JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Optimization Accuracy of Lung Diseases Prediction using Convolution Neural Network

[1]Shradha Shrivastava, [2]Prof. Ruchi Dronawat, [3]Dr. Ritu Shrivastava
M. Tech. Scholar, Department of Computer Science and Engineering, SIRT, Bhopal[1]
Professor, Department of Computer Science and Engineering, SIRT, Bhopal[2]
Head of Dept., Department of Computer Science and Engineering, SIRT, Bhopal[3]

Abstract: Nowadays nanotechnology is gaining more advantages and widely used in many real-life applications including minute tumor detection and effective diagnosis. Nanoscale imaging technique significantly increases the precision, accuracy of tumor detection, and classification of the tumor into benign or malignant. X-rays & Computerized Tomography (CT) scans are important for lung cancer diagnosis and research because it gives accurate segmentation results. To detect the presence of lung cancer, the basic steps to be followed are preprocessing and segmentation. Preprocessing is required to ensure that the dataset is consistent and displays only relevant features information. Regarding collect the data of available lung cancer images, the tumor type, and the efficiency measure of the dataset preprocessing techniques are applied to make the segmentation more accurate. Further, the lung tumor boundary area is measured by using a nanotechnology-based detection scheme and then the convolution neural network classifier is used to classify the lung cancer image.

Index Terms – Lung Disease, Cancer, Convolution Neural Network

I. INTRODUCTION

Lung cancer has been identified as the main cause of death and is very challenging to detect early as most symptoms appear during the advanced stages. More cases of deaths occur due to lung cancer than other types of cancer, such as breast cancer, colon, or prostate cancers. Various techniques are available to detect lung cancer, such as the Sputum Cytology, Chest Radiograph (x-ray), Magnetic Resonance Imaging (MRI), and Computed Tomography (CT). However, most of these are expensive and require a lot of time to complete [1, 2]. Various techniques detect lung cancer in progressive phases, but the chances of survival of patients are low. A tool consisting of good quality for improving manual cancer analysis can be image processing. Various medical researchers with an analysis of sputum cells can identify lung cancer in its early stages.

The histologic classification of lung cancer includes non-small cell and small cell lung cancers. It can have common symptoms like dyspnea, cough or hemoptysis, and systematic symptoms like anorexia leading to weight loss. Chest radiography is the most common diagnostic procedure used by high-risk patients, also computed tomography, and positron emission tomography is used as a likely alternative. Diagnostic evaluation is carried out only if the suspicion for the disease is high. The diagnostic evaluation determines prognosis and leads to the plan of care. It should be kept in mind that the least invasive method should be utilized. Treatment of lung cancer involves the teamwork of pulmonologists, oncologists, and radiation oncologists, radiologists, pathologists, and thoracic surgeons [3].

It is the duty of the family physician to assure proper plan of care to the patient with moral support to the family. In case of necessity, a coordinated plan of care is done until the end of life. Palliative care, even at the point of disease detection, can improve life‟s quality and prolong the chance of survival. Primary detection of lung cancer and ensuring tobacco cessation at every visit should be the prime responsibility of the family physician. There has always been a dispute between the U.S. Preventive Services Task Force and the American Academy of Family Physicians concerning the type of screening that has to be implemented regarding the high-risk patient. It should be ideal if both the doctor and the patient themselves decide the appropriate screening process. In the USA, the leading cause of death is due to cancer. Two hundred twenty-one thousand (221,000) new cases of this disease are diagnosed in 2011, while one hundred fifty-six thousand and nine hundred (156,900) individuals died due to the disease, and the prognosis is that just 15% survived at the end of five years after their initial diagnosis. Eighty-five percent (85%) of all lung cancer is Non-Small Cell Lung Cancer (NSCLC) [4, 5]. Unfortunately, only fifteen percent (15%) are diagnosed at an early stage, i.e., when the cancer is restricted to its primary site, and almost 56% of the patients are diagnosed after the patient has reached metastasis. This group has the poorest prognosis, and their assessed survival rate after five years is just 3.6% [6].

II. LUNG CANCER

Studies show that in the United Kingdom, over 3000 deaths per year take place due to lung cancer, which is high when related to new European countries, though the reason is not clear. However, stress is laid upon early diagnosis; experts also opine that symptoms of lung cancer remain silent and could not be detected until is quite advanced at which stage it becomes beyond redemption so that saving the patient becomes out of question. Though the health professionals are trying to hasten up the early

diagnosis of the disease, due to the aforesaid reasons, delay has become unavoidable. Both doctors and patients can attribute to delay in diagnosis, as there can be a delay in the commencement of investigations, or there can also be failures at the hospital end in mobilizing records, which can, in turn, lead to delay in treatment. Moreover, patient‟s records have been the prime source of information, and the delay by the patient in visiting the doctor at the advent of symptoms can be the main reason for the delay [7].
It has been stressed the importance of continued effort in reducing the delays in diagnosing and treating patients with lung cancer. They are screened at primary care; not much information is available on the role of general practitioners in conducting surgeries for lung cancer. Moreover, differentiating patients with lung cancer and ones with nonspecific symptoms of minor illness pose difficulty to the general practitioners. Though the symptoms and diagnosis are quite known to the medical fraternity, the various types and stages are not known, and also very little information is available on how these arise or how patients answer back to them before diagnosis. Screening aims at increasing life expectancy and refining the quality of life, but unfortunately, decades of screening of lung cancer have no survival benefits.
Moreover, Soan-Kettering Study and Johns Hopkins Study added sputum technology to chest radiography, with both failed to begin advantages with respect to survival time. Though chest radiography and sputum cytology are recommended annually, the Mayo Project tried an intensive program of chest radiographs and sputum cytology taken every four months. Again, after a median follow-up of 20.5 years, there was an increase in the survival rate in the intervention arm by five years; the specific morality remained unchanged. In comparison to the usual plan of care, lung, colorectal, ovarian cancer screening trials were randomized to baseline for the participants with three annual trials for smokers plus two for non-smokers. The anterior and posterior chest radiographs were compared with usual care. Over 150,000 participants between the age group of 55-74 who were enrolled from 1993 to 2001 were tested. The results published in 2011 did not show any benefits for chest radiology and the histology where stage distribution remained similar in both the arms [8, 9].

III. LUNG CANCER DETECTION

In the field of medicine, while detecting disease, a number of image processing techniques can be applied. While detecting lung cancer using CT images, four main steps are involved. At every step, several techniques are applied whose accuracies are different in the detection of lung cancer. The initial step is to remove the noises that exist by pre-processing the lung CT image; secondly, segmentation of the image is done to get Region of Interest (ROI).Third step is to apply feature extraction to extract features such as entropy, energy, variance, and finally, the various classification Algorithms are applied on the features of the lung tissue that are extracted from the lung CT image. All the steps involved in detecting cancer are depicted in Fig. 1.

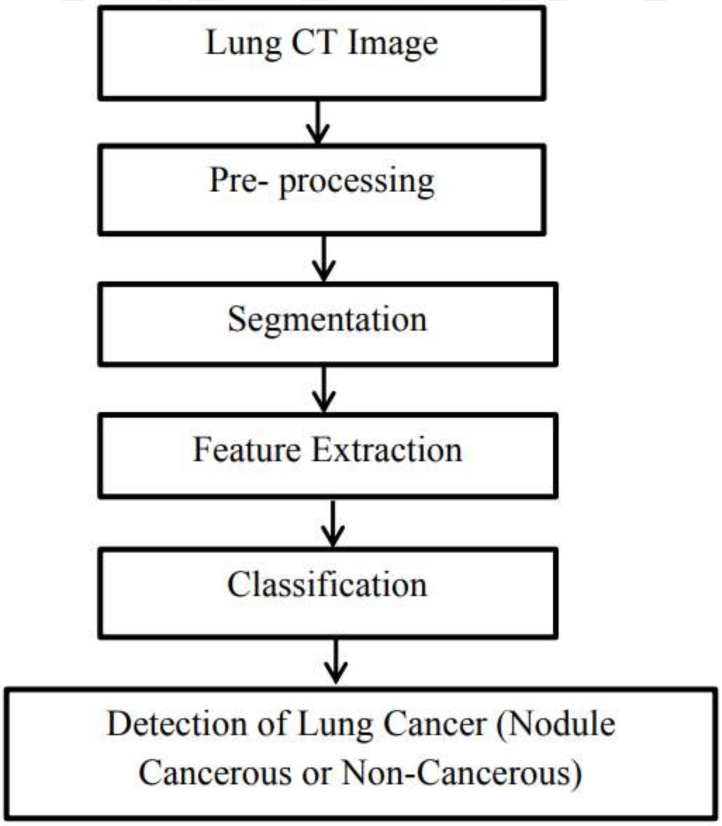

Fig. 1: Steps involved in detecting Lung Cancer

Pre-Processing of Lung Cancer: Unwanted distortions in the image data are improved during pre-processing, or otherwise, it improves certain image features for further processing. Moreover, it is required to minimize the effects of distortion that are present in the imaging device, including light fluctuation, which removes the blueness. It enhances image features such as line, boundary and texture of the image so that it can divide the contents of images easily in two parts, i.e., wanted and redundant image.

IV. CONVOLUTION NEURAL NETWORK

Lung cancer is a leading cause of death worldwide. Currently, in the differential diagnosis of lung cancer, accurate classification of cancer types (adenocarcinoma, squamous cell carcinoma, and small cell carcinoma) is required. However, improving the accuracy and stability of diagnosis is challenging. The main goal of this thesis is to automate the classification using Convolutional Neural Networks (CNN). With the introduction of Convolutional Neural Networks, the field of pattern recognition broadened. The classical way of designing and extracting hand-made features for classification is substantially different from letting the computer itself decide which features are of importance; the new approach was enabled by CNNs. This together with groundbreaking results on benchmark image sets has made CNN's a well-used method in pattern recognition. In this research work, an automated classification scheme for lung cancers is discussed in CT images using a convolutional neural network (CNN), which is a major deep learning technique. The CNN used for classification consists of three convolutional layers, three pooling layers, and two fully connected layers. In evaluation experiments conducted, the CNN was trained using our original database with a graphics processing unit CT images were first cropped and resampled to obtain images with a resolution of 256 × 256 pixels and, to prevent over fitting, collected images were augmented via rotation, flipping, and filtering.

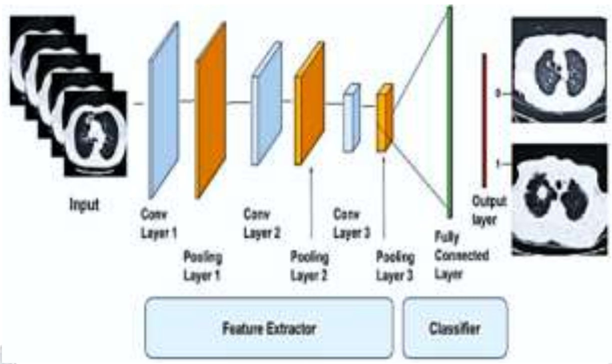

Fig. 2: Steps in CNN classifier

The Convolution layer is always the first. The image (matrix with pixel values) is entered into it. Imagine that the reading of the input matrix begins at the top left of the image. Next, the software selects a smaller matrix there, which is called a filter (or neuron, or core). Then the filter produces convolution, i.e. moves along the input image. The filter‘s task is to multiply its values by the original pixel values. All these multiplications are summed up. One number is obtained in the end. Since the filter has read the image only in the upper left corner, it moves further and further right by 1 unit performing a similar operation. After passing the filter across all positions, a matrix is obtained, but smaller than an input matrix.

V. PROPOSED METHODOLOGY

DenseNet is modern architecture of CNN for visual object recognition that has acquired state-of-the-art with fewer parameters. With some principal modifications, DenseNet is very similar to ResNet. DenseNet, along with its concatenated (.) attributes, combines the previous layer output with a future layer, while ResNet uses an additive attribute (+) to merges the previous layer with the future layers.

The DenseNet Architecture aims to fix this problem by densely connecting all layers. Among the different DenseNet (DenseNet-121, DenseNet-160, DenseNet-201), this study employed DenseNet-121 [5 + (6 + 12 + 24 + 16) × 2) = 121] architecture.

Details of the DenseNet-121 is following: 5—convolution and pooling layers, 3—transition layers (6,12,24), 1—Classification layer (16) and 2—denseblock (1 × 1 and 3 × 3 conv).

Dataset was divided into two datasets (70%/30%, training/testing) to avoid any bias in training and testing. Of the data, 70% was used to train the ML model, and the remaining 30% was used for testing the performance of the proposed activity classification system.

Precision provides a measure of how accurate your model is in predicting the actual positives out of the total positives predicted by your system.

Recall provides the number of actual positives captured by our model by classifying these as true positive.

F-measure can provide a balance between precision and recall, and it is preferred over accuracy where data is unbalanced.

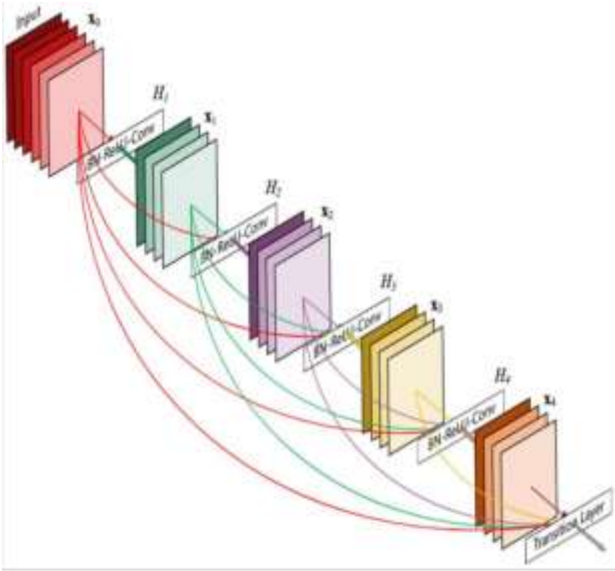

Fig. 3: DenseNet Architecture

Step 1	Importing the libraries and packages
Step 2	Initializing the parameters:
Step 3	Reading the path of input files and initialize the output data
Step 4	Pre-processing the lung disease data for giving them as the input to the model
Step 5	Separating the train data and test data
Step 6	Defining a model and its respective learning model
Step 7	Compiling the model
Step 8	Fitting the data into the compiled model, i.e., training the model using the initially defined parameters
Step 9	Print the Classification Report, Confusion Matrix of the training process and Accuracy

VI. SIMULATION RESULTS

Steps as follow

- We use covid 19 x-ray image dataset with COVID and NON-COVID classes, covid class consisted 1252 images and non covid has 1229 images.
- Preprocessing
- For preprocessing
- Log transformation

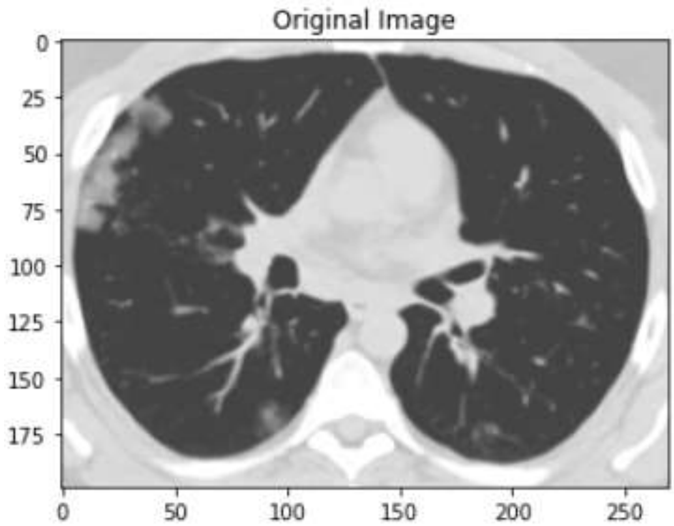

Fig. 4: Input Image

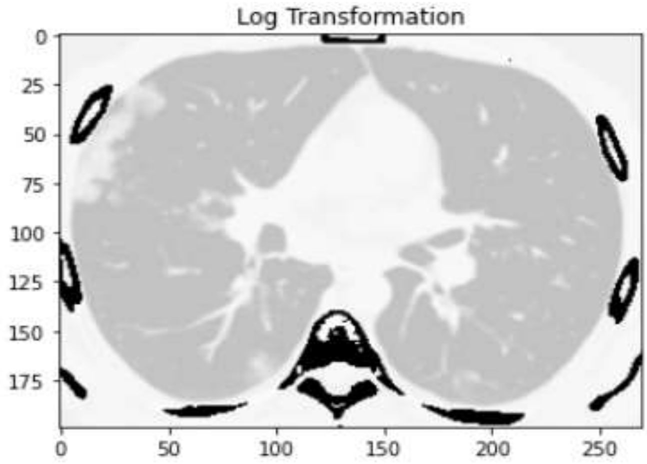

Fig. 5: Log Transformation

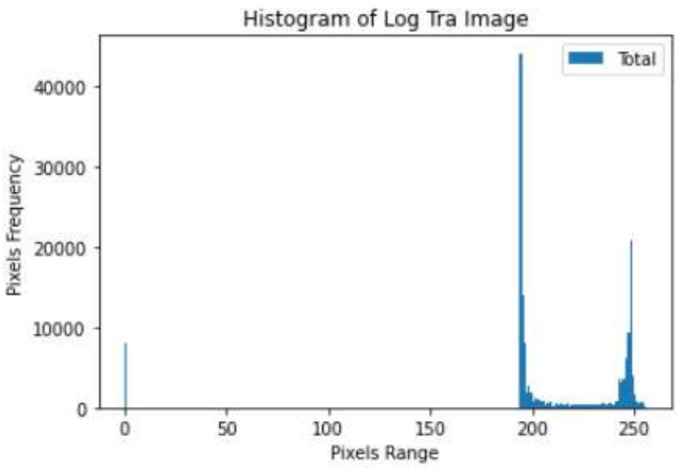

Fig. 6: Histogram of Log Tra Image

- Creating data frame of classes and randomize all images with their path
- Plot frequency Histogram of each class
-

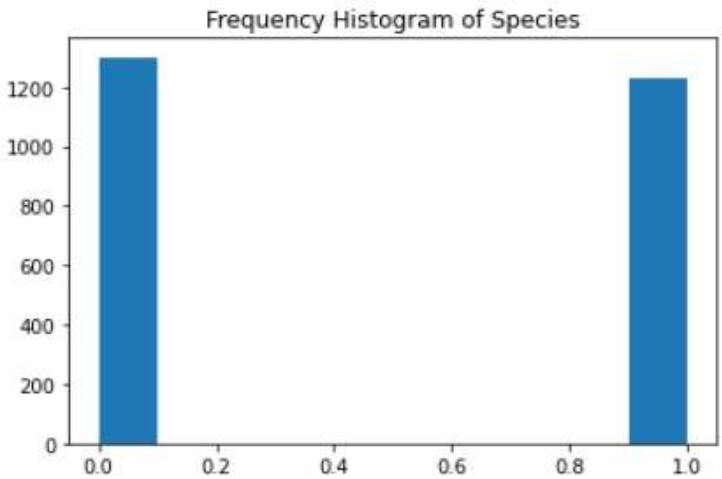

Fig. 7: Frequency Histogram

- Resize and reshape images into 128*128 height and width for training and testing
- Preparing data for training, normalize images using pixel division technique (./255) and convert it into labels to categorical.
- Splitting data images into 80/20 ratio for training and testing.
- Modeling
- For modeling we use DenseNet121 and resnet101 for training preprocess data which was generated using Log Transformation.

Table 1: Image Data Generator

Rotation range	360
Width shift range	0.2
Height shift range	0.2
Zoom range	0.2
Horizontal Flip	True
Vertical Flip	True

Table 2: Build DenseNet121 and ResNet101 with transfer learning

Model	DenseNet121, ResNet101
Weights	ImageNet
Shape	128*128*3
Layers	Conv2D, Dropout
Pooling	Global Average Pooling
Normalization	Batch normalization
Activation	Relu
Output Function	Softmax
Optimizer	ADAM
Learning rate	0.002
Loss	Categorical cross entropy

Table 3: Log Transformation

Model	**Train Accuracy**	**Test Accuracy**	**Loss**	**Precision**	**Recall**	**F1 Score**
ResNet101	**90.61**	**92.09**	**23.46**	**91**	**94**	**92**
DenseNet121	**95.10**	**93.08**	**12.16**	**96**	**89**	**93**

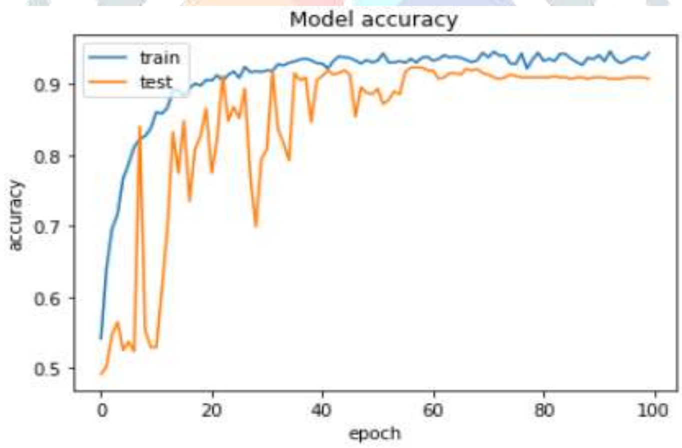

Fig. 8: Model Accuracy

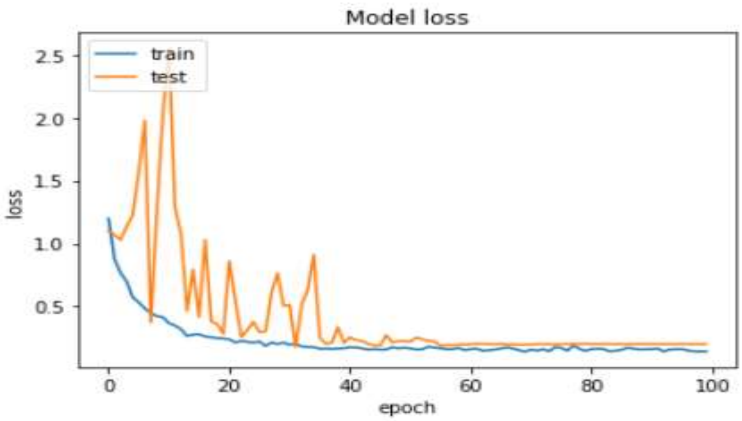

Fig. 8: Model Loss

VII. CONCLUSION

Worldwide, many people are affected by lung cancer. Overall, the chance that a man will develop lung cancer in his lifetime is about 1 in 15; for a woman, the risk is about 1 in 17. These numbers include both smokers and non-smokers. For smokers, the risk is much higher, while for non-smokers the risk is lower. Medical image processing methods and advanced detection algorithms are used to detect the abnormalities present in the lung. With the help of medical image processing, lung cancer can be detected at earlier stages and it increases the immortality rate of the affected persons. In the review system, various types of image processing techniques and classification algorithms are applied for examining and detecting the tumor cells present in the lung.

REFERENCES

[1] Shimpy Goyal and Rajiv Singh, "Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques", Journal of Ambient Intelligence and Humanized Computing, Springer 2021.

[2] Bharati S, Podder P, Mondal R, Mahmood A, Raihan-Al-Masud M. Comparativeperformance analysis of different classification algorithm for the purpose ofprediction of lung cancer. Advances in intelligent systems and computing, vol. 941.Springer; 2020. p. 447–57.

[3] Bhandary Abhir, et al. Deep-learning framework to detect lung abnormality – astudy with chest X-Ray and lung CT scan images. Pattern Recogn Lett January2020;129:271–8.

[4] Rousan LA, Elobeid E, Karrar M, Khader Y (2020) Chest x-ray findings and temporal lung changes in patients with COVID-19 pneumonia. BMC Pulm Med 20:245.

[5] Asuntha A, Srinivasan A (2020) Deep learning for lung cancer detection and classification. Multimed Tools Appl.

[6] Singh D, Kumar V, Yadav V, Kaur M (2020) Deep convolutional neural networks based classification model for COVID-19 infected patients using chest X-ray images. Int J Patt Recogn Artif Intell.

[7] Bharati S, Podder P, Paul PK. Lung cancer recognition and prediction according torandom forest ensemble and RUSBoost algorithm using LIDC data. Int J HybridIntell Syst 2019;15(2):91–100.

[8] Liang C-H, et al. Identifying pulmonary nodules or masses on chest radiographyusing deep learning: external validation and strategies to improve clinical practice.Clin Radiol 2019.

[9] Chouhan V, et al. A novel transfer learning based approach for pneumoniadetection in chest X-ray images. Appl Sci 2020;10(2):559. https://doi.org/10.3390/app10020559.

[10] Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutationprediction from non–small cell lung cancer histopathology images using deeplearning. Nat Med 2018;24:1559–67.

[11] Mondal MRH, Bharati S, Podder P, Podder P. "Data analytics for novel coronavirusdisease", informatics in medicine unlocked, 20. Elsevier; 2020. p. 100374.

[12] Kuan K, Ravaut M, Manek G, Chen H, Lin J, Nazir B, Chen C, Howe TC, Zeng Z,Chandrasekhar V. Deep learning for lung cancer detection: tackling the Kaggle datascience bowl 2017 challenge.

[13] Sun W, Zheng B, Qian W. Automatic feature learning using multichannel ROI basedon deep structured algorithms for computerized lung cancer diagnosis. ComputBiol Med 2017;89:530–9

[14] Song Q, Zhao L, Luo X, Dou X. Using deep learning for classification of lung noduleson computed tomography images. Journal of healthcare engineering 2017:8314740.

[15] Sun W, Zheng B, Qian W. Computer aided lung cancer diagnosis with deep learningalgorithms. In: Proc SPIE. Medical Imaging, 2016, 9785. Computer-AidedDiagnosis; 2016. 97850Z.