International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

IJIEMR Transactions, online available on 05th Apr 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/35

Title **AN EFFECTIVE DEEP LEARNING APPROACH FOR PULMONARY NODULE CLASSIFICATION**

Volume 12, ISSUE 04, Pages: 261-267

Paper Authors

Mr. B. Manikanth, B.Anishitha, A. Madhu Bindu, D. Basava Chandrika, A. Maha Lakshmi

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

An Effective Deep Learning Approach for Pulmonary Nodule Classification

Mr. B. Manikanth, Assistant Professor, Department of ECE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
B.Anishitha, **A. Madhu Bindu**, **D. Basava Chandrika**, **A. Maha Lakshmi**
UG Students, Department of ECE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
manikanth439@gmail.com, anishithabokka26@gmail.com,
amadhubindu1@gmail.com, chandud2708@gmail.com,
mahaamla11@gmail.com

Abstract
One of the deadliest malignancies in the world is lung cancer. The survival percentage for lung cancer can be considerably increased by early identification. Pulmonary nodules are the small growths of cells inside the lung. The size of the nodule determines the lung as cancerous or non- cancerous. Detection of malignant lung nodules is necessary at an early stage for necessary treatment. There are numerous techniques available for diagnosing lung cancer. Using CT (Computed Tomography) and CNN (Convolutional Neural Network) with image segmentation is one of the simplest methods. CNN is one of the deep structured algorithms widely used to analyze the ability to visualize and extract the hidden features of image datasets. The suggested approach will work well for early detection of cancer.

Keywords: malignancy, CT (Computed tomography), CNN(Convolutional neural network)

Introduction

Now-a- day's many people are dying to lung cancer than other types of cancers. Lung cancer is a type of cancer that begins in the lung. In general lung cancer is spotted based on the existence of the lung nodules. Lung nodules are small masses of tissue. The size of the cancerous nodules is greater than that of healthy nodules. Lung cancer typically doesn't cause signs and symptoms in the early stages. So, it is necessary for the trained radiologists to identify the cancer accurately in the early stage to reduce the deaths of humans.

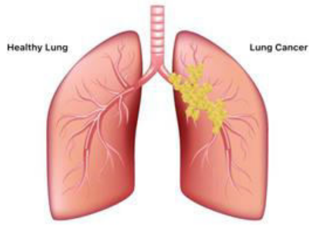

Fig.1. Healthy lung Vs Malignant lung

Therefore, it is very difficult to identify lung nodules of both cancer and non-cancer at an early stage. Radiologists have recently been manually analyzing lung CT scan images to look for potential nodules and distinguish between cancerous and non-cancerous nodules in those nodules. This extremely time-consuming and labor- intensive operation necessitates a high level of expertise in lung nodules. The computer-aided diagnosis (CAD) system can be used to identify lung nodules and categorize them as either cancerous or non-cancerous, which will help to resolve the issue. The radiologist will use this technology as a second opinion to find and examine lung nodules. Computed tomography (CT), contrast-enhanced computed tomography (CECT), low-dose computed tomography (LDCT), and positron emission tomography are non- invasive imaging techniques that can detect lung cancer (PET). The detection of distinctive characteristics in images is done using conventional image processing algorithms. So, it is necessary to construct hand-crafted characteristics that are learned through a manual procedure. Hence, it is exceedingly difficult to distinguish between malignant and non-cancerous nodules using this method. Hence, deep learning may avoid all of these issues and can deal with issues like picture and video recognition, voice and natural language processing, and other issues. Manual feature extraction demands the designer to have a thorough understanding of lung cancer. Deep learning will be able to learn every characteristic present in the photos. In particular, Convolutional Neural Network (CNN) recovers the features of the input images using one or more hierarchically organized layers of convolution, sub sampling, or max-pooling. Hence, appropriate deep learning algorithms were applied to process these CT scan images from the medical field.

Literature Survey

The paper by Masud M, Sikder N, et al. classifies the image using a CNN-based model. The dataset being used is histopathological, which is the microscopic analysis of an invasive biopsy. Our method favors using CT scans, a non-invasive method of cancer detection. Also, we'll be using the LUNA 16 dataset.

In research that was published, Sajja T, Devarapalli R, et al. demonstrated how to use the pre-trained CNN model known as Google-Net to identify lung cancer. The model still has to be tested using different dropout rates to see whether it performs more accurately. Our strategy attempts to create a streamlined CNN model to categorize cancer.

In a work published by Tripathi P, Tyagi S, et al., they test the ability of four distinct image processing segmentation algorithms to identify lung cancer. The comparison research reveals that CT scans typically offer the best likelihood of finding cancer. So we are using deep learning on CT scan images.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Problem Identification

The existing CAD system used for early detection of lung cancer has been unsatisfactory. Earlier systems mostly used invasive methods for the detection of cancer. Segmentation techniques used for the image processing did not give satisfactory results. And In most of the existing methods data was overfitted. Classification of nodules was not done correctly. Above all the accuracy of the existing systems was low.

Methodology

The proposed model is a convolutional neural network (CNN) approach based on lung segmentation on non-invasive image modalities called CT (Computed Tomography) scan images. Our proposed model is a sequential model which allows us to specify a neural network, precisely, sequential from input to output, passing through a series of neural layers, one after the other. Convolutional neural networks are made to classify images by reducing the number of parameters and modifying the network's architecture. Layers in convolutional neural networks are arranged in accordance with their characteristics and functions. The design of a CNN closely resembles the neuronal connection network found in the human brain.. The first step is to make lung segmentation of the dataset with a watershed algorithm. Watershed algorithm highlights lung part and makes binary masks for lungs using semantic segmentation approach. We are also going to use VGG16-net Transfer learning model to compare our proposed model. VGG16 is one of the best and simple models of CNN. VGG stands for Visual Geometry Group, the 16 in VGG16 is the name given to 16 layers with weights. Only sixteen of VGG16's total of 21 layers—13 convolutional layers, five Max Pooling layers, and three dense layers— are weight layers, or layers with learnable parameters. The most notable aspect of VGG16 is that it constantly used the same padding and maxpool layer of a 2x2 filter with stride 2, as well as a 3x3 filter with stride 1's convolution layers rather than a lot of hyper-parameters. The convolution and max pool layers are consistently arranged throughout the architecture.

Implementation

At first we have gathered the dataset and we need to split the dataset into training dataset and testing dataset. After that CNN takes this data (training dataset)

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

and uses filters on the image to learn detailed patterns. The images from the dataset passes through various layers in the CNN model and the CNN model extracts the required features from the data and classifies the image as 0 or 1 i.e non-cancerous or cancerous. This is how the training process takes place. After that we have to test the model by giving some data (testing dataset). We will see whether we are getting satisfied output or not. If we are getting satisfactory output then our model is ready to use. Otherwise, we have to work on the errors.

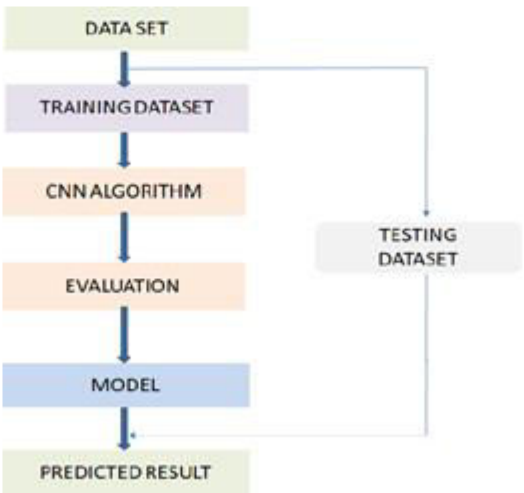

Fig.2. Work flow of proposed model

We are using a LUNA 16 dataset which consists of 6700 images. In which 5100 images are used for training and 1600 images are used for testing.

Folder	Benign images	Malignant images
Training	3250	1850
Testing	1320	280

Fig.3. Dataset classification

Results

The proposed Sequential model is the Deep Convolutional Neural Network with Max Pooling and Fully connected layers in the end. Simulation results of the proposed sequential model:

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 50, 50, 32)        320
_________________________________________________________________
batch_normalization (BatchNo (None, 50, 50, 32)        128
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 50, 50, 32)        9248
_________________________________________________________________
batch_normalization_1 (Batch (None, 50, 50, 32)        128
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 25, 25, 32)        0
_________________________________________________________________
dropout (Dropout)            (None, 25, 25, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 25, 25, 64)        18496
_________________________________________________________________
```

```
_________________________________________________________________
conv2d_8 (Conv2D)            (None, 3, 3, 128)         147584
_________________________________________________________________
batch_normalization_8 (Batch (None, 3, 3, 128)         512
_________________________________________________________________
conv2d_9 (Conv2D)            (None, 3, 3, 128)         147584
_________________________________________________________________
batch_normalization_9 (Batch (None, 3, 3, 128)         512
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 1, 1, 128)         0
_________________________________________________________________
dropout_4 (Dropout)          (None, 1, 1, 128)         0
_________________________________________________________________
flatten (Flatten)            (None, 128)               0
_________________________________________________________________
dense (Dense)                (None, 128)               16512
_________________________________________________________________
batch_normalization_10 (Batc (None, 128)               512
_________________________________________________________________
dropout_5 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_1 (Dense)              (None, 128)               16512
=================================================================
Total params: 914,144
Trainable params: 911,968
Non-trainable params: 2,176
_________________________________________________________________
```

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Fig.4. Sequential model summary

The model graph of proposed sequential model representing training Vs validation accuracies and training Vs validation loses

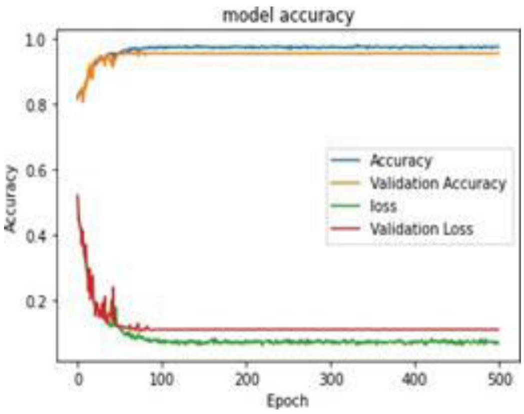

Fig.5. model graph

Confusion matrix of the proposed sequential model is

```
confusion matrix
[[1311   29]
 [  43  239]]
```

Fig.6. Confusion matrix Classification report of our proposed Sequential model is

Classification report

	precision	recall	f1-score	support
0	0.97	0.98	0.97	1340
1	0.89	0.85	0.87	282
accuracy			0.96	1622
macro avg	0.93	0.91	0.92	1622
weighted avg	0.95	0.96	0.96	1622

Fig.7. Classification report

Comparison of existing VGG 16 model ad proposed sequential model based on accuracies and loses is

Sequential model

Training Accuracy: 98.53%

Training Loss: 0.0442

Validation Accuracy: 96.00%

Validation Loss: 12.8

VGG 16 model

Training Accuracy: 99.44%

Training Loss: 0.0046

Validation Accuracy: 92%

Validation Loss: 28.26

Conclusion

Both the Sequential model and the VGG 16 model performed well during model training, achieving respectable levels of test accuracy and test loss, while Sequential fared better in validation in terms of accuracy and loss when compared to VGG 16.

In order to facilitate a quicker and more accurate diagnosis of lung cancer, the proposed sequential model uses a convolutional neural network (CNN) technique based on lung segmentation on CT (Computerized Tomography) scan pictures, a non-invasive image modality.

Future Scope

There is a scope to improve the efficiency of the model by using various hybrid algorithms and different segmentation techniques like cluster based segmentation and edge-based segmentation etc. and we can use 3d CT scan images and also increase the efficiency of model by using different datasets.

References

[1] Sahu AK, Sharma S, Raja R (2022) Deep learning-based continuous authentication for an IoT-enabled healthcare service. Computer Electr Eng99:107817.ISSN.0045-7906. https://doi.org/10.1016/j.compelec eng.2022.107817

[2] Qinghai Zhang and Xiaojing Kong, "Design of automatic lung nodule detection system based on multi- scene deep learning framework", IEEE Access, vol. 8, pp. 90380–90389, 2020, 10.1109/ACCESS.2020.2993872.

[3] Wenkai Huang and Lingkai Hu, "Using a noisy U-net for detecting lung nodule candidates", IEEE Access, vol. 7, pp. 67905–67915, 2019, 10.1109/ACCESS.2019.

[4] Wangxia Zuo, Fuqiang Zhou, Zuoxin Li and Lin Wang, "Multiresolution CNN and knowledge transfer for candidate classification in lung nodule detection", IEEE Access, vol. 7, pp. 32510–32521, 2019, 10.1109/ACCESS.2019.2903587

[5] Tanzila Saba, "Automated lung nodule detection and classification based on multiple classifiers voting", Microscopy Research and Technique, vol. 82, no. 9, pp. 1601– 1609, 2019.

[6] Xiaojie Huang, Junjie Shan and Vivek Vaidya, "Lung Nodule Detection in CT using 3D Convolutional Neural Networks", *IEEE*, 2017.

[7] Jia Ding, Aoxue Li, Zhiqiang Hu and Liwei Wang, *Accurate Pulmonary Nodule Detection in Computed Tomography Images Using Deep Convolutional Neural Networks*, 2017

[8] hhttps://drive.google.com/drive/fold ers/1lmMymeoxie_yl1aNKZJdEOg1i XPP_Epp?usp=sharing

[9] Hongyang Jiang, He Ma, Wei Qian, Mengdi Gao and Yan Li, "An automatic detection system of lung nodule based on multigroup patch based deep learning network", IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 4,

pp. 1227–1237, 2017.

[10] Kumar D, Wong A, Clausi DA (2015) Lung nodule classification using deep features in CT images. In: Proceedings—2015 12th conference on computer and robot vision, CRV 2015, pp133–138. https://doi.org/10.1109/CRV.2015. 25