

# Classification and Stage Prediction of Lung Cancer using Convolutional Neural Networks

K. Narmada, G. Prabakaran, S. Mohan

**Abstract:** *In recent years, digital image processing is widely used for the medical treatment classification and diagnosis. Lung cancer is the most leading cause of death in all over the world nowadays. Based on the signs and symptoms it can't be diagnosis and treatment classified at the early stage. However it can be identified through the symptoms like coughing up blood and chest pain, the stages and risk factors of the cancer cannot be identified through the symptoms. The CT scanned lung images should be involved in image classification processing for earlier prediction of stages and treatment diagnosis. In existing, machine learning treatment classification can be done through the SVM classification. In case of large set of training samples, this will not be in accurate manner and it has less accuracy because of improper feature extraction techniques. Thus the performance of the classification based on the segmented features obtained in preceding sections. The extracted fine-grained training data through deep learning are utilized for the classification using Convolution Neural Network (CNN). In this paper, we propose a novel framework to classify both small cell and large cell lung cancer and predict its type and treatment using CNN. It is also concentrates on the preprocessing and segmentation processes to accomplish the accuracy in prediction. The experiment results in Python - TensorFlow with Kaggle image dataset show that compared to state of the art of classification and prediction methods, the proposed scheme can obtain much higher accuracy in type prediction and treatment diagnosis.*

**Key terms:** SVM classification, Convolution Neural Network, Deep learning

## I. INTRODUCTION

In this medical world, the early classification and prediction plays a vital role in the medical diagnosis and treatment prediction based on the complexity. Even, the complexities and also the stages of the cancer cells are also predicted nowadays through the implementation of the machine learning and deep learning techniques. The cancer is the most treacherous disease in the medical field, because it should be detected at earlier stage to reduce the complexity of the diagnosis and treatment. In this work, it is concentrated on the most vital lung cancer diagnosis and treatment classification through the deep learning techniques.

Here the machine learning techniques [4] [5] like Naïve Bayes and Support Vector Machine classifications are widely used for the implementation of the stage classification

**Revised Manuscript Received on August 05, 2019**

**K. Narmada**, Research Scholar, Department of Computer Science and Engineering, Annamalai University.

**Dr. G. Prabakaran**, Associate Professor, Department of Computer Science and Engineering, Annamalai University.

**Dr. S. Mohan**, Assistant Professor, Department of Computer Science and Engineering, Annamalai University.

and treatment prediction. The accuracy of the results obtained through the machine learning techniques are need to be improve for further technology enhancements. The Convolutional Neural Networks can be used for the implementation of the deep learning classification and stage identification. In this proposed work, the lung cancer identification, classification of the disease and the stage prediction can be performed through the implementation of the Convolutional Neural Networks with the wide range of training samples. Thus the efficiency of the proposed model can be identified through the confusion matrix and accuracy based on implementations.

## II. BACKGROUND KNOWLEDGE

The some of the domain knowledge should be required to obtain the details and description of the problem and the problem solving methodologies. This can be explained in brief as following.

### A. Lung Cancer Detection and Prediction

Lung cancer is a type of cancer that begins in the lungs. Your lungs are two spongy organs in your chest that take in oxygen when you inhale and release carbon dioxide when you exhale. Lung cancer is the leading cause of cancer deaths in the United States, among both men and women. Lung cancer claims more lives each year than do colon, prostate, ovarian and breast cancers combined.

People who smoke have the greatest risk of lung cancer, though lung cancer can also occur in people who have never smoked. The risk of lung cancer increases with the length of time and number of cigarettes you've smoked. If you quit smoking, even after smoking for many years, you can significantly reduce your chances of developing lung cancer. Early detection of this cancer is most important. Even the manual diagnosis of the disease and the stage identification are too tedious and risky. The image segmentation like mean shift, machine learning like Naïve Bayes and SVM classification, deep learning like ANN and CNN are used to perform the early detection of lung cancer using digital CT images.

### B. Convolutional Neural Networks

A Convolutional Neural Network (CNN/ConvNet) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a

ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area. This can be achieved through the three different layers. They are Convolution Layer, Pooling/ Sub-Sampling Layer, Fully Connected Layer and Soft-Max Layer. This can be achieved through the two different steps. Initially the feature learning can be performed in the convolution and pooling layers and the most important classification can be performed in the fully connected and soft max layer.

### C.Related works

In existing, F Thaher and R Sammouda [1] have proposed the lung cancer detection through the segmentation based technique such as Artificial Neural Networks and Fuzzy clustering methods. It has been performed through the Computer Aided Diagnosis (CAD). Even the machine learning techniques like SVM classifier [2] [14], the lung cancer can be performed based on the classification on SVM and clustering techniques. As like as Werghi N et al [3] have proposed the lung cancer detection and segmentation using the Bayesian Classification and mean shift segmentation algorithms used for clearly identify the sputum cells.

Even a novel algorithms have also proposed based on the machine learning techniques like [7] [8], the authors have designed and developed their own novel algorithms for the implementations of the lung cancer detection. In [9] [10], the authors Tefti et al and Iyer .A et al have designed and developed the deep learning techniques to improve the accuracy of the lung cancer detection rather than the existing.

In [11], Cengil E and Cinar A have proposed the lung cancer identification through the deep learning for classification and also it can be implemented using the tensorflow and 3D CNN architecture of deep learning classification and prediction of stages. Vas M and Dessai A [12] performed the image processing for lung cancer detection using mathematical morphological operations for segmentation of lung region of interest. Potghan et al [13] have been proposed a methodology to identify the lung cancer with tumorous or non tumorous. The improvements can be done through the innovative implementations to the segmentation by threshold and k means clustering. Even the segmentation can be performed in this scheme for two different ways that is for lung volumes and lung nodules. Then the identified features can be extracted through the Gray Level Co occurrence matrix. Thus the classification can be performed to identify the lung nodules through the K-Nearest Neighbor and Multi Layer Perception algorithm with 98.30 and 98.31% of accuracy. Alam. J [15] et al have proposed an efficient lung cancer detection and prediction methodology with the modified multi-class SVM classification technique. As like this proposed mechanism, in

every stage of classification, image enhancement and segmentation have been done separately.

### IV. PROPOSED SYSTEM MODEL

The proposed model architecture has been illustrated in the figure 4.1 with this layer details of Convolutional Neural Networks. Even it only represents the part of implementation of CNN itself. The input image should be preprocessed to obtain accurate results. The input image can be scaled to the 500 x 500 pixels in size and the image should be applied to Gaussian or Gradient filter to make de-noised image for the implementation.

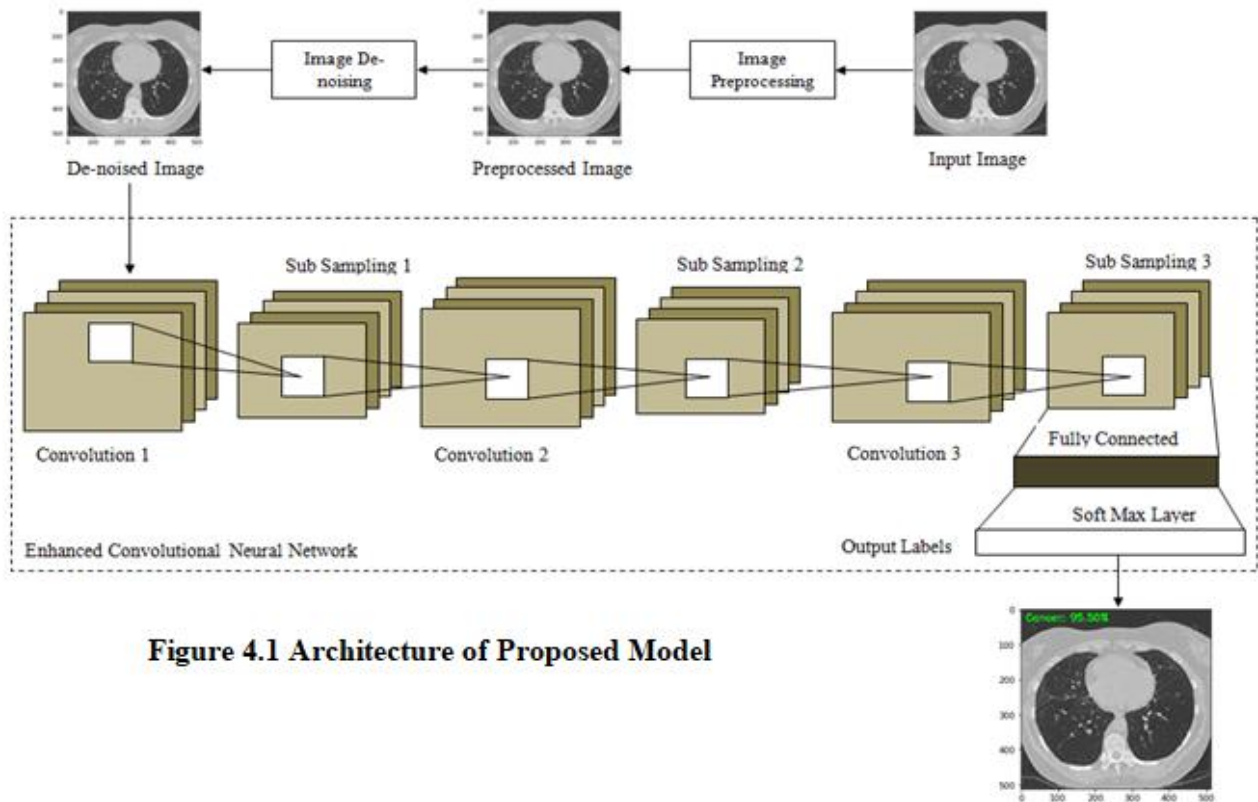
The convolution neural network can be applied to the input image as layer by layer. The image can be initially applied to the convolution layer and sub sampling layer for obtain the feature learning. After the identification of features, the classification can be performed in the fully connected and soft max layer of this CNN architecture. The output labels can be obtained from the implementation of soft max layer.

The input CT lung image can be deployed to the initial preprocessing phase to obtain the preprocessed image. Here the image can be scaled up to the range of pixels. Thus the image needs to identify and remove the noises using the Gaussian or Gradient Filter. This can be chooses under the pre-checking of the quality of the CT images.

The de-noised image can be applied to the feature extraction phase of CNN. Here the features can be extracted through the convolution and pooling layer of this mechanism. The iteration goes up to three times to obtain the creamy level of features and then it sent to the fully connected and soft max layer to get the output labels. This output labels displays the cancer identification with percentage and also not recognized output labels. The pseudo code for this proposed model can be explained in detail for the implementation and easy understanding.

#### Pseudo code for this proposed model:

- Step 1: Obtain the input CT lung image from the user
- Step 2: Preprocess the image to get the appropriate scale 500 x 500 pixels
- Step 3: Apply Gaussian/ Gradient filter to obtain denoised image for the preprocessed image
- Step 4: Apply Convolution Neural Network to obtain the resultsStart initial iteration of Convolution and Pooling

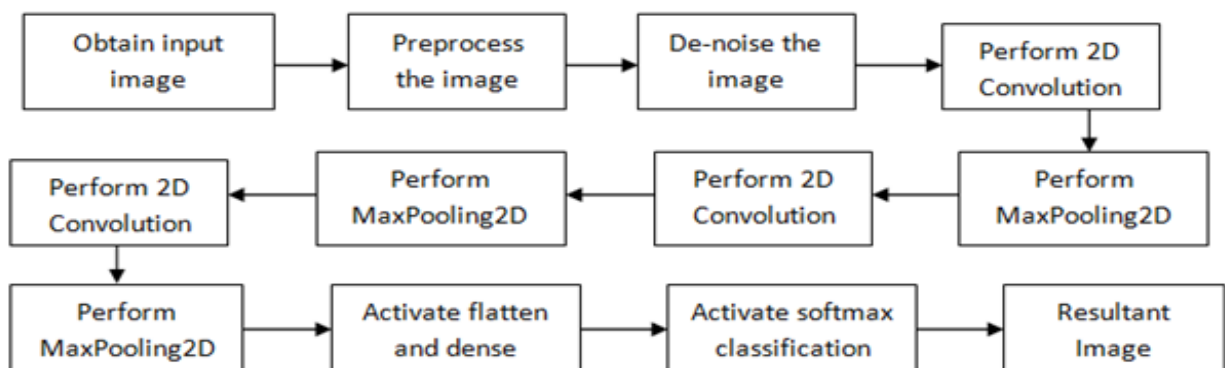


**Figure 4.1 Architecture of Proposed Model**

- a. Perform convolution using conv2D with pixels, padding and input\_shape parameters
- b. Activate the relu layers
- c. Perform sub sampling using MaxPooling2D with poolSize and Strides
- d. Perform convolution using conv2D with pixels and padding parameters
- e. Activate the relu layers
- f. Perform sub sampling using MaxPooling2D with poolSize and Strides
- g. Perform convolution using conv2D with pixels and padding parameters
- h. Activate the relu layers
- i. Perform sub sampling using MaxPooling2D with poolSize and Strides

Start second iteration of Convolution and Pooling

Step 5: Perform first set of fully connected layer using Flatten and Dense functions



**Figure 4.2 Flow of proposed model**

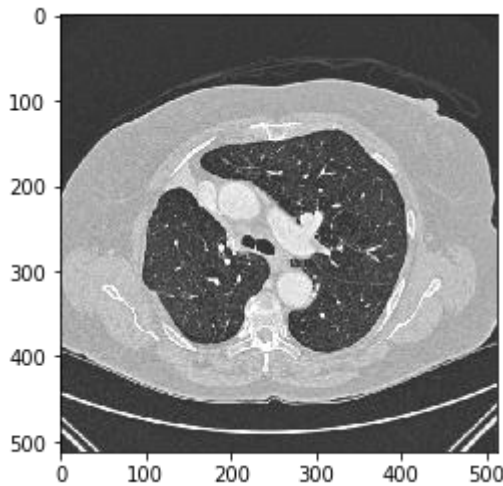
- d. Perform convolution using conv2D with pixels and padding parameters
  - e. Activate the relu layers
  - f. Perform sub sampling using MaxPooling2D with poolSize and Strides
  - Step 6: Perform softmax classification using dense and activation of softmax function
  - Step 7: Display the resultant image with the stage classification
- This proposed model can be graphically designed to describe the implementation flow of the work as in figure 4.2.

Start third iteration of Convolution and Pooling

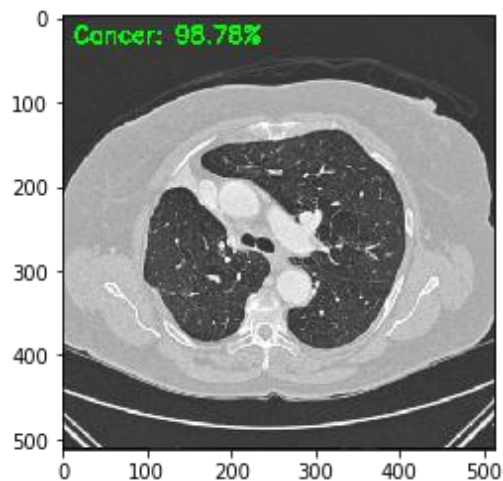
This model can be experimentally verified in the Python environment to get the detection results.

**V. IMPLEMENTATION RESULTS**

Here the deep learning can be performed in the Python TensorFlow environment. It can be obtained using the keras.models classes from the TensorFlow environment.



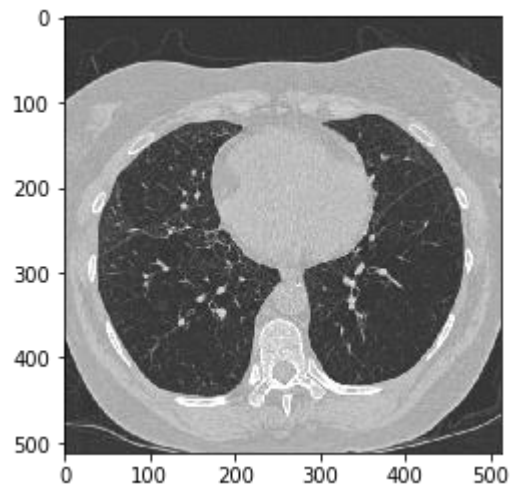
**Figure 5.1 Input image\_1**



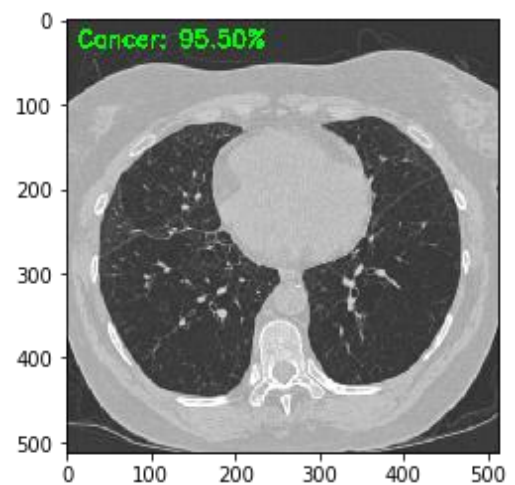
**Figure 5.2 Obtained output\_1**

The Kaggle dataset can be collected and utilized for the implementation of the proposed detection model. Python Syntax: Using TensorFlow backend;

The above two results that show the positive results of the cancer identification and it also shows the how far the lung cancer cells can be spread over the image in percentage. In the figure 5.5 and figure 5.6, shows the input and the resultant images with the negative sign of the lung cancer and it shows the result as not recognized.

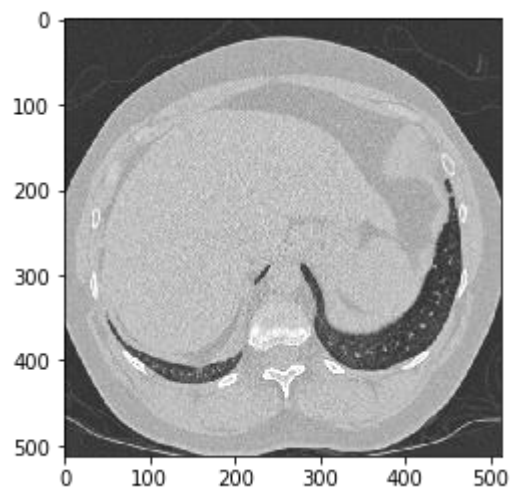


**Figure 5.3 Input image\_2**



**Figure 5.4 Obtained output\_2**

The accuracy can be obtained with the comparison of training samples.



**Fig.5.5 Input image\_3**

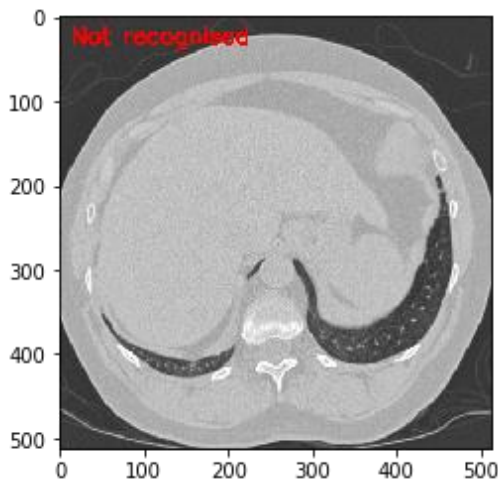


Fig.5.6 Obtained output\_3

The confusion matrix can also be obtained through the results as from predicted labels and true labels of cancer cells and it has been illustrated in the figure 6. Based on the training samples, the number of test samples were tested to obtain this confusion matrix with true positives (TP), true negatives (TN), false positives (FP) and false negative (FN). Let us calculate the precision, recall, true positive rate and false positive rate at the threshold of 0.5. Then we make the confusion matrix has illustrated in fig.7.

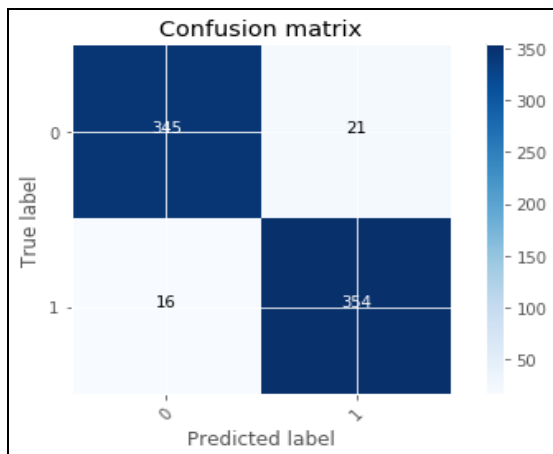


Figure 5.7 Confusion Matrix

The performance metrics such as precision, recall and f1-score can be obtained through the following notations based on the results obtained through the confusion matrix.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FN}) = 345 / (354 + 21) = 0.92$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FP}) = 345 / (354 + 16) = 0.93$$

$$\text{F-Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = 0.91$$

The accuracy for this proposed mechanism can be obtained through the below formula using the values through the confusion matrix.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TN} + \text{TP} + \text{FP} + \text{FN})$$

$$= (345 + 354) / (345 + 354 + 21 + 16)$$

$$= 699 / 736$$

$$= 94.97$$

Accuracy of this proposed mechanism is 94.97% obtained through the identified matrix has illustrated in the fig.5.8. Thus the accuracy can be varied from the usual CNN model used in the exiting works.

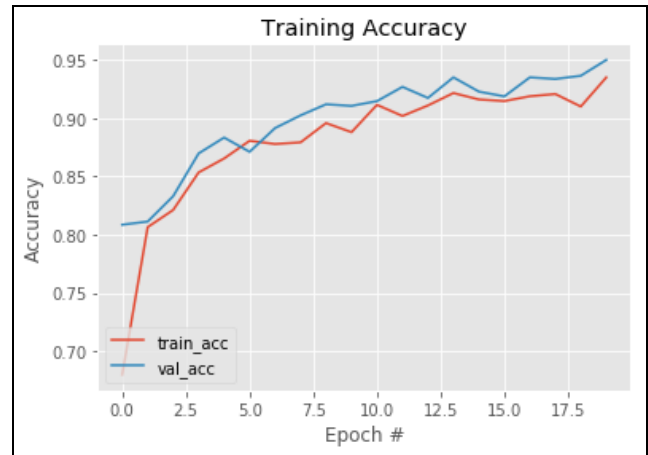


Figure 5.8 Performance graph based on accuracy

From these results, the obtained results can be improved through the modification in the feature extraction obtained from the convolution and pooling layer of the CNN model.

## VI. CONCLUSION

Thus the stage identification of the Lung cancer can be performed through the effective implementation of the proposed mechanism using Convolution Neural Networks. In this proposed works, four layers of enhancements may helps to perform the deep learning based classification. Thus the layers are convolution layer, sub sampling layer / pooling layer, fully connected layer and soft max layer. Initial two layers are iterated three times to obtain the accurate results from the model. The predicted results can obtained in average of 95% accuracy and it should comparatively good rather than the existing works. In future, it is to improve through the enhancements in the security of the training samples to obtain the more accurate solutions for the classification and prediction.

## REFERENCES

1. Taher F., & Sammouda R., "Lung cancer detection by using artificial neural network and fuzzy clustering methods", 2011 IEEE GCC Conference and Exhibition (GCC), 2011, Arab Emirates, pp.294-298.
2. Cabrera J., Dionisio A., & Solano G., "Lung cancer classification tool using microarray data and support vector machines", 2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA), 2015, Greece.
3. Werghi N., Donner C., Taher F., & Al-Ahmad H., "Detection and segmentation of sputum cell for early lung cancer detection", 19th IEEE International Conference on Image Processing, 2012, USA, pp.2813-2819.
4. Kanitkar S. S., Thombare N. D., & Lokhande S. S., "Detection of lung cancer using marker-controlled watershed transform", 2015 International Conference on Pervasive Computing (ICPC), 2015, Pune.
5. Punithavathy K., Ramya M. M., & Poobal S., "Analysis of statistical texture features for automatic lung cancer detection in PET/CT images", 2015 International Conference on Robotics, Automation, Control and Embedded Systems (RACE), 2015, India.
6. Riti Y. F., Nugroho H. A., Wibirama S., Windarta B., & Choridah L., "Feature extraction for lesion margin characteristic classification from CT Scan lungs image", 2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2016, Indonesia, pp.54-58.

7. Dass M. V., Rasheed M. A., & Ali M. M., "Classification of lung cancer subtypes by data mining technique", Proceedings of The 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC), 2014, India, pp.562-568.
8. Kulkarni A., & Panditrao A., "Classification of lung cancer stages on CT scan images using image processing", 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, 2014, India, pp.1384-1388.
9. Tafti A. P., Bashiri F. S., La Rose E., & Peissig P., "Diagnostic Classification of Lung CT Images Using Deep 3D Multi-Scale Convolutional Neural Network", 2018 IEEE International Conference on Healthcare Informatics (ICHI), 2018, USA, pp.412-418.
10. Iyer A., Vyshnavi A M H., & Namboori P K K., "Deep Convolution Network Based Prediction Model for Medical Diagnosis of Lung Cancer - A Deep Pharmacogenomic Approach : deep diagnosis for lung cancer", 2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAIECC), 2018, India.
11. CENGIL E., & CINAR A., "A Deep Learning Based Approach to Lung Cancer Identification", 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018, Turkey.
12. Vas M., & Dessai A., "Lung cancer detection system using lung CT image processing", 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), 2017, India.
13. Potghan S., Rajameenakshi R., & Bhise A., "Multi-Layer Perception Based Lung Tumor Classification", 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, India, pp.492-502.
14. Taher F., Werghi N., & Al-Ahmad H., "Rule based classification of sputum images for early lung cancer detection", 2015 IEEE International Conference on Electronics, Circuits, and Systems (ICECS), 2015, Egypt, pp.29-32.
15. Alam J., Alam S., & Hossan A., "Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classification", 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018, Bangladesh.