TECH-CHRONICLE
AN INTERNATIONAL E-JOURNAL ON
EMERGING TRENDS IN SCIENCE, TECHNOLOGY AND MANAGEMENT

# Lungs Tumor Detection Using Convolutional Neural Networkon Histopathological Images

VinayMotghare[1],Prof.JayantAdhikari[2],Prof.JayantRohankar[3] Department of Artificial Intelligence & Machine Learning
TulsiramjiGaikwad-PatilCollegeofEngineering&Technology, Nagpur
vinaymotghare70@gmail.com

***Abstract:***

**Diseases that affect the lungs have the effect of making breathingdifficult. Oneof thebiggestcauses ofdeathamong people is lung cancer.**

**Human survival rates can be increased through early detection. In the event of a positive diagnosis in lung cancer patients has increased from 14% to 49% throughout that period. When computed tomography (CT) is more accurate than X-ray images, complete diagnosis requires the use of many imaging modalities that complement one another. In this study, we create and test a deep neural network to identify lung cancer in CT scans. An adaptive boosting algorithm and a densely connected convolutional neural network (DenseNet) were used todetermine whether agiven lung image was healthy orcancerous. For this task, we use a dataset containing 201 images of the lungs, with 85% of the images are use for training and 15% use for testing and classifying. According to experimental data, the suggested approach has a 90.85% success rate.**

***Keywords*-- Convolution Neural Networks (CNN), Image Processing, Deep Learning, Machine Learning,Artificial Intelligence**

## I. Introduction

One of the leading causes of death globally is lung cancer [1].It's one of the worst kind of cancer to strike people. It has the highest mortality rate of any tumor and is the leading cause of cancer-related death of men and women [2, 3]. Nearly 2.1 million new cases of lung cancer are diagnosed each year (14.3 percent of all cancers), with 1.8 million deaths occurring as a result (20.4 percent among all cancers). When abnormal cells in the lungs grow and divide uncontrollably, a tumor forms. Lung cancer has thehighest mortalityrate of all cancers. Around 85% of male and 77% of female lung cancer cases are caused by smoking cigarettes. With a mortality rate of 20.4 percent, lung cancerranksamongthedangerous diseases plaguingdeveloping nations. A growing number of people lost their lives every year to lung cancer, making it one of the deadliest forms of the disease [4-6]. The Pros of Early forecasts using fuzzy logic will lead to outcome-focused investigation [5]. Lung cancer survival rates after diagnosis are proportional to the disease's stage. Nonetheless, people's chances of success are highest in their younger years. Lymph fluid, which bathes the lung tissues, acts as a conduit for the dissemination of cancer cells into the bloodstream. The lymph travels through lymph vessels before draining to the lungs and chest lymph nodes. One of the biggest challenges facing humans today is examining and treating lung disease.Thesurvivalrateofpeoplewithtumorscanbe significantlyi ncreasedthroughconsistentearlydetection efforts. This study presents a technique for classifying lung cancers as malignant or benign using a convolutional neural network (CNN). CNN's success rate is 81%, making it the neuralnetwork with the best results. The lung is a common site for malignant tumors to form. The result is lung tissue damage. Lung cancer is characterizedbythe uncontrolled growth of abnormal cells in thelungs..Both lungs are equally at risk. Lung cancer symptoms include coughing up blood and having trouble breathing. Smokers are disproportionately affected by this problem. Cigarette smoke isa major contributor to lung cancer. Cigarette smoking is a leading cause of death in the United States [14, 15, 16]. The World Health Organization estimated in May2020 that tobacco use causes the deaths of more than 8 million people each year. Tobacco consumption is responsible for the deaths of about seven million people worldwide, while secondhand smoke contributes to the deaths of an additional 1.2 million [17].

More than four thousand chemicals have been identified in cigarette smoke, according to [18]. Most of these chemicalshave been linked to cancer. People who smoke more than one pack of cigarettes per day have a 20-25 times higher risk of developing lung cancer than nonsmokers [19], but early detection increases the chance of survival by73%. Around90% of lung cancer fatalities are attributable to tobacco use or smoking [1]. Although cigarette smoking is still the leading cause of lung cancer, there are other risk factors to consider. Lung cancer risk factors include environmental pollution (primarily air pollution), excessive alcohol consumption, and exposure to tobacco smoke (passive smoking).

Manypeople wouldbenefit from earlydetection or prediction of lung cancer. This might be an efficient preventative measure. Lung cancer diagnosis often involves a battery of tests,including computed tomography, chest radiography, sputum cytology, magnetic resonance imaging, etc. Late or delayed cancer identification, however, dramatically reduces a patient's likelihood of survival, and most of these procedures are expensive and time-consuming. [21] A technological solution for early diagnosis is required to address this challenge, since it will lead to more rapid outcomes.

The system uses deep learningmethods for cancer detection. To distinguish between cancerous and noncancerous tissue, the system employs the Convolutional Neural Network algorithmon whole slide images [3]. [23]The algorithm in this system for early detection is responsible for extracting features. The machine will make a lung cancer diagnosis based on these characteristics.The system was developed tosupplya quickand cheap answer for lung cancer early detection. Certain criteriaandthresholds, tobedetailedlater [22], willbeused tomakethe detection. Medical image processing [12, 13] is a popular application of Deep Learning.

## II. ExistingSystem

Histopathological classification of lung cancer now frequently makes use of medical imaging [5]. In this study, we provide a novel CNN-based architecture for classification of lung cancer images usinghistopathology. One input layer, various blocks of convolutional layers, a drop-out layer, a max-pooling layer, and a fully connected layer are all part of the CNN architecture presented in this study. Sixteen convolutional layers, five dropout layers, fivemax-poolinglayers, andonefullyconnected layer make up the total network architecture. The proposed architecture thus validates the superior performance when compared to baseline models.

MRI images are used as a dataset in [6]. Nevertheless, thedataset is quite tiny, hence transfer learning is applied. Insteadof starting with a blank slate, models built using transferlearning make advantage of pre-trained, deep convolutional Neural Network architectures.

AlexNet, GoogleNet, VGG-19,VGG16,ResNet-18, ResNet-50, ResNet-101,ResNet-inception-v2,andSENetarejustsomeof architectures used on the dataset in the paper. Epochs of 25, 50, and 90 are used to generate the results. The findingsarethen comparedwith oneanother. Finally, theresults from the three selected epochs demonstrate that VGG16, VGG19, and some AlexNet layers outperform ResNet and ResNet-Inception-v2. Their respective rates of accuracy were 98.71, 98.55, and 98.55 percent. After 50 iterations, AlexNetand VGG16 both reached 98.55% precision. Nonetheless, VGG16 was slower than AlexNet and took more time to train.

In , the author explains how neural networks have changed over time. Using examples of classification, detection, and segmentation on pulmonarymedicalimages, thispaper provides an introduction to the application of deep learning to medical image analysis. The steps involved in using deep learning to identifythepulmonarynodulearealsodetailed. Lung conditions such as pulmonary nodule disease, pulmonary embolism, pneumonia, and interstitial lung disease are all treated using these strategies.

In order to categorize pneumonia, a chest X-ray dataset is usedin . They used three different methods on this data set, the first of which was a linear support vector machine model with local rotation and orientation free features. Second, we use two different convolutional neural network models to perform transfer learning.Thecapsulenetwork isthethirdmethod. Since the dataset used was insufficient, various forms ofaugmentation are applied to the same data in an effort to boost accuracy. Results from all three experiments pointed to CNN-based transfer learning as the most effective method. Superior efficiency is achieved by capsule networks compared to ORB and SVM classifiers.

## III. ProposedModel

The architecture of convolutional neural networks is optimized to perform image classification with a small number of parameters. A convolutional neural network is built up from severalinterconnectedlayers,eachofwhichperformsaspecific architecture of a ConvNet is a neuronal connectivity pattern that is reminiscent of the Visual Cortex. When data is "augmented," its volume and complexity grow. Insteadofconvertingthe existingdata, we will collect newdata. Since we need large amounts of data for deep learning, and in some cases it's not really feasible to gather millions of images, data augmentation is a crucial step. These data optimize the dataset magnitude and add ambiguity inside the patients, herethereferences.

## IV. Preliminaries

Several machine learning and deep learningmethods on diverse datasets may be used to diagnose lung cancer. Various studieson lung cancer detection has already done by researchers on different datasets. Brief explanations of the studies reported in publications about methods for detecting lung cancer are provided below. The UCI machine learning repository and data world have yielded varying outcomes for the various classifiers used to detect lung cancer. Classifiers such as DT, LR, NB, and SVM were used to put the findings into practice. It has been shown that SVM outperformed other classifiers on the dataset . In this research, the author conducted experiments using the Lung Imaging Database Consortium image collection(LIDC-IDRI) using a deep neural network and auto encoder to identify lung cancer. Accuracy levels of 79%, 81%, and 79% were attained by them, respectively . They used Random Forest, an ensemble of U-Net, XGBoost, and ResNet, to obtain an accuracy of 84% on the Lung Imaging Database Consortium image collection (LIDC-IDRI), proposing this way of creating deep residual networks and classifiers . This research proposes a CNN architecture for identifying lung cancer from whole-section histopathology pictures. They performed a VGG anda reset. Receiver Operating Characteristic (ROC) plotswere used to evaluate VGG and ResNet's final results. In terms of patch-level accuracy, the VGG16 model scored 75.41% and the restingmodelscored72.05%.Theauthorsofthispaper[8]

TECH-CHRONICLE
AN INTERNATIONAL E-JOURNAL ON
EMERGING TRENDS IN SCIENCE, TECHNOLOGY AND MANAGEMENT

describe how theyported a CNN network over to the LC25000. Using a combination of lung and colon histopathologicalimages, theauthorstraineda CNNmodel todistinguish between three distinct types of lung cancer. The constructed CNN model hastrainingandvalidationaccuracies of96.11and97.2percent, respectively.

## V. OperationalTrainingusingCNN

The Convolutional Neural Networks (CNNs or ConvNets) utilized for the categorization and recognition of images were built using a linear stack of layers. Convolutional layers with kernel filters, max pooling, and fullylinked layers were used on both the training and testing pictures. The item was put througha classification process using the softmax algorithm. The model wasdevelopedandtestedusinghardwarewith theGPUmoniker provided by the Google Colaboratory. For this assignment, we used a neural network with three hidden layers, one input layer, and one fully connected layer. To facilitate both training and evaluation, images are divided 90:10. The input layer received images of size (180, 180) pixels. Each convolutional layer useda kernel matrix of size (3, 3) with the activation function (ReLU(x) = max (0, x)). To lessen the load on the subsequent convolutionlayer'scomputingparameters,a maxpoolingsizeof (2, 2) was introduced. The model used a dropout value of 0.1, which is rather little. The class probabilities for the final output classes were calculated using a dense value of three and the sigmoid activation function. Adam, an optimizer based on adaptive moment estimation, was used to determine the rates of parameter learning. To determine whether the model could correctly identify malignant lesions on radiographs, a detection performance test was run on a per-lesion basis using the test dataset.Achestradiographlesion'slikelihoodofbeing malignant was converted to a number between 0and 255 by the model. True positives were determined by whether or not the median of the model's output fell within the range of the gold standard (TP). The remaining outcomes were all FPs. When the model suggested many TPs for the same underlying fact, we aggregated them into a single TP. For every missing result for a single ground truth, the model returned one FN. A.S. and D.U., two radiologists, looked at the radiograph and CT after the fact to see what structures were picked up by the FP output. Segmentation accuracy was also measured using the dice coefficient.

There are two primary components to the method suggested for detecting lung cancer. As a preliminary step, we do preprocessing on the photos before feeding them to 3D convolutionalneuralnetworks. InordertoidentifyCT imagesas positive or negative for lung cancer, we firsthad todiscover the nodule candidate that would be utilized to train 3Dconvolutional neural networks.

*1) PreprocessingandSegmentation*

For each patient, we first convert thepixel valuesin each image to Hounsfield

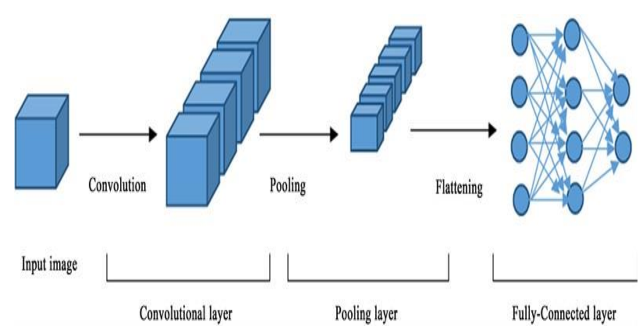


Fig1.PreprocessingofImagesinConvolutionalNeuralNetwork

## VI. Descriptionofdataset

Kaggle dataset is used in thisstudy. There area total of 250,000 images in this dataset, with 5,000 images from each of five categories. Three forms of lung cancer and two of colon cancer are present. Histopathological images are the focus of this data set. All of the JPEG files are 768 pixels on the longest side. Histopathology is the study of disease through microscopic analysis oftissue. In histopathologyand clinical medicine refers specificallyto the examination of a processed andhistologically sectioned surgical or biopsy specimen on glass slides by a pathologist. Histopathologists are specialists in cancerdiagnosis; they are the ones who examine the tissues and cells taken from suspected "lumps and bumps," determinewhether or not they are malignant, and then report back to the treating physician with details about the cancer's type, grade, and response to treatment. The pancreas and retro peritoneum, for example, were previously inaccessible for biopsy tissueretrieval, but modern imaging techniques have made this possible. After the tissue has been collected, it is usually processed overnight before being examined microscopically (Figure 1).

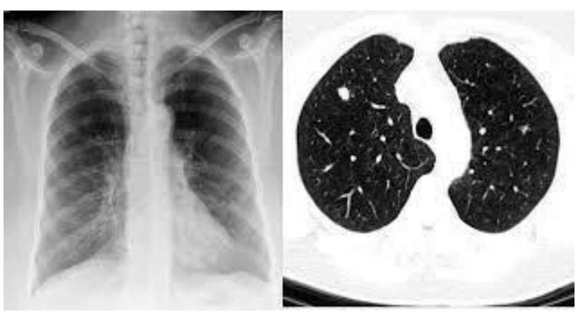

Fig2.ImageofLungs

***Imageresizing:***In computer vision,resizingpicturesisacrucial part of the preprocessing phase. Our machine learning algorithms,ingeneral,benefitmostfromtrainingonlower resolutionphotos.Inaddition,ourrawgatheredphotosmaynot all be the same size, yet this is a requirement of many deep learning model designs. In order for the CNN to function properly, all photos must be scaled to the same dimensions before being fed into it. In the dataset, the original image size is a sizable 768 pixels by 768 pixels. Computer vision is used to resize the photos to 224 by 224 pixels so that the model may be trained more quickly and at a lower computational cost.

***Dataset splitting:*** Separating your data into a train, validation, andtest setisamustifyou wanttosucceed in machinelearning. The rationale for this practice is elementary. The model should be tested using the same data it has already seen if we did not separatethedataintovarious sets. Hence, issues likeover fitting may arise throughout training without our awareness. Hone inon The Effective Net-B0 model was used. Pre-trained on Image.Net, this model was fine-tuned by swapping out the last layer alongwith a seriesoflinear anddropout ones. Inthiscase, we use a ReLu activation function in 2 dense layers. Dropout layer with a 0.2 percent rate is used in ReLu (f(x)=max (0, x)). Dropout should be utilized to prevent the model from being overfit. Once a last dense output layer was placed, the process was considered complete.

Table1:DetailsofSpecification

| | |
|---|---|
| Model | MATLABR2014a |
| UsedSoftware | 224*224 |
| ImageSize | 32 |
| BatchSize | 1.00E-03 |
| LearningRate | 15 |
| Epochs | 5 |
| Additional Layers | 0.2 |
| Dropout Rate | Adam |
| Optimizer | CrossEntropy |
| LossFunction | Loss |

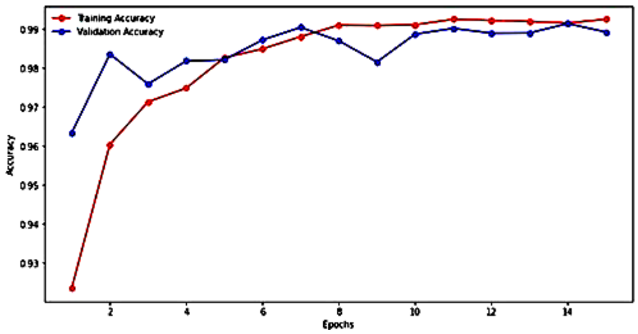


**Figure3:**ModelTrainingandvalidationaccuracy*vs*.epoch

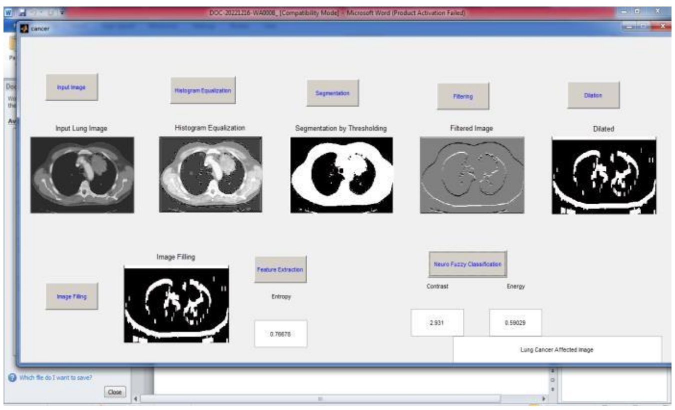


Fig3OverallsystemGUI

## VII. RESULTANDDISCUSSION

Above, Fig. 2 and Fig. 3shows the plot of model accuracy vs. epoch and model loss vs. epoch for training and validation images.The images are train for 20 epochs with batch size 65 with 212 steps in each epoch. The model achieved a training accuracy of 97.11% and a validation accuracy of 98.20% in the final approach.

## VIII. CONCLUSION

Usinghistopathological scans, thisstudydemonstrateshowlung cancer may be diagnosed. One picture was trained using a CNN to determine if it belonged to the benign, Adenocarcinoma, or squamous cell carcinoma categories. The model's training accuracywas97.11percent,anditsvalidationaccuracywas 98.20percent.Modelperformancewasevaluatedusinga numberof metrics, including precision, f1- score, recall, and a confusion matrix plot.

## References

1. Wadood Abdul, An Automatic Lung Cancer Detection and Classification (ALCDC) System Using Convolutional Neural Network 2020 13th International Conference on Developments in eSystems Engineering (DeSE) IEEE Xplore: 14 June 2021

2. Amrit Sreekumar; Karthika Rajan Nair; SnehaSudheer; H Ganesh Nayar; Jyothisha J Nair Malignant Lung Nodule Detection using Deep Learning 2020 International Conference on Communication andSignal Processing (ICCSP) ( Volume: 39, Issue: 12, December 2020) IEEE Xplore

3. Nusraat Nawreen; Umma Hany; Tahmina Islam Lung Cancer Detection and Classification using CT Scan Image Processing 2021 International Conference on Automation, ControlandMechatronicsfor Industry4.0 (ACMI) IEEE Xplore: 08 September 2021

4. Aman Agarwal; Kritik Patni; Rajeswari D LungCancer Detection and Classification Based on Alexnet CNN 2021 6th International Conference on Communication and Electronics Systems (ICCES) IEEE Xplore: 02 August 2021

5. Mohseena Thaseen; S.K. UmaMaheswaran; Darshana A Naik; Mrunal Swapnil Aware A Review of Using CNN Approach for Lung Cancer Detection Through Machine Learning 2022 2nd International Conference on Advance Computing and Innovative Technologiesin Engineering (ICACITE) IEEE Xplore: 18 July2022

6. D. Monisha; N. Rino Nelson; Detection of lung tumor using SVM and Bayesian classification 2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP) IEEE Xplore: 25 April 2022

7. Rao, P., Pereira, N.A. and Srinivasan, R. (2016) Convolutional Neural Network for Lung Cancer Screening in Computed Tomography (CT) Scans. 2nd International Conference on Contemporary Computing and Informatics (IC3I), Noida, 14-17 December 2016, 489-493.

8. Tan,J.X.,Huo,Y.M.,Liang,Z.R.andLi,L.H.(2017) AComparisonStudyontheEffectofFalsePositive ReductioninDeepLearningBasedDetectionfor

Juxt apleural Lung Nodules: Cnn vs Dnn. Proceedings of the Symposium on Modeling and Simulation in Medicine (MSM), San Diego, 23-26 April 2017.

9. Golan, R., Jacob, C. and Denzinger, J. (2016) Lung Nodule Detection in CT Images Using Deep Convolutional Neural Networks. International Joint Conference on Neural Networks (IJCNN), Vancouver, 24-29 July 2016, 243-250.https://doi.org/10.1109/IJCNN.2016.7727205.

10. W. D. Travis, et al. International association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary classification of lungadenocarcinoma. Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer 6 (2) (2011) 244-85

11. JLortet-Tieulent ISoerjomataram JFerlay MRutherford EWeiderpass FBray ‘‘International trends in lung cancer incidence by histological subtype: Adenocarcinoma stabilizing in men but still increasing in women Lung Cancer 84(1): 13-22 2014

12. DSharma GJindal ‘‘Computer aided diagnosis system for detection of lung cancer in CT scan images International Journal of Computer and Electrical Engineering 3(5): 714-718 2011

13. DBhatnagarAKTiwariVVijayarajanAKrishnamoorthy ‘‘Classification ofnormalandabnormal imagesoflung cancer IOP Conference Series: Materials Science and Engineering Vol 263 2017

14. Narendra Mohan Tumor Detection From Brain MRI Using Modified Sea Lion Optimization Based Kernel Extreme Learning Algorithm International Journal of Engineering Trends and Technology 68(9)(2020)84-100.

15. A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25 (2012). doi: 10.1145/3065386.

16. AA. Borkowski, MM. Bui, LB. Thomas, CP. Wilson, LA. DeLand, SM. Mastorides. Lung and Colon Cancer Histopathologica l Image Dataset. (LC25000). ArXiv: 1912.12142v1 [eess.IV], (2019)

17. SRS Chakravarthy and H. Rajaguru. Lung Cancer Detection using Probabilistic Neural Network with modifiedCrowSearch Algorithm.AsianPacificJournal of Cancer Prevention, 20 (7) (2019) 2159-2166. doi: 10.31557/APJCP.2019.20.7.2159

18. W. Rahane, H. Dalvi, Y. Magar, A. Kalane and S. Jondhale, Lung Cancer Detection Using Image Processing and Machine Learning HealthCare, International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, (2018)1-5

19. T. Atsushi, T. Tetsuya, K. Yuka, and F. Hiroshi. Automated Classification of Lung Cancer Types from Cytological Images Using Deep Convolutional Neural Networks”.BioMedResearchInternational.(2017) 1-6. doi:10.1155/2017/4067832

20. W. Ausawalaithong, A. Thirach, S. Marukatat, and T. Wilaiprasitporn, Automatic Lung Cancer Prediction from Chest X-ray Images Using the Deep Learning Approach, 2018 11th Biomedical Engineering International Conference (BMEiCON), Chiang Mai, (2018) 1-5. doi: 10.1109/BMEiCON.2018.8609997

21. D. Bazazeh and R. Shubair, Comparative study of machinelearningalgorithmsfor breastcancerdetection and diagnosis, 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), Ras Al Khaimah, (2016) 1-4 doi: 10.1109/ICEDSA.2016.7818560.

22. K. Yu, C. Zhang, G. Berry, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun 7 (2016) 12474 doi: 10.1038/ncomms12474