International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org ||Volume 9 Issue 1 Series. II || Jan.2020 || PP 63-66

Automatic Classification of Lung Nodules on Computed Tomography Images Using a Pre-Trained Convolutional Neural Network.

Oluwadare Adebisi[1], Olukayode Busari [2] , Yemisi Oyewola[3], Iyabo Adeaga [4]
[1, 2, 3, 4]*(Department of Computer Engineering Technology, The Polytechnic Ibadan, Nigeria)*
Corresponding Author: Oluwadare Adebisi

Abstract : *Computed Tomography (CT) is the most common imaging technique that is used to capture lung images in order to differentiate between begnin and malignant lung nodules. However, interpretation of the lung nodules images is a complicated process. Therefore, visual interpretation of CT scan images result in subjective interpretation, inter-observer variability, and time consuming process. Computer aided diagnosis system assist radiologist in interpreting images by detecting and classifying images in short time. Recent advancements in deep learning enhance the classification of images. This paper thus presents Computer Aided Diagnosis (CAD) system for classification of lung nodules into begnin and malignant using a pre-trained Convolutional Neural Network. The developed CAD system achieved an accuracy of 95%, sensitivity of 94.73% and specificity of 98.38 %.*

Keywords - *Computed Tomography, begnin, malignant, Computer aided diagnosis system, Convolutional Neural Network*

Date of Submission: 20-01-2020 Date of Acceptance: 05-02-2020

I. Introduction

Lung nodule refers to a spot on the lung. It is round or oval shape. Lung nodule can be begnin or malignant. Research has shown that 26% of cancer deaths result from lung cancer [1]. Therefore, it is necessary to detect and classify abnormalities in lung nodules. Although, excisional biopsy is the best for differentiation of lung nodules into begnin or malignant, it is very labor intensive for large scale screenings. Unnecessary biopsy will make patients more anxious and increase health cost. Therefore, a non –invasive method is necessary for lung nodules diagnosis.

Several non-invasive imaging techniques are used to capture lung nodules images such as X-ray, Computed Tomography (CT) and Ultrasound (US). CT scan is commonly used to capture lung nodules images because of its ability to form three-dimensional (3-D) chest images which result in greater resolution. However CT scan images are difficult to interprete because of similar CT images features. Computer Aided Diagnosis (CAD) systems assist to detect earliest signs of abnormality in patients [2] evaluate, and interpret information obtained from medical imaging in short time for proper detection and diagnosis of diseases [3].

II. Literature Review

Kawata et al. [4] developed CAD system to classify lung nodules based on toplogical and histogram features. The voxel in the nodule were aggregated through shape and histogram to quantify how much shape classes were present in the nodule. Dandil et al. [5] applied Artificial neural network (ANN) to develop a CAD system to classify lung nodules and obtained a accuracy of 90.63%, sensitivity 92.3% and specificity of 89.47% respectively. Xie et al.[6] presented an algorithm that fuses texture, shape and model information (FUSE-STD) at the decision level in order to classify lung nodules into begnin and malignant. The FUSE-STD algorithm achieved a AUC of 96..65%, 94.45% and 81.24%. Jiang et al. [7] applied a novel Pixel Value Space Statistical Map (PVSSM) to classify lung nodules into begnin and malignant. Set of features from the created feature maps were extracted by Singular Value Decomposition (SVD) technique. A classification accuracy of 77.29%, 80.72% and 84.21 % respectively with K-Nearest Neighbor (KNN), Random Forest ad Support Vector Machines were obtained.

Many methods have been applied to classify lung nodules into begnin and malignant, but there are still room for improvement in their performance. Deep learning is an area of machine learning that has produced great improvement in image classification [8] but there are still challenges of large scale training medical set. Transfer learning solves this problem by transferring knowledge from a large dataset to a smaller dataset. Transfer learning is usually achieved by pre-trained models. Most pre-trained models used in transfer learning are based on convolutional neural networks [9]. This paper aims to develop a system for classification of lung nodules on CT images using a pre-trained Convolutional Neural Network.

III. Methodology

3.1 Data Acquisition

800 Lung images were obtained from the Lung Image Database Consortium image collection (LIDC-IDRI) [10]. LIDC-IDRI consists of diagnostic and lung cancer screening thoracic computed tomography (CT) scans. The database was initiated by National Cancer Institute (NCI). It is an accessible resource for development, and evaluation of computer-assisted diagnostic (CAD) methods for lung nodules detection and classification. The dataset was created by seven academic centers and eight medical imaging companies and contain 1018 cases. The LIDC-IDRI lung nodules is represented as XML file with radiologist's annotation and diagnostic description of suspicious lung lesions performed by four experienced thoracic radiologists. Samples of lung nodules images are shown in Fig. 1. The dataset consists of lung nodules which are grouped to either begnin or malignant.

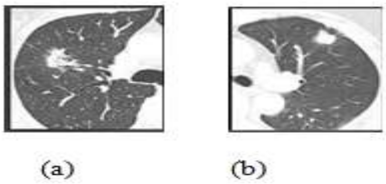

Fig. 1: Sample images of Lung nodules

(a) Begnin (b) Malignant

3.2 Preprocesing

The lung nodules dataset was divided into training set (60%), validation set (20%), and test set (20%). The test set only contains samples from the original dataset, not augmented data. Histogram Equalization was used to enhance the lung nodules image contrast while Adaptive Filtering was applied to remove the image noise.

3.4 Classification of Lung Nodules

MATLAB 2018 was used for the whole simulation process, with initial start-up of loading images into the system. Alexnet is the pre-trained convolutional neural network that was modified to classify lung nodules into begnin and malignant. Training a model from scratch requires a lot of labeled training data and computing power . A pretrained network is usually much faster and easier than training a new network because learned features can be applied to a new task using a smaller number of training images. Alexnet is a convolutional neural network that was developed by Krizhevsky et al.[11] and trained on more than a million images from the imagenet database [12] and can classify images into 1000 object classes. The network consists of 8 convolutional layers and input size of 227 × 227×1.

The Alexnet transfer learning network was modified to classify lung nodules as shown in Fig. 2. It consists of five convolutional layers and three fully connected layers. Each convolutional layer is followed by rectified linear units. A softmax layer was added after the final layer to classify the lung nodules and a max pooling layer was applied after each convolution layer to reduce the network size. To fit Alexnet, the original size of the lung images were resized to 227× 227×1 which is the Alexnet input size. Furthermore, Alexnet requires 3-channels input data, therefore, the lung images were converted from grayscale to RGB (227 × 227 ×3) by concatenating their channel three times. The last three layers were replaced with new layers applied to lung nodules. A new layer with two output neurons which corresponds to the two of classes of the lung nodules. was used to replace the final fully connected layer.

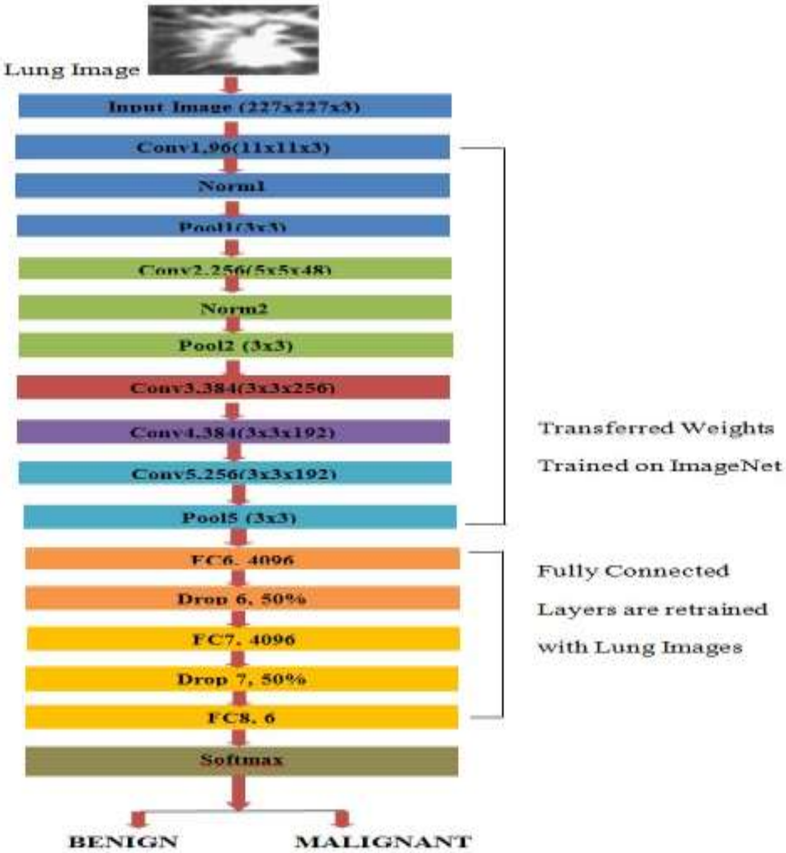

Fig. 2 : Alexnet transfer learning network for classification of lung nodules

Data augmentation was applied to multiply the number of lung images in order to avoid overfiting of the network. The Lung images were rotated left and right and then flipped 70, 160 and 270 degrees, resulting to a total number of 3000 images. The training options were specified and the Alexnet was trained with training lung images The modified Alexnet was simulated in MATLAB environment . The network was trained on a laptop computer system with an intel core processor i5, 2.8 CPU processor speed, 8 GB of RAM.

3.5 Performance Evaluation of the Classification System

Accuracy, True Positive Rate (Sensitivity**)** and True Negative Rate (Specificity) were determined in order to evaluate the performance of the classification system [13]. Accuracy is the ratio of number of correct predictions to the total number of input lung images samples which is expressed in equation (1) as:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Number of Samples}} \qquad (1)$$

Sensitivity (True Positive Rate) is the proportion of positive data points that are correctly considered as positive, with respect to all positive data points. Sensitivity is defined in equation (2):

$$\text{True Positive Rate} = \frac{\text{True Positive}}{\text{False Negative} + \text{True Positive}} \qquad (2)$$

Specificity corresponds to the proportion of correct negative predictions divided by the total number of negatives defined in equation (3) as

$$\text{True Negative Rate} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \qquad (3)$$

IV. Result And Discussion

The confusion matrix for the classification of lung nodules is as shown in table 1. For begnin, lung nodules, 90 were correctly classified while 5 were misclassified. Similarly, For malignant lung nodules, 62 were correctly classified while 5 were misclassified. The modified Alexnet results in accuracy, sensitivity and specificity of 95%, 95% and 93 % respectively as shown in Table 3. The results indicate a good performance of the developed CAD system.

Table 1: Confusion Matrix for the Classification of Lung Nodules

		Predicted Class	
		Begnin	Malignant
Actual Class	Begnin	90	5
	Malignant	3	62

Table 2: Result of the Performance Evaluation of the Classification System

Metrics	Accuracy %	Sensitivity %	Specificity %
	95	94.73	95.38

V. Conclusion

An efficient computer aided diagnosis system for classification of lung nodules into begnin and malignant has been developed using a pre-trained CNN (Alexnet). The computer aided diagnosis system will be an aid to radiologist by providing second opinion for characterisation of lung nodule which will consequently leads to reduction of false diagnosis of lung nodules. However, only lung nodules obtained from CT scan have been considered. Further research work can be done to classify lung nodules obtained from other imaging modalities into begnin and malignant using modified Alexnet .

References

[1]. X. Wang, K. Mao, L. Wang, P. Yang, D and Lu, D.. An Appraisal of Lung Nodules Automatic Classification Algorithms for CT Images., *Sensors* 19(1), 2019, 94.

[2]. A. Lostumbo, K. Suzuki, H. Dachman, Flat lesions in CT colonography. *Abdom Imaging.* 35, , 2010,578–583

[3]. A.*Taki, A. Kermani*, S.M *Ranjbarnavazi, and A. Pourmodheji*, Computing and Visualization for Intravascular Imaging and Computer-Assisted Stenting, Elsevier Academic Press , *2017,106*

[4]. Y. Kawata, N. Niki, H.Ohmatsu, M Kusumoto, R.Kakinuma, K.Mori, H.Nishiyama, Computer-aided differential diagnosis of pulmonary nodules based on a hybrid classification approach. Med. Imaging: Image Process. 2001;4322:1796–1806.

[5]. E. Dandil, M Çakiroglu., Z Eksi., M.Ozkan, O.K. Kurt, A. and Canan, Artificial neural network-based classification system for lung nodules on computed tomography scans; Proceedings of the 2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR); Tunis, Tunisia. 11–14 August 2014, 382–386

[6]. Y.T. Xie, J.P. Zhang, S.D. Liu, W.D. Cai and Y. Xia, Lung Nodule Classification by Jointly Using Visual Descriptors and Deep Features; Proceedings of the Medical Computer Vision and Bayesian and Graphical Models for Biomedical Imaging; Athens, Greece. 21 October 2016, 116–125.

[7]. H.Y.Jiang., H Ma., W. Qian, G.H Wei, X.Z. Zhao and M.Gao, A novel pixel value space statistics map of the pulmonary nodule for classification in computerized tomography images; Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jeju Island, Korea. 11–15 July 2017; 556–559

[8]. Y Bengio, Y. *LeCun*,.G..and Hinton, ."Deep Learning". Nature. *521 (7553)*, 2015, *436–444.*

[9]. Y. LeCun, L.Bottou, Bengio Y., and P. Haffner Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE 86, 1998, 2278-2324

[10]. https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI, Date Accessed 3rd Sepetember, 2019

[11]. Krizhevsky A., I. Sutskever, and G. Hinton (2012) ImageNet Classification with Deep Convolutional Neural Networks, *Advances in neural information processing systems* 1-5, 2012

[12]. http://www.image-net.org, Date Accessed 2nd Sepetember, 2019

[13]. Mishra, Metrics to Evaluate your Machine Learninng Algorithm,https://towardsdtasciece.com/metrics-to-evaluate-your-machine-f10bba6e38234,2018,Date accessed 3rd September, 208

Dr. Nlerum, Promise A"Machine Learning: Application To Pervasive Computing Systems " *International Journal of Engineering Science Invention (IJESI)*, Vol. 09(01), 2020, PP0 63-66