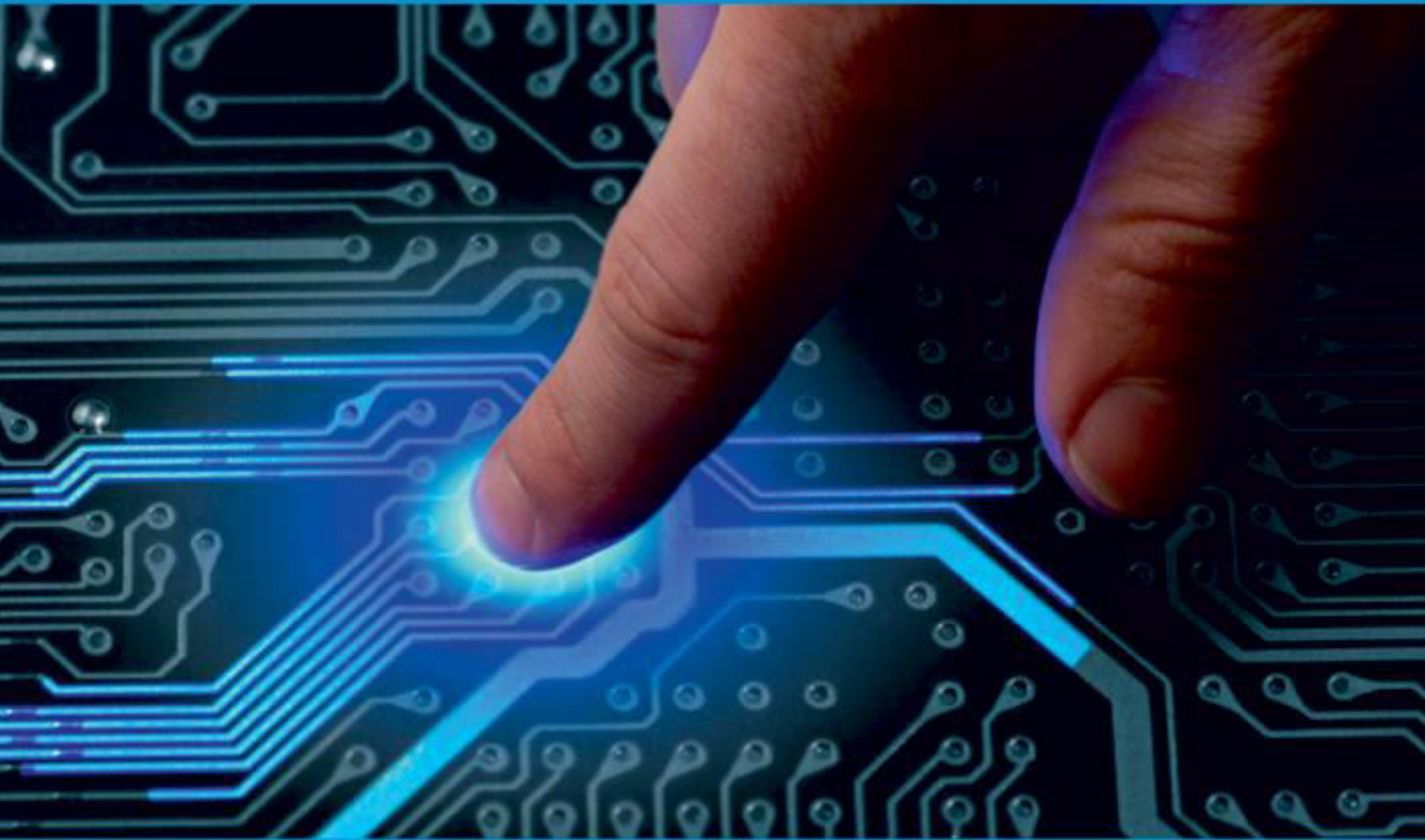
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006199|

Lung Cancer Detection Using Deep Convolution Neural Network

S.Charanjeeth, T.S.Kartick Raj, G.Manikandan M. Mohamed Jasim, P.B.Arun Prasad

UG Students, Department of Computer Science and Engineering, Saranathan College of Engineering, Tiruchirappalli, Tamil Nadu, India.

Assistant Professor, Department of Computer Science and Engineering, Saranathan College of Engineering, Tiruchirappalli, Tamil Nadu, India.

ABSTRACT: Cancer is the root cause for a large number of deaths worldwide, out of which lung cancer is the cause of the highest mortality rates. Computer tomography scan is employed by radiologists to detect cancer in the body and track its growth. Visual interpretation of database can lead to cancer detection at later stages, thus leading to late treatment of cancer which only boosts up the cancer death rates. Therefore, image processing tools can be used for early detection of cancer. In this paper, a lung cancer detection algorithm is proposed using mathematical morphological operations for segmentation of the lung region of interest, from which Haralick features are extracted and used for classification of cancer by artificial neural networks. This work has been done by providing a CT image of Lungs and to identify weather the provided image is a Benign or Malignant cases. If it cannot identified the tumor cells in the provided Dataset Then it is said to be Normal.

I. INTRODUCTION

Lung cancer disease is the second largest death threat to the world after heart attack, as this cancer is responsible for the largest number of deaths, compared to the number of deaths caused by any other cancer type. Lung cancer is the uncontrolled growth of the cells, thus leading to the formation of lung nodules. It is reported that lung cancer is responsible for around 19% deaths globally mostly due to alcohol and tobacco consumption. The rate of survival is assured by only 15% survival chances, for a survival period of 5 years. The main cause of such high death rate is the detection in later stages, thus leading to delayed treatment. Machine learning might be able to provide a broader class of more flexible alternative analysis methods better suited to modern sources of data. It is imperative for statistical agencies to explore the possible use of machine learning techniques to determine whether their future needs might be better met with such techniques than with traditional ones. In the past few decades, Deep Learning has proved to be a very powerful tool because of its ability to handle large amounts of data. The interest to use hidden layers has surpassed traditional techniques, especially in pattern recognition. One of the most popular deep neural networks is Convolutional Neural Networks. In deep learning, a convolutional neural network (CNN/ConvNet) is a class of deep neural networks, most commonly applied to analyze visual imagery. Now when we think of a neural network we think about matrix multiplications but that is not the case with ConvNet. It uses a special technique called Convolution. Now in mathematics convolution is a mathematical operation on two functions that produces a third function that expresses how the shape of one is modified by the other. Despite the power and resource complexity of CNNs, they provide in-depth results. At the root of it all, it is just recognizing patterns and details that are so minute and inconspicuous that it goes unnoticed to the human eye. But when it comes to understanding the contents of an image it fails. Let's take a look at this example. When we pass the below image to a CNN it detects a person in their mid-30s and a child probably around 10 years. But when we look at the same image we start thinking of multiple different scenarios. Maybe it's a father and son day out, a picnic or maybe they are camping. Maybe it is a school ground and the child scored a goal and his dad is happy so he lifts him.

II. RELATED WORK

The goal of this paper is to present a critical review of major Computer-Aided Detection systems (CADe) for lung cancer in order to identify challenges for future research. CADe systems must meet the following requirements: improve the performance of radiologists providing high sensitivity in the diagnosis, a low number of false positives

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006199|

(FP), have high processing speed, present high level of automation, low cost (of implementation, training, support and maintenance), the ability to detect different types and shapes of nodules, and software security assurance. Further research is needed to improve existing systems and propose new solutions. For this, we believe that collaborative efforts through the creation of open source software communities are necessary to develop a CADe system with all the requirements mentioned and with a short development cycle. In addition, future CADe systems should improve the level of automation, through integration with picture archiving and communication systems (PACS) and the electronic record of the patient, decrease the number of false positives, measure the evolution of tumors, evaluate the evolution of the oncological treatment, and its possible prognosis.(Macedo Firmino,2020) This study proposes a lung cancer diagnosis system based on computed tomography (CT) scan images for the detection of the disease. The proposed method uses a sequential approach to achieve this goal. Consequently, two well-organized classifiers, the convolutional neural network (CNN) and feature-based methodology, have been used. In the first step, the CNN classifier is optimized using a newly designed optimization method called the improved Harris hawk optimizer. This method is applied to the dataset, and the classification is commenced. If the disease cannot be detected via this method, the results are conveyed to the second classifier, that is, the feature-based method. This classifier, including Haralick and LBP features, is subsequently applied to the received dataset from the CNN classifier. Finally, if the feature-based method also does not detect cancer, the case study is healthy; otherwise, the case study is cancerous(Zhiqiang Guo. 2021)

In the past decade, the introduction of molecularly targeted agents and immunecheckpoint inhibitors has led to improved survival outcomes for patients with advanced-stage lung cancer; however, this disease remains the leading cause of cancer-related mortality worldwide. Two large randomized controlled trials of lowdose CT (LDCT)-based lung cancer screening in high-risk populations the US National Lung Screening Trial (NLST) and NELSON have provided evidence of a statistically significant mortality reduction in patients. LDCT-based screening programmers for individuals at a high risk of lung cancer have already been implemented in the USA. Furthermore, implementation programmers are currently underway in the UK following the success of the UK Lung Cancer Screening (UKLS) trial, which included the Liverpool Health Lung Project, Manchester Lung Health Check, the Lung Screen Uptake Trial, the West London Lung Cancer Screening pilot and the Yorkshire Lung Screening trial. In this Review, we focus on the current evidence on LDCT-based lung cancer screening and discuss the clinical developments in high-risk populations worldwide; additionally, we address aspects such as cost-effectiveness. We present a framework to define the scope of future implementation research on lung cancer screening programmers referred to as Screening Planning and Implementation RAtionale for Lung cancer (SPIRAL) (Matthijs Oudkerk ,2021).

III. METHODOLOGY AND DISCUSSION

3.1 Existing System:

A lung cancer diagnosis system based on computed tomography (CT) scan images for the detection of the disease. The proposed method uses a sequential approach to achieve this goal. Consequently, two well-organized classifiers, the convolutional neural network (CNN) and feature based methodology, have been used. In the first step, the CNN classifier is optimized using a newly designed optimization method called the improved Harris hawk optimizer. This method is applied to the dataset, and the classification is commenced. If the disease cannot be detected via this method, the results are conveyed to the second classifier, that is, the feature-based method. This classifier, including Haralick and LBP features, is subsequently applied to the received dataset from the CNN classifier. Finally, if the feature-based method also does not detect cancer, the case study is healthy; otherwise, the case study is cancerous. The performance of the proposed classification approach is compared with the existing techniques and the proposed approach outperforms the others. The performance of LESH+CC is quite comparable with the proposed approach. Though the accuracy rates of LESH+CC are greater, the sensitivity and specificity rates are not up to the mark. Besides this, the time consumption of this work is maximal, as it has some difficulty in feature extraction.

3.2 Proposed System:

CNN is of the well-regarded machine learning method in the literature. CNN reduce the parameters of a given problem using spatial relationship between them. One of the reason of its popularity is due to the automatic hierarchical feature representation to recognizing object and images This describes the method of classifying lung tumor cells as malignant or benign a (CNN)Convolutional neural network with alexnet algorithm. Alexnet is one of the most used and transfer learning models. This model of proposed system achieve a high degree of accuracy which is more effective .In the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006199|

First stage we should extract the green channel of original colour CT image. In the next step we should extract the cancer regions of lung in the CT image see in fig.1 and 2

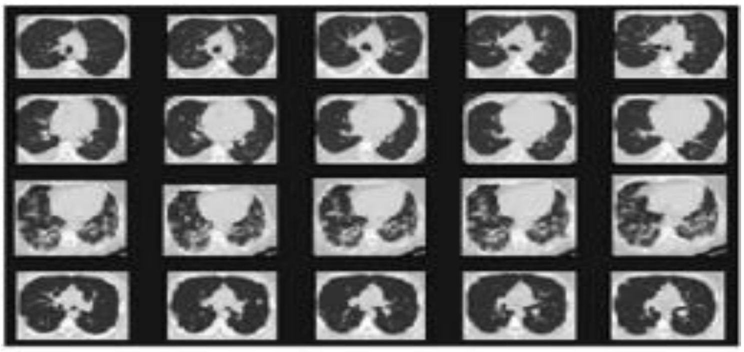
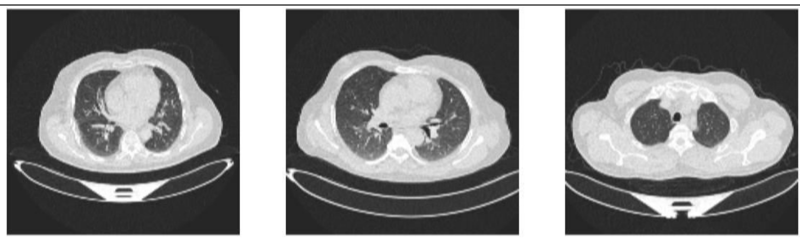


Fig. 1 CT trained image database



(a) Benign (b) Normal (c) Maligant

Fig. 2 Categorization of affected Lungs Disease

In the third step we should identify the affected and non-affected regions in the lungs using segmentation process. At the final step the tumor regions are fed into the alexnet to identified weather it is a benign or malignant cases. The alexnet is the combination of feature extraction and classification in one model. The deep learning classification contains the maxpooling layer, convolutional layer, fully connected layer and softmax layer was show in fig 3 and 4. The learning property of the convolutional layer is to break the pixels of particular image that we provided. Then that pixel is segmented the image into minor pixel boxes by maxpooling layer. kernal and filter operations are performed in this deep learning operations. All parameters are passed effectively by using relu activation function. Unused parameters are dropped out in the pooling layer and the dimensions of the feature map are reduced. In the max pooling layer it performs actions on maximum number of element in the feature map. The average of input data presented in the size of feature map is calculated by the average pooling layers. Each network in the feature map can be reduced by global pooling layer and it convert it to a signal value. And the transformed vector matrix can be collected by fully connected layer.

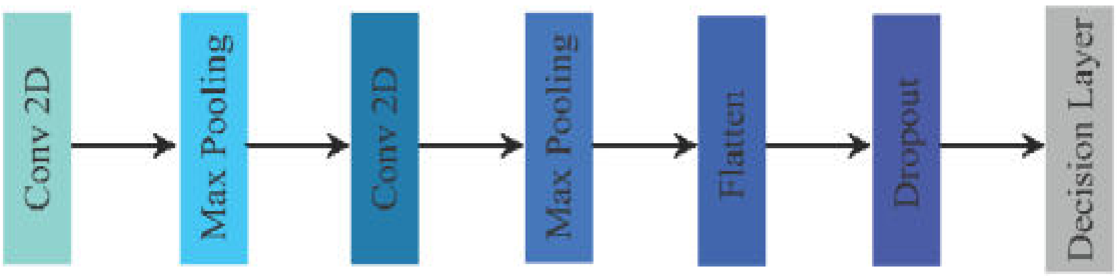


Fig. 3Convolution Layers

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006199|

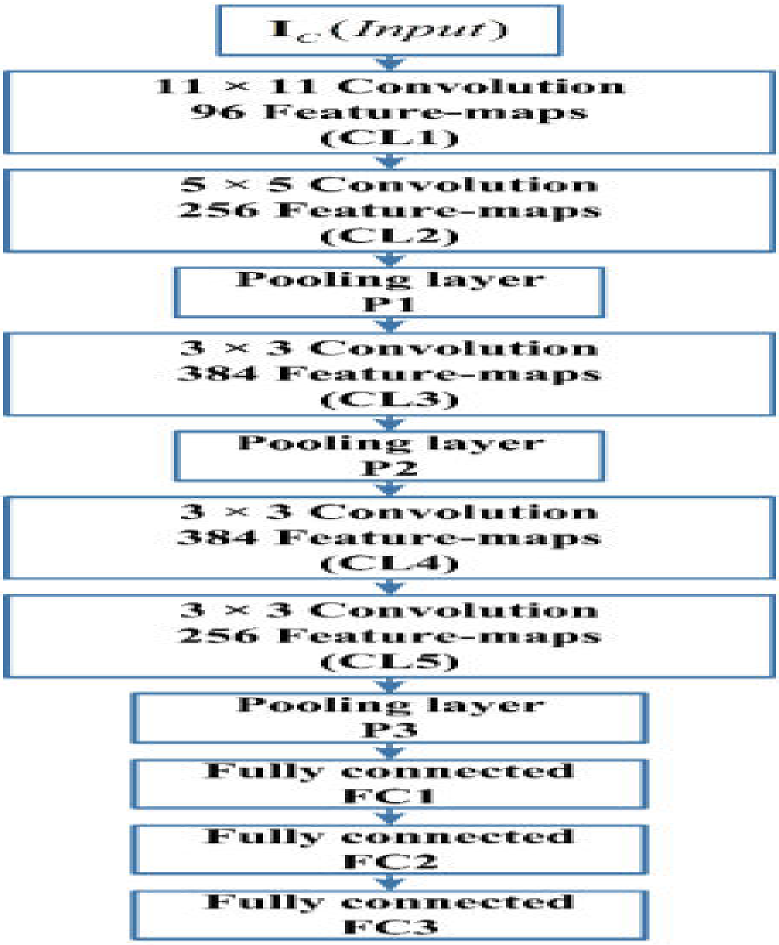


Fig.4 System Architecture

Load the dataset which contains 244,527 images of the 1010 cases. It loads 2705 images for training the AlexNet. The function Split the provided dataset by categories such as 70% of the images for training and 30% for Validating Load the AlexNet network which will provide the basic information of the network. In table 1 shows the flow of AlexNet provided with different weights, bias, padding applicable in convolutional, ReLU and pooling layers the process start in the layer called convolutional layer. This layer pre-process the image converts into pixels of 96*55 then the pixel moves through the maxpooling layer this layer will segment the pixel into different regions. When it arrives in the fully connected layer each cases of result will be assigned some place value like 0,1,2 it stored the pixel values in the matrix format. Softmax layer is the final layer of the architecture In this step, replacement of convolutional layers with FC layers provides the classification output and SoftMaxlayer whereas it already stores the features of the image During the training, the function 'train-Network (imds, layers, options)' used for classifying the images. The term imds – assembles the input images, layers - clarifies the network configuration, options – likelearningrate-0.0001,accuracy-9.91%,Maximum epochs-6,Validation was briefly discussed explained in table1.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006199|

Table 1 calculation of each layer in ALEX NET

Size / Operation	Filter	Depth	Stride	Padding	Number of Parameters	Forward Computation
3* 227 * 227						
Conv1 + Relu	11 * 11	96	4		(11*11*3 + 1) * 96=34944	(11*11*3 + 1) * 96 * 55 * 55=105705600
96 * 55 * 55						
Max Pooling	3 * 3		2			
96 * 27 * 27						
Conv2 + Relu	5 * 5	256	1	2	(5 * 5 * 96 + 1) * 256=614656	(5 * 5 * 96 + 1) * 256 * 27 * 27=448084224
256*2727						
Max Pooling	3*3		2			
256*13*13						
Conv3 + Relu	3 * 3	384	1	1	(3 * 3 * 256 + 1) * 384=885120	(3 * 3 * 256 + 1) * 384 * 13 * 13=149585280
384 * 13 * 13						
Conv4 + Relu	3 * 3	384	1	1	(3 * 3 * 384 + 1) * 384=1327488	(3 * 3 * 384 + 1) * 384 * 13 * 13=224345472
384 * 13 * 13						
Conv5 + Relu	3 * 3	256	1	1	(3 * 3 * 384 + 1) * 256=884992	(3 * 3 * 384 + 1) * 256 * 13 * 13=149563648

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006199|

256 * 13 * 13						
Max Pooling	3 * 3		2			
256 * 6 * 6						
Dropout (rate 0.5)						
FC6 + Relu					256 * 6 * 6 * 4096=37748736	256 * 6 * 6 * 4096=37748736
4096						
Dropout (rate 0.5)						
FC7 + Relu					4096 * 4096=16777216	4096 * 4096=16777216
4096						
FC8 + Relu					4096 * 1000=4096000	4096 * 1000=4096000
1000 classes						
Overall					62369152=62.3 million	1135906176=1.1 billion
Conv VS FC					Conv:3.7million (6%) , FC: 58.6 million (94%)	Conv: 1.08 billion (95%) , FC: 58.6 million (5%)

IV. RESULT AND DISCUSSIONS

The data set is collected from the Kaggle website, Data set divided into three category A training set, A validation set, testing set This will split our dataset into training, validation, and testing sets in the ratio mentioned above- 80% for training (of that, 10% for validation) and 20% for testing. The original dataset consisted of 162 slide images scanned at 40x. An imbalance in the class data with *over 2x* the number of negative data points than positive data points Preprocessing is the process of image reduce the dimension of image. We specify the input image volume shape to our network where depth is the number of color channels each image contains. The network we'll build will be a CNN (Convolutional Neural Network) and call it Cancer Net. This network performs the following operations. Use 3×3

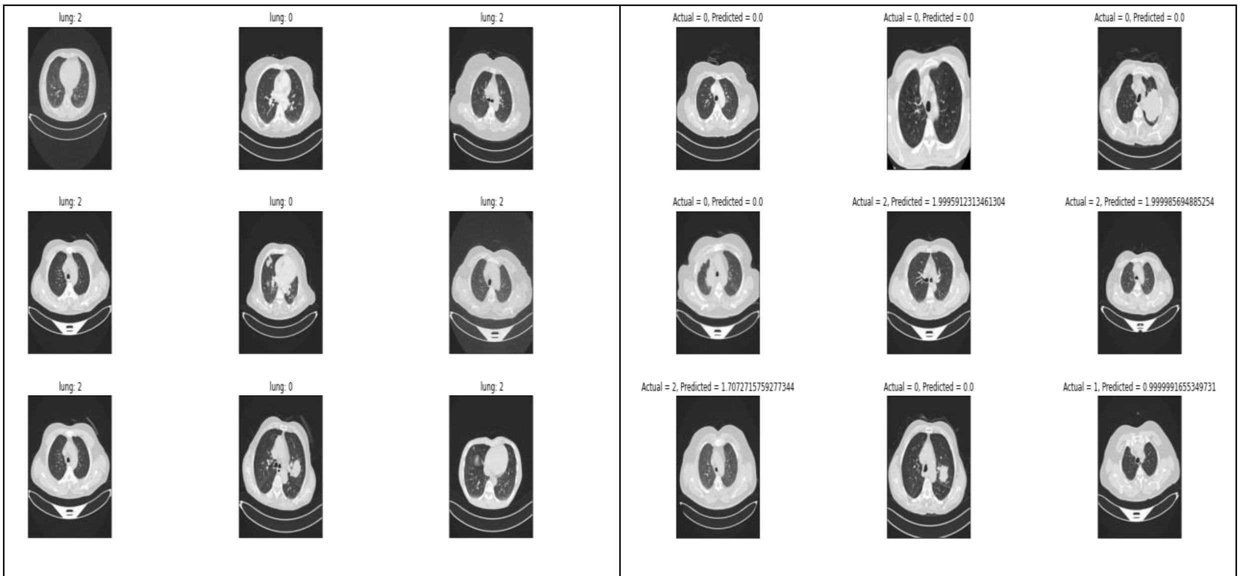
International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006199|

CONV filters Stack these filters on top of each other Perform max-pooling Use depth wise separable convolution (more efficient, takes up less memory). Its consists of input layers, convolution layers, ReLu layer, maxpooling layers for extract the features of images of build model. Feature extraction train the model the build the model. The training process is implemented for the Adam Adaptive momentum as optimizer for gradient with epochs is implemented training process. Its breast cancer, sorted by size, and the items at the beginning are more likely to be benign, and the ones at the end are more likely to be malignant, then you'll be training on benign data, and testing on malignant, which isn't representative. Based on feature vectors we build the model using kera'S system. The testing process is implemented this function we can split the model with a test set of 30% of the original data set. The input jusy specify the size of the input and is called D (see the code above X_ train shape). The dense layer is instead where the real work happens: it takes the input and does a linear transformation to get an output of size 1. The linear transformation we want to apply is the sigmoid activation function so that in output we are in a range of 0 and 1. Loss per iteration, training loss, validating loss is implemented in module. Accuracy and sensitivity of the analyzed. The image resize according the deep learning layer size of rows and column of image Cancer can be detected in the given image that can be shown in fig.5



(a) Output of Training images (b) Output of Testing images

Fig.5 Comparison of Output images

V. CONCLUSION

The methodology adopted in this project aims to develop an automated system for lung cancer detection. Application of median Filter to eliminate impulse noise in the images proved to be a success. The morphological operations also contributed towards satisfactory results in the process of segmentation. Artificial neural networks proved to be a good classifier with acceptable accuracy. The methodology adopted in this project resulted in an accuracy of 99% for the hospital database. this system aims at increasing the accuracy and speed of the lung cancer detection system. It also helps in detecting the cancer at earlier stages. The accuracy of the cancer detection system can be improved by using a different segmentation technique like p-tile thresholding and watershed segmentation followed by binary morphology. Using different feature set like curve let transformation features together with morphological features other than Haralick features, may have a positive impact on the accuracy of the system.

REFERENCES

1. Aggarwal T, Furqan A and Kalra K 2015 Feature extraction and LDA based classification of lung nodules in chest CT scan images Int. Conf. on Advances in Computing, Communications and Informatics (ICACCI) pp. 1189-1193.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006199|

2. Alam J, Alam S and Hossan A 2018 Multi-stage lung cancer detection and prediction using multi-class svm classifier Int. Conf. on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2) pp. 1-4.
3. Anita Chaudhary, Sonit Sukhraj Singh “Lung cancer detection on CT images using image processing”, computing sciences 2012 international conference, IEEE, 2012.
4. Badrul Alam Mia, Mohammad Abu Yusuf , "Detection of lung cancer from CT image using image processing and neural network", International conference on Electrical Engineering and Information Communication Technology (ICEEICT) ,IEEE ,May, 2015.
5. Chen W, Wei H, Peng S, Sun J, Qiao X and Liu B 2019 HSN: hybrid segmentation network for small cell lung cancer segmentation. IEEE Access 75591-603.
6. Fakoor R, Ladhak F, Nazi A and Huber M 2013 Using deep learning to enhance cancer diagnosis and classification. In Proceedings of the Int. Conf. on machine learning vol. 28 (New York, USA).
7. Farahani FV, Ahmadi A and Zarandi MF 2015 Lung nodule diagnosis from CT images based on ensemble learning IEEE Conf. on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) pp. 1-7.
8. Gawade Prathamesh Pratap, R.P Chauhan, "Detection of lung cancer cells using image processing techniques ", 1st IEEE International Conference on Power Electronics, Intelligent control and energy systems, IEEE, 2016
9. Huang PW, Lin PL, Lee CH and Kuo CH 2013 A classification system of lung nodules in CT images based on fractional brownian motion model Int. Conf. on system science and engineering (ICSSE) pp. 37-40.
10. K.Punithavathy, M.M.Ramya, Sumathi Poobal, "Analysis of statistical texture features for automatic lung cancer detection in PET/CT images", International Conference on Robotics, Automation, Control and Embedded systems(RACE),IEEE ,18-20 February 2015.
11. Mohsen Keshani, Zohreh Azimifar and Reza Boostani, "Lung nodule segmentation using activecontour modeling” MVIP, IEEE, 2016
12. Muhammed Anshad, S.S Kumar , Recent methods for the detection of tumor using computer aided diagnosis", International Conference on Control, Instrumentation, Communication and Computational technologies(ICCICCT), IEEE transactions,2014.
13. Nooshin Hadavi, Md Jan Nordin, Ali Shojaeipour , “Lung cancer diagnosis using CT-scan images based on cellular learning automata”, International conference on Computer and Information Sciences(ICCOINS), IEEE , 2014.
14. Rahman MS, Shill PC and Homayra Z 2019 A New Method for Lung Nodule Detection Using Deep Neural Networks for CT Images Int. Conf. on Electrical, Computer and Communication Engineering (ECCE) pp. 1-6.
15. Robert M Haralick, K.Shanmugam, Itshak Dinstein, “Textural Features for Image Classification” IEEE Transactions on Systems, Man and Cybernetics, 3(6), pp. 610 - 621, 1973.
16. Rodrigues MB, Da Nóbrega RV, Alves SS, Rebouças Filho PP, Duarte JB, Sangaiah AK and De Albuquerque VH 2018 Health of things algorithms for malignancy level classification of lung nodules IEEE Access 6 18592-601.
17. Sanc M, Russo M, Stella M and Sikora M 2019 CNN-based method for lung cancer detection in whole slide histopathology images Int. Conf. on Smart and Sustainable Technologies (SpliTech) pp. 1-4.
18. Tran GS, Nghiem TP, Nguyen VT, Luong CM and Burie JC 2019 Improving accuracy of lung nodule classification using deep learning with focal loss. Journal of Healthcare Engineering.
19. Xie Y, Xia Y, Zhang J, Song Y, Feng D, Fulham M and Cai W 2018 Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT. IEEE transactions on medical imaging. 38(4) 991-1004.