Lung Cancer Detection Using Convolutional Neural Network

Dunke Siddhi[1], Tarade Swapna[2], Waghule Pratiksha[3]

[1]Dunke siddhi, JCOE kuran
[2]Tarade Swapna, JCOE kuran
[3]Waghule Pratiksha, JCOE kuran
[4]Prof. Kolase Sachin, Department of Computer Engineering, JCOE, Kuran, Maharashtra

---

Abstract- Lung cancer is commonly cause of cancer death in the world. detection of lung cancer will help to save the patient. The CNN, a method that good describe a deep learning models featuring filter that can be trained with the local pooling operations being incorporated on input CT images in an alternating manner and create an array of hierarchical of complex features. This paper presents an approach which uses a Convolutional Neural Network (CNNs) to classify cancer or normal seen in lung cancer screening computed tomography scans as malignant or benign.

In this project, we have implement a CNN suitable for the analysis of CT scans with CT images, using domain knowledge from both medicine and neural networks.This result shows where patient has cancer or normal.

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

1.INTRODUCTION

Lung Cancer is the most common cancer among men and the third most common cancer in women.Cancer is the growth of abnormally and uncontrolled cells. It can damage the surrounding tissue spread far from its origin[1]. Malignant cause death and it could grow from every cell type in the human body.The prognosis of the disease has not been very favorable and this is largely due to the latens in detecting the presence of the malignancy. It has been reported that patients who had lung cancer treated at stage 1 have a better survival rate than those who are at the advanced stage of the disease[2].

Deep learning is use for the classification of CT Scan Images as cancerous/non-cancerous. The process of a feature extraction in Convolution Neural Networks is such that features are defined and computed by the algorithm itself. During the training stage, input and an output label are provided. Based on the given data, the algorithm analyses the features/patterns and for a training data, forms a set of parameters and feature extraction[3].

Convolution layers are used to define features and parameters.Pooling layers bring together the computations with similar permutation. The convolution filter will form a spatially dense output by assigning a common value to a set of matrix pixels These values decide the output for that image[4].

2.LITERTAURE SURVEY

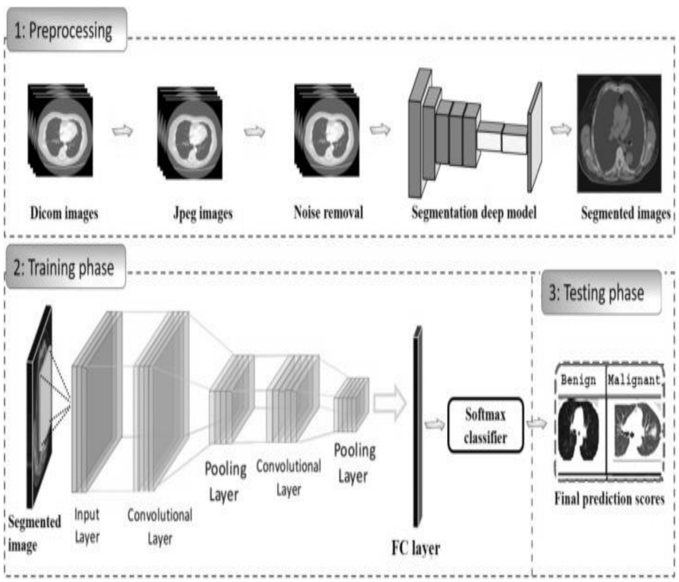
Convolutional neural network (ConvNet/CNN), Deep Learning or other machine learning algorithms have been used various researchers to perform experiments on different types of lung cancer detection. Shrinivas Arukonda used techique covolutional deep neural networks. In this the system capable to detecting lung cancer in its earlier stages because of its survival rate[1].

The dataset is used from the Data Science Bowl 2017 Kaggle competition ,LUNA16. first step, they detected the annotated nodules. Cube of size 32 X 32 X 32 is around the nodules, with the nodule in the center.Region of Interest mask is applied for lungs. Then cubes are made around the predicted nodules and the prediction is done using a second 3D CNN. The precision achieved in the model is found to be 94.30%[2].

Rohit Y. Bhalerao has create a novel approach for detecting of lung cancer using image. He has used convolutional deep neural networks. It is simple easy to understand and time consuming[3].

A. Badnjevic M. Crifrek has create cancer lungs detection using Aritificial neuralnetwork(ANN). In this ability to learn info in data can be used for classification. It is difficult to understand algorithm[4].

3. SYSTEM ARCHITECTURE



4. METHODOLOGY

4.1.Data Set:-

The database used is obtained from Lung Image Database LUNA16,Data Science Bowl 2017 This is a lung nodule classification database containing the scans of a total of 1018 patients. Each patients' CT scan in turn is comprising of around 150 to 550 dicom format images. The database provides four classifications namely-(i)Unknown, (ii)Benign, (iii)Malignant, and (iv)Metastatic

4.2.Convolutional Neural Network

The convolution layer of a produces a feature map by convolving different sub regions of the image with a learned kernel. Further, non-linear activation functions such as a sigmoid, tanh or rectified linear (ReLu) can also be applied. Another method for reducing computations is the pooling layer, where a region of the image/feature map is chosen and the maximum among them is chosen as the representative pixel. Hence, a 2x2 or 3x3 grid can be reduced to a single scalar value. traditional fully connected layer can also be used in conjunction with the convolutional layers, and are usually used towards the output stage.

4.3.Convolution Layer 1:

The data in 3-D hdf5 format forms the input to the first convolution layer. This layer of kernel size 50x50 with a stride of 6. The output of this layer produces 78 features. The weight filler is set to a 0.01 Gaussian distribution change and the bias is set at constant zero. This output is then fed to the Rectified Linear (ReLu) layer to bring all the negative activations to zero. The primary application of this layer is to detect the lowest level features, e.g., whether there is classification in some area of the image.

4.4. Convolution Layer 2:

The first Convolution layer output is fed into the second having a kernel size of 3x3 and a stride of 1. This layer pads the data with one enclosure of zeros. The weight filler is the same as convolution layer 1 and the bias is set to a constant value of 1. Also, this layer is followed by a ReLu layer. This layer is intended to make use of the information predicted from the previous layer and detect the pattern of calssification - e.g., popcorn, diffuse etc. From the training phase, it will hence learn as to which among the patterns are benign, and which are malignant. In this way the CNN achieves two objectives - it learns features hierarchically, and it eliminates the need for specific feature engineering.

4.5. Max-pooling Layer:

After the convolution layer 2 comes the max-pooling layer where the most responsive node of the given kernel is extracted. The kernel size used in the proposed network is 13x13 with stride shift of 13. This is primarily intended to reduce the computational effort. Since each CT scan composes of 500 images, if we have a batch size of 50, the number of required computations can be significantly large, leading to frequent memory overload. The max-pooling layer is used particularly to ease memory and data bottlenecks by reducing the image dimensions.

4.6 . Dropout layer:

The dropout layer is used in the network to prevent over-fitting. This is done by switching off random neurons in the network. Our proposed network uses a dropout layer with a drop ratio of 0.5. The intent of this layer is to improve the classification quality on test data that has not been seen by the network earlier.

4.7. Fully connected layer:

A fully connected layer which provides two outputs is used. It uses Gaussian weight filler of 0.5 and a constant bias filler of 0. The two output neurons from this layer gives the classification of benign or malignancy.

This layer is mainly intended to combine all the features into one top level image and will ultimately form the basis for the classification step.

5. IMPLEMENTATION

1. Download CT Scan Images

2. Load Dataset into python

3. Create a pandas dataframe

4. Map annotations to image filenames

5. Extract images

6. A.1 Understand the 3D image data

   A.2 Visualize images

   B.1Segment lungs & cancer

   B.2 Save Segmented image into a file.npy

7. Upload the segmented images into a file.npy

8. Train the U-Net Model

9. Evaluate model

10 .Save model as .hdfs

10. Download model from floyhub

11. Create flsak backend

12. Create front end

13. Load model into web application

14. Deploy

6. FUTURE WORK

Doctor's who work in this field are prone to observer fatigue from viewing so many CT scan images. The research on that suggests that observer fatigue increases the risk of errors that can be made by doctors while analyzing these scans. Many images in a CT scan also are irrelevant to Doctors e.g. for 200-300 images only 3 scans would show cancer depending on the stage of the patient. Although this feature was not implemented on the website, a more efficient deep learning model would be capable of alleviating these additional challenges.

7. CONCLUSION

This chapter details technical implementation of the project. Although the deep learning model only performs on a 65% accuracy on the training set it is still able to create masks and find cancer within new instances given the risk of high false positives. The model was integrated into a web application and performs the main functionality of this project.

8. REFERENCES

[1].Srinivas Arukonda, *S.Sountharrajan Investigation of Lung Cancer detection Using 3D Convolutional Deep Neural Network(2020)

[2].Rohit Y. Bhalerao A novel approach for detection of Lung Cancer using Digital Image Processing and Convolution Neural Networks.(2019)

[3].Lilik Anifah Cancer Lungs Detection on CT Scan Image Using Artificial Neural Network Backpropagation Based Gray Level Coocurrence Matrices Feature(2017).

[4]Prajwal Rao∗, Nishal Ancelette Pereira†, and Raghuram Srinivasan Convolutional Neural Networks for Lung Cancer Screening in Computed Tomography (CT) Scans(2016).