.Mathematical morphological operation enables accurate lung

segmentation and detect lung region.The training accuracy was 96% & testing accuracy was 92%[3].

In paper Radhika P R,Rakhi A.S.Nair, mainly focused on prediction and classification of medical imaging data. They used UCI machine learning repository and data world dataset. Used various deep learning algorithm for comparative study and found that support vector machine gives higher accuracy 99.2%. [4].

In paper K.Mohanambal, Y. Nirosha et al studied structural co-occurrence matrix(SCM) to extract the feature from the images and based on these feature categorized them into malignant or benign. The SVM classifier is used to classify the lung nodule according to their maligancy level(1 to 5)[5].

III.CNN Architecture



Model Development

We adopted the CNN architecture using segmentation method. The separation method method outputs more information than the detection method (which present a bounding box) or the classification method (which determine the malignancy from a single image). Maximal diameter of the tumor is specially important in clinical practice. Since the largest diameter of the tumor again and again coincides with an oblique direction, not the horizontal nor the vertical direction, it is tough to measure with detection methods which present a bounding box. Our CNN architecture was based on the encoder-decoder architecture to the output segmentation. The encoder-decoder architecture has a bottleneck structure, which overcome the resolution of the feature map and improves the model hardiness to noise and overfitting.

In addition, one characteristic of this DL-based model is that it used both a normal chest radiograph and a black-and-white transformation of a chest radiograph. This is an augmentation that makes use of the experience of radiologists. It is known that black-and-white transformation makes it easier to confirm the presence of lung lesions overlapping blind spots. We considered that this augmentation could be effective for this model as well, so we applied a CNN architecture to each of the normal and upent images and then an ensemble model using these two architectures. Supplementaryonline shows detailed information of the model.

Using the chest radiographs from the training dataset, the model was trained and validated from scratch, utilizing five-fold cross-validation. The model when the value of the loss function was the minimum within the 100 epochs using Adam (learning rate = 0.01, beta_1 = 0.09, beta_2 = 0.99, epsilon = 0.000001, decay = 0.0) was adopted as the best-performing.

V.Project Description

1.Pre-processing:

Pre-processing mention to the transformations apply to our data before feeding it to the algorithm.Data preprocessing is a method that is used to transform the raw data into a clean data set. In other words, whenever the data is assembled from different sources it is collected in raw format which is not achivable for the analysis.

2. Data training set:

The training data is used to make sure the machine recognizes of patterns in the data, the cross-validation data is used to ensure good accuracy and effectiveness of the algorithm used to train the machine, and the test data is used to see how good the machine can predict new answers based on its training.

3.Classification:

The classification process is used to recognize the category of the data. The classification is used to recognizedifficult data combinations, missing data's, out of range value, etc. The classification is used to delete the damaged data, and the empty data in the overall dataset.

3. Feature Extraction:

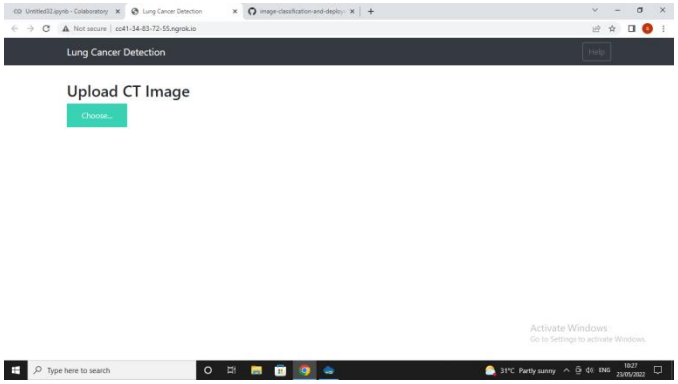
Feature extraction is used to overcome the amount of resources required to explain a large set of data. It starts from an starting set of measured data and builds derived values. When the input data to an algorithm is too big to be processed and it is think to be redundant, then it can be transformed into a reduced set of features. Determining a subset of the initialfeatures is said to be feature selectionExtraction.

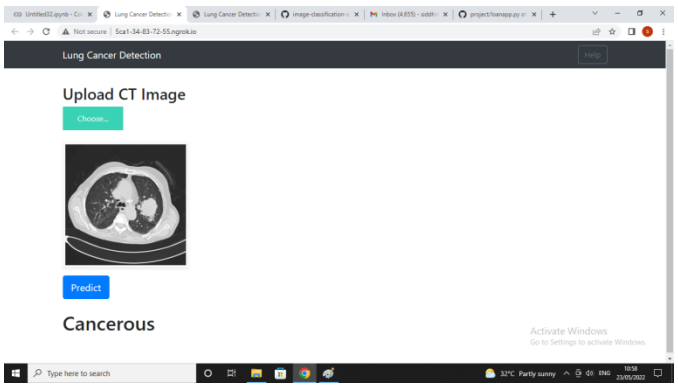
VI. Conclusion

The many algorithms and different methods for detection of lung cancer were studied in this observe. This work aims at detection of lung cancer using digital image processing techniques to get an better images of lung CTs and feed forward back propagation convolutional neural network which consists of input, hidden, output layer is trained to differentiate cancerous and non-cancerous images.

Result

Screenshots:





References

1. Bray F, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 2018;**68**:394–424. doi: 10.3322/caac.21492

2. Sagawa M, et al. The efficacy of lung cancer screening conducted in 1990s: Four case–control studies in Japan. *Lung Cancer.* 2003;**41**:29–36. doi: 10.1016/s0169-5002(03)00197-1.

3. Fontana RS, et al. Lung cancer screening: The Mayo program. *J. Occup. Med.* 1986;**28**:746–750. doi: 10.1097/00043764-198608000-00038.

4. Kubik A, et al. Lack of benefit from semi-annual screening for cancer of the lung: Follow-up report of a randomized controlled trial on a population of

high-risk males in Czechoslovakia. *Int. J. Cancer*. 1990;**45**:26–33. doi: 10.1002/ijc.2910450107.

5. Raghu VK, et al. Feasibility of lung cancer prediction from low-dose CT scan and smoking factors using causal models. *Thorax*. 2019;**74**:643–649. doi: 10.1136/thoraxjnl-2018-212638.