ASIAN JOURNAL OF
INFORMATION TECHNOLOGY
Journal Home Page: https://medwelljournals.com/journalhome.php?jid=1682-3915

# Lung Cancer Detection and Classification using Convolutional Neural Network

Deepa Sajjanar, B.S. Rekha and G.N. Srinivasan
*Department of Industrial System Engineering, Rashtreeya Vidyalaya College of Engineering, Bengaluru, India*

**Key words:** Computed tomography, Convolutional Neural Network (CNN), computer-aided diagnosis, CT scan, pooling layer

**Corresponding Author:**
Deepa Sajjanar
*Department of Industrial System Engineering, Rashtreeya Vidyalaya College of Engineering, Bengaluru, India*

Page No.: 233-240
Volume: 18, Issue 10, 2019
ISSN: 1682-3915
Asian Journal of Information Technology
Copy Right: Medwell Publications

**Abstract:** In the field of medicine, identification and treatment of cancer is considered as one of the biggest challenge in treatment of chronic illness. The survival of patients depends on timely detection and cure. Experts use the CT scan or computed tomography scan images of patients to detect and classify nodules, before proceeding with advanced treatment procedures. The present day advances in artificial intelligence, machine learning based on deep learning model can be used to develop sophisticated computer aided diagnosis systems to detect cancerous nodules. The proposed system is based on convolutional neural networks to categorize nodules detected in CT scan images as malignant or benign. Image processing and neural networks have been extensively used in the detection and classification of cancerous nodules. Hence, CNNs are more appropriate, for the task of nodule detection and classification. CNN's have more properties like multiple feature extraction. When convolution layer, subsampling or pooling layer, fully connected layer such layers are combined, leading to deep CNNs, it helps in increasing the accuracy of classification. The proposed CNN Model will be suitable for the early detection and classification of CT scans images containing nodules with accuracy of 93.52% using the domain knowledge of the CT scan images of lung in the field of medicine and neural network.

## INTRODUCTION

Nowaday's lung cancer has become one of the leading causes of cancer related deaths of humans. So, it is necessary for the trained radiologists to identify the cancer accurately in the lung in the early stage as possible to reduce the deaths of humans (Ramaswamy and Truong, 2006). So, this is very complicated job to detect the lung nodules which is affected by cancer and non-cancer in its early stages.

In the past days, radiologist were manually analyzing the CT scans images of lung, looking for the potential nodules and identify the cancerous and non-cancerous in those nodules. This process needs a high knowledge about the lung nodules which is very tedious and time consuming. So, this can resolved by introducing the Computer Aided Diagnosis (CAD) system to detect the lung nodules and classification of nodules as either cancerous or non-cancerous. This system will be as second opinion to the radiologist to detect and analyze the nodules of the lung.

In US, lung cancer has become the cause for the cancer related deaths. Approximately 229,447 new cases of lung cancer were there and in that 159,124 related deaths. So, early diagnosis of lung cancer can improve the effectiveness of treatment and increase the chance of the survival of patient's. The lung cancer can be identified by the noninvasive imaging modalities those are Computed Tomography (CT), Contrast-Enhanced Computed Tomography (CE-CT), Low-Dose Computed Tomography (LDCT) and Positron Emission Tomography (PET) (El-Baz *et al*., 2013).

In the past years, CAD system have been developed for both the nodule segmentation and also classification of lung nodules as cancerous or non-cancerous. But this system for segmentation will generate a very good detection of lung nodules but however in the process they lead to many false positive, when actual positive compared to this false positive segmentation the ratio can be in hundreds (El-Baz *et al*., 2013). So, this system can lead many false positive when dealing with lung cancer. It's better to label anything that looks like nodule in the image and the radiologists need to fallow the manual methods of detecting the lung cancer.

Traditional algorithms of image processing used to detect unique features of images. So, this requires the hand crafted features has to be created which learns the features at manual process. So, in this method it is very complex to differentiate the features of cancerous and non-cancerous nodules, so, the deep learning can avoid all these problems and are capable of tackling with the problems like image recognition, video recognition, speech recognition and natural language processing, etc. (Gu *et al*., 2018).

Manual feature extraction requires expert knowledge of the lung cancer to the designer. Deep learning will have the properties to learn all the features in the images. In particular, Convolutional Neural Network (CNN) extracts the features of input images by using one or more layers of convolution and subsampling or max-pooling layer that are in the hierarchical manner. The general CNN consists of three layers convolution, max-polling and fully connected. The classification of images is done by CNN by extracting features in each layer and producing a final model. When other test image is given to model compares the features of both and classify the images to different classes by the accuracy.

In this study, we have used the CNN for the classification of CT scan images of lung cancer as cancerous or non-cancerous.

**Literature review:** Rao *et al*. (2016) in this study the researchers has designed known CNN system called as "CanNet" for classification of lung nodules. The CanNet is compared with the results of Artificial Neural Network (ANN) and LeNet got by other researchers. The results are shown in graph form for all different iterations. The accuracy of the proposed CanNet is improved by 45% when compared to LeNet and improved by 14% when compared to ANN. So, CanNet outperforms the both ANN and LeNet Architectures accuracy for the classification of lung nodules.

Ramaswamy and Truong (2006) in this study the researchers aims to enhance the existing system of detection of lung nodules in the CT scans. The existing system focuses on algorithms of segmentation and diagnosis through the traditional image processing techniques to detect the lung nodules in CT scans. But they proposed the CAD system by generally using Convolutional Neural Network (CNN). The accuracy got is 89.6% for classification of nodules and reducing the false negative results as compared to existing system.

El-Baz *et al*. (2013) in this study the researcher has explained about the most important and challenging problem in the computer science which is lung cancer detection in the patients and diagnosis it at the early stage of growth. This can be done by developing the CAD (Computer-Aided Diagnosis) system for the early detection of lung cancer and increase the survival of patients for the effect of early detection of lung cancer. So, the creating of CAD system has been investigated in the large number of research studies. The main four steps in the development of lung cancer detection and classification CAD system includes first the segmentation of the lung area fields, second is detection of nodule inside the lung area field, third is segmentation of those detected nodule of the lung and last is diagnosis or classification of the nodules as benign or malignant. So, in this study the researcher overviews current techniques and methods for the development of CAD system steps. For each of the techniques, there are different methodologies, conditions, technical issues, database format for the training and testing and the methods of the verification and validations and also the overall performance. The researcher also addressed the many issues, challenges that are faced by the researches in implementing each step of the development of CAD system and even the author has marked the advantages and disadvantages of existing methodologies of lung cancer system.

Golan *et al*. (2016) in this study the researcher has implemented the CAD system which uses the Lung Image Database Consortium (LIDC) and Image Database Research Initiative (IDRI) database includes 1018 CT scans images with nodules of different shape and size and the deep Convolutional Neural Network (CNN) for the detection of nodules in the CT scans which uses the back propagation algorithm for extraction of features of input scan images. The results were achieved to each test scan image is 78.9% of true positive rate and 20 False Positives (FPs) or the 71.2% of true positive and 10 False Positives (FPs). In this proposed research, the CAD system does not use any segmentation and additional FPs reduction procedures involved.

Yang *et al*. (2016) in this project the researcher has proposed the two architectures namely the DCNN-5 (deep convolutional neural network with 5 hidden layers) and DCNN-7 (Deep Convolutional Neural Network with 7 hidden layers) for the automatic lung tumor classification. These architectures performance is compared with SVM (Support Vector Machine) and ELM

(Extreme Learning Machine) results got by other researchers. The accuracy results of proposed CNN architectures overcomes the SVM and ELM for the classification of tumors of lung to cancerous or non-cancerous.

Rossetto and Zhou (2017) in this study, the researcher has presented an ensemble of Convolutional Neural Network (CNN) using the number of preprocessing methods to increase the accuracy level of the automated labeling of CT scans images. This is done by implementing the CNN's with the voting system to get the consensus of two networks. This method got the accuracy of 97.5% and <10% false positive for classification of tumors.

Gu *et al*. (2018) in this study the researcher made the survey on recent advances of CNN. The researcher has discussed the improvements of CNN in different aspects, namely, activation function, fast computation, loss function optimization and CNN layer design. And also, they discussed the applications of CNN on many tasks like image classification, object detection, object tracking, text detection, scene labeling, speech and natural language processing.

Miah and Yousuf (2015) in this study the researcher has implemented the system to detect the lung cancer in the premature stage. It includes many image processing steps like preprocessing, thresholding, segmentation, features extraction and neural network. Here, multilayer feedforward neural network has been used to classify the lung CT scans as cancerous and non-cancerous. The proposed system results the accuracy of 96.76% for the classification.

Jin *et al*. (2016) in this study, the researcher has proposed the method of segmentation based on morphology process and also Region of Interest (ROI) extraction based on morphology and circular filter to reduce the false positive to increase the accuracy.

Jin *et al*. (2017) in this study the researcher has explored the early lung cancer detection by learning the deep spatial lung features. For this they has designed the 3D CNN architecture by improving the initial 2D architecture of CNN. First, they have segmented the 2D CT scans and visualized that in the 3D form to view the segmented lung volume. The CNN is used to classify the CT scans as cancerous and non-cancerous by training and testing the segmented CT scans.

**Dataset:** For the proposed work the dataset used in the LUng Nodule Analysis (LUNA) which is derived from Lung image Database Consortium and Image Database Research Initiative (LIDC/IDRI) database. Out of 1024 patient data available, 888 patient data is used to excluding the slices of thickness >2.5 mm. The dataset consists of the DICOM format CT scan images where they are showed in MHD and RAW image format. MHD file format is the meta image that is based tagged file format for the medical images and RAW file contains the processed data from the image sensor of either camera, scanner.

The dataset includes the annotations of the CT scans provided in the dataset. These annotations are obtained from the expertized radiologists including all the information (diameter, X, Y, Z co-ordinates of the images) about the CT scans images of lung. The dataset LUNA was the challenge as the part of 2016 IEEE International Symposium on Biomedical Imageing.

## MATERIALS AND METHODS

The proposed research includes following methodology as showed in the below system architecture diagram (Fig. 1). The system is implemented by the following methods:

**Nodule extraction and pre-processing:** The dataset include the information of the nodules in the candidate.csv file which consists of serial nodules of lung image, co-ordinates of the nodules present in the lung image. The LUNA16 dataset:

- Load LUNA16 candidates.csv
- Load lines with class = 1 (Positives )
- Load lines with class = 0 (Negatives)
- Load x, y, z co-ordinates and label
- Split training, validation and testing dataset
- For record in dataset:
    - Extract nodule based on x, y, z co-ordinate
    - Normalize the extracted image to filter unwanted voxels
    - Save nodule in JPEG format

Training and testing in convolutional neural network the nodules are extracted from the annotations given in the dataset by the radiologists and normalize those nodules and save in the JPEG format. These images are given as input to CNN for classification of image as cancerous and non-cancerous.

The pre-processing of images is not explicitly done in CNN because CNN classifier implicitly does all preprocessing of images.

**Training/testing/validation dataset creation:**

- Augment the training images by applying rotation to increase dataset variations and improve overall module accuracy
- Package training, validation, test images into HDF5 files

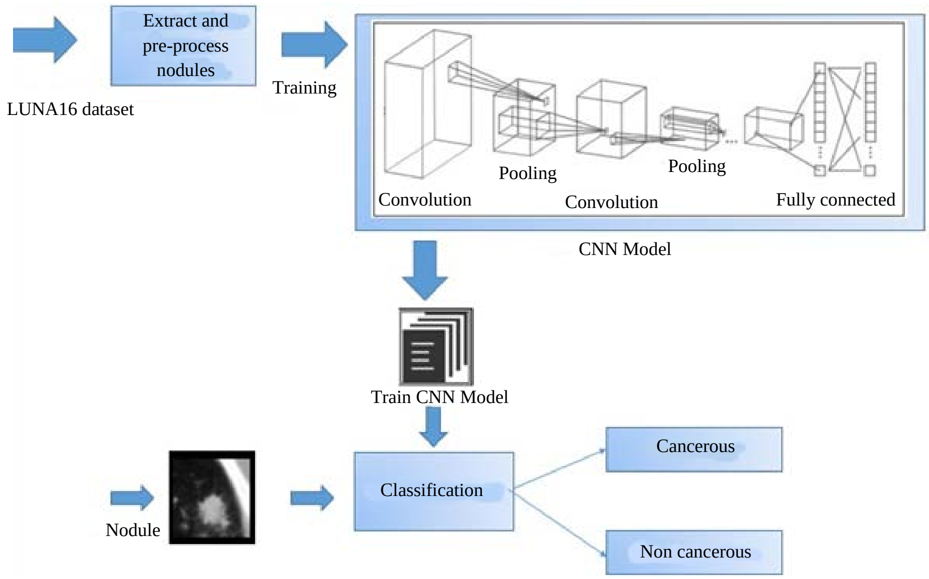


Fig. 1: System architecture of lung cancer detection using convolutional neural network

**Training/validation:**

- Load training and validation dataset from h5 files
- Define tensorflow CNN Model
- Define and initialize tensorflow deep neural network model
- Invoke model fit with training and validation dataset for specific number of epochs
- Save the trained model

**Testing:**

- Select a patient ID
- Retrieve the nodules for selected patient ID
- Load the tensor flow trained CNN Model
- Invoke model.predict to classify the test nodules
- Invoke model.eval to determine the accuracy
- Check the predicted value and accuracy

**Concepts of convolutional neural network:** A typical Convolutional Neural Network (CNN) is the type of neural network used as classifier to make image classification. All the set of dataset images are loaded on to CNN where each layer reads the features of images and forward to next layers.

In the general CNN Model of image recognition, a hand developed feature extractor collect the data or information from the input image and eliminate the irrelevant variables. This extractor is followed by classifier which is to train the feature vectors into classes.

The first layer is convolution layer which plays a rule of feature extractor and second layer is pooling layer where it eliminates some of the negative features that are extracted in previous layer. Next is fully connected layer which combines all the features of previous layer and fix the model. Softmax which compares and eliminates the features of irrelevant variables of the test image.

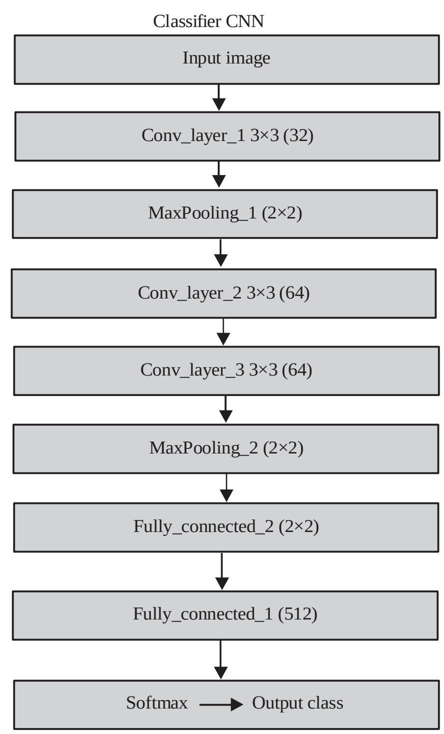


Fig. 2: The proposed CNN architecture for classification

**The proposed CNN Model:** Convolutional neural network has been proposed for the detection and classification of the lung nodules as the malignant or benign (Fig. 2). The proposed CNN Model includes different layers of combination where the image features of one layer are passed to next layers. The next layers extracts the more features and pass those to the next layer and soon. At softmax layer (it is used in final layer of the classifier to compare the features of the previous layer with input image to eliminate the unwanted feature) the features collected are filtered by eliminating the unwanted features and make a trained model (Fig. 2).

The proposed CNN has some conditions which gives the improved accuracy are explained.

**Learning rate:** It is the parameter used in training process of images where it controls weight and bias that changes in learning the training algorithm. The learning for the proposed system is 0.001.

**Epoch:** The value of epoch is determined when the training will stop once the iterations reaches the epoch. When training by minimum error, the represents the maximum number of iterations. Epoch for the proposed system is 70.

**Minimum error:** It is the value of mean square error of the epoch specified. It differentiate the accuracy got for the training and testing.

## RESULTS AND DISCUSSION

The results of lung cancer detection and classification using convolutional neural network system are shown in the graphical user interface that has created.

Figure 3 shows the user interface of giving the test lung image for detection and classification of nodules as cancerous or non-cancerous. And marking the nodule in the lung image by the annotations given by radiologists. This results is obtained for learning rate of 0.001, epoch 70 and 3×3 filter size.

Figure 4 shows the extracted nodule of the one test lung image. This nodule is preprocessed to normalization of extracted nodule image. This results is obtained for learning rate of 0.001, epoch 70 and 3×3 filter size.

Figure 5 shows the other test slice of the particular patient id which has the nodule in the slice will be marked the image. This results is obtained for learning rate of 0.001, epoch 70 and 3×3 filter size.

Figure 6 shows the extracted nodule of the other test lung slice. This nodule is preprocessed to normalization of extracted nodule image. This results is obtained for learning rate of 0.001, epoch 70 and 3×3 filter size.

Figure 7 shows the classification of lung cancer as cancerous where the expected is non-cancerous, so, it false negative results. So, the accuracy for the particular nodule is 0. This results is obtained for learning rate of 0.001, epoch 70 and 3×3 filter size.

Figure 8 shows the classification of lung nodule as non-cancerous where the expected was also non-cancerous. So, the results obtained are correct or accurate with the accuracy of 100% for the particular nodule. This results is obtained for learning rate of 0.001, epoch 70 and 3×3 filter size.

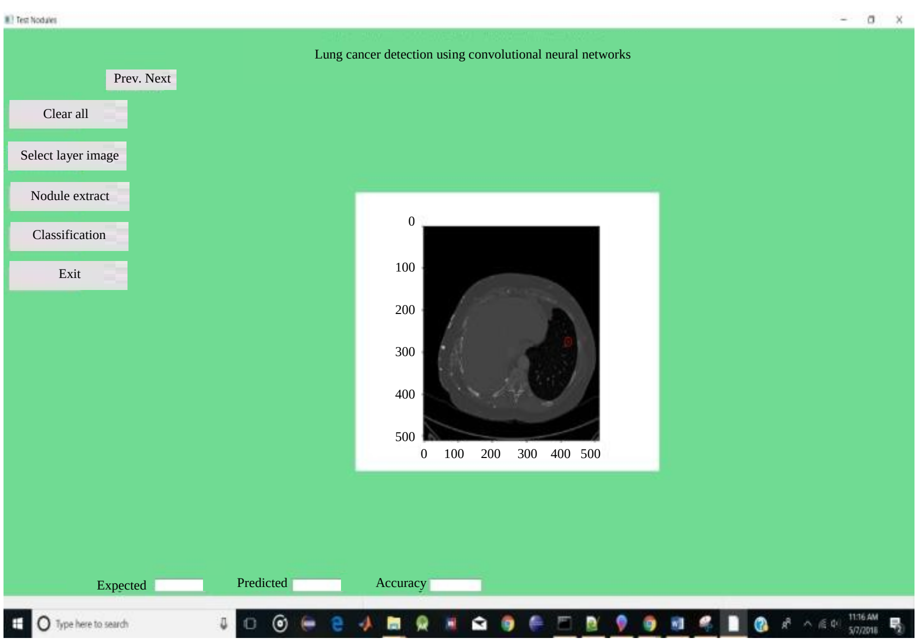


Fig. 3: User interface of one test lung CT scan slice in the particular patient ID

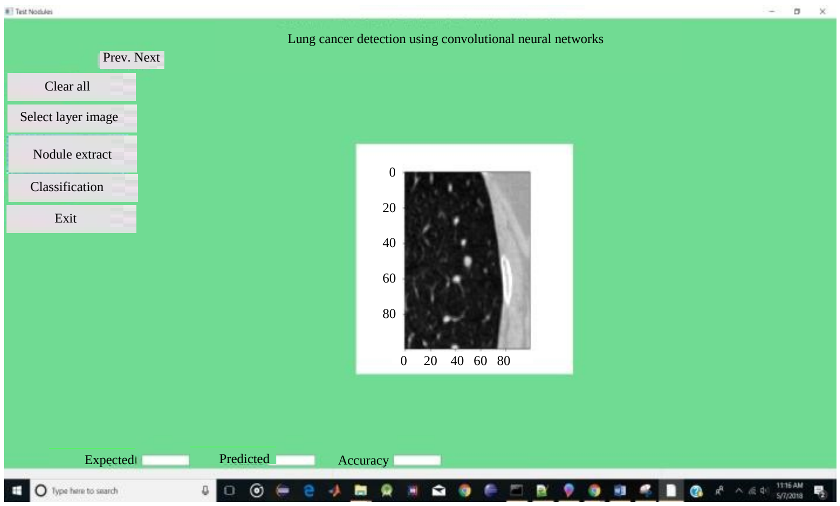


Fig. 4: User interface of one lung slice extracted nodule with normalization

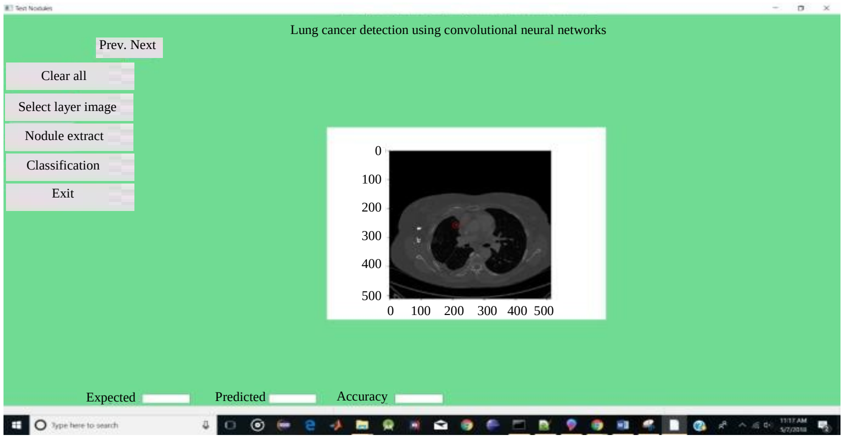


Fig. 5: User interface of next test lung CT scan slice in the particular patient ID

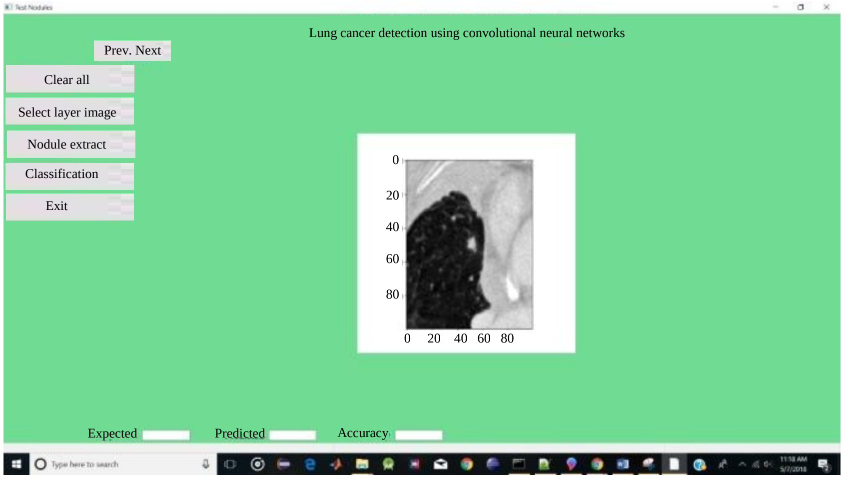


Fig. 6: User interface of other lung slice extracted nodule with normalization

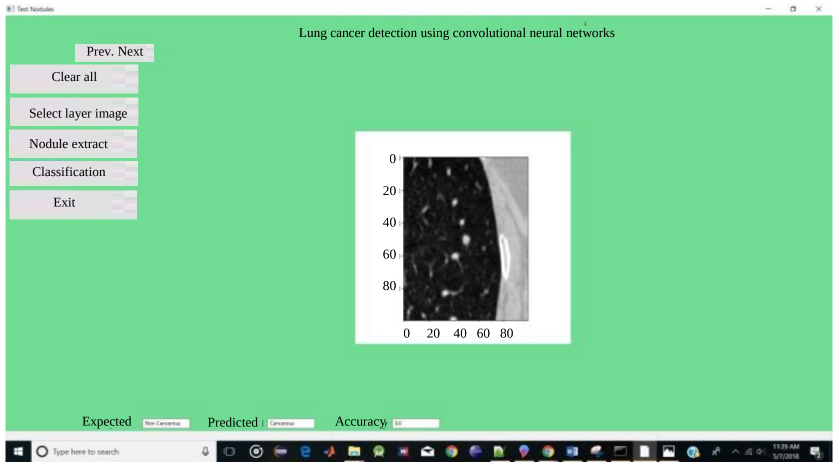


Fig. 7: Nodule classification of one lung slice of CT scan with false negative results

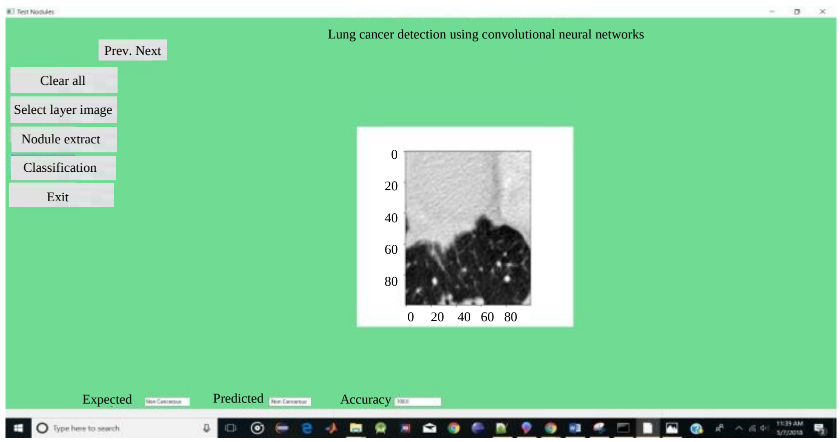


Fig. 8: Nodule classification of one lung slice of CT scan with accurate result

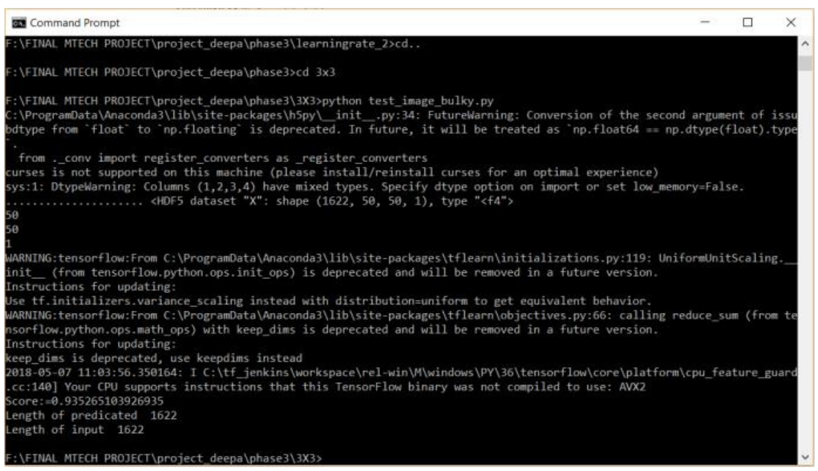


Fig. 9: The accuracy percentage of the 3×3 filter size CNN for all test images

Figure 9 shows the bulk test images (1622) result of accuracy of 93.52% which is the classification accuracy of the test images as per trained network model with learning rate of 0.001, epoch 70 and 3×3 filter size.

The results are analyzed by considering the different modifications of the values such as learning rate, epoch and the filter size. So, based on the results obtained for different values, the accuracy can be changing. By considering filter size 3×3 learning rate 0.001 and epoch 70, the results were better in reaching the accuracy than others.

## CONCLUSION

Based on the work done, a lung cancer can be detected and classified using the neural network. This helps for the doctor to improve treatment in the early stage of cancer and avoid many deaths of patients with the effect of lung cancer detection in early stage. The average percentage accuracy for the proposed system is reached 93.52% for detection and classification of lung cancer using CNN.

## RECOMMENDATIONS

The future enhancement of the proposed work can be done as below:

- The system could be enhanced to classify the CT scans by increasing layers in network to increase the accuracy
- The system can be improved by including different types of dataset (UCI, TCIA, etc.)
- The system can be improved by incorporating 3D lung cancer images for classification

## REFERENCES

El-Baz, A., G.M. Beache, G. Gimel'farb, K. Suzuki and K. Okada *et al*., 2013. Computer-aided diagnosis systems for lung cancer: Challenges and methodologies. Int. J. Biomed. Imaging, Vol. 2013, 10.1155/2013/942353

Golan, R., C. Jacob and J. Denzinger, 2016. Lung nodule detection in CT images using deep convolutional neural networks. Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), July 24-29, 2016, IEEE, Vancouver, Canada, pp: 243-250.

Gu, J., Z. Wang, J. Kuen, L. Ma and A. Shahroudy *et al*., 2018. Recent advances in convolutional neural networks. Pattern Recogn., 77: 354-377.

Jin, T., H. Cui, S. Zeng and X. Wang, 2017. Learning deep spatial lung features by 3D convolutional neural network for early cancer detection. Proceedings of the 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), November 29-December 1, 2017, IEEE, Sydney, Australia, pp: 1-6.

Jin, X.Y., Y.C. Zhang and Q.L. Jin, 2016. Pulmonary nodule detection based on CT images using convolution neural network. Proceedings of the 2016 9th International Symposium on Computational Intelligence and Design (ISCID) Vol. 1, December 10-11, 2016, IEEE, Hangzhou, China, pp: 202-204.

Miah, M.B.A. and M.A. Yousuf, 2015. Detection of lung cancer from CT image using image processing and neural network. Proceedings of the 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), May 21-23, 2015, IEEE, Dhaka, Bangladesh, pp: 1-6.

Ramaswamy, S. and K., Truong, 2016. Pulmonary nodule classification with convolutional neural networks. Comput. Math. Methods Med., Vol. 1.,

Rao, P., N.A. Pereira and R. Srinivasan, 2016. Convolutional neural networks for lung cancer screening in Computed Tomography (CT) scans. Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), December 14-17, 2016, IEEE, Noida, India, pp: 489-493.

Rossetto, A.M. and W. Zhou, 2017. Deep learning for categorization of lung cancer ct images. Proceedings of the 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), July 17-19, 2017, IEEE, Philadelphia, USA., pp: 272-273.

Yang, X., S.Y. Yeo, J.M. Hong, S.T. Wong and W.T. Tang *et al*., 2016. A deep learning approach for tumor tissue image classification. IASTED Biomed. Eng., Vol. 1, 10.2316/P.2016.832-025