IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

Lung Cancer Prediction using Deep Neural Networks

Smt. K S L Sruthi[1], K Divya Sri[2], MD Khairunnisa[3], N Leela Vyshnavi[4]
Assistant Professor, Department of Information Technology[1]
U. G. Students, Department of Information Technology[2,3,4]
S.R.K.R. Engineering College, Bhimavaram, Andhra Pradesh, India

Abstract: *Lung cancer is one of the leading diseases caused all over the world. In India one of the most occurring diseases is lung cancer and lot of people die due to the reason that it can only be cured during its initial stages. It is caused by the uncontrollable growth of cells in the lung tissues. It can only be treated in its early stages, when therapy is started. Computed Tomography (CT) scans and blood test data are used to identify this. The tumor is diagnosed by a blood test after individuals have been impacted for at least four years.*

CT scanning is used to determine the early stage of cancer. The CT pictures are divided into two categories: normal and pathological. Focusing on the tumor part of the picture reveals the aberrant image. The collection consists of Computed Tomography (CT) pictures in jpg format. The Convolutional Neural Network is used to train the suggested model (CNN). During the training, picture enlargement techniques such as zooming, cutting, horizontal filling, and twisting were used on the dataset to improve the classification success rate.

Lung cancer is detected using pre-trained ImageNet models such as LeNet, AlexNet, and VGG-16. The suggested model is based on the AlexNet model, and the features extracted from the network's final fully connected layer were used as distinct inputs to the SoftMax classifier. The combination of AlexNet and the SoftMax layer resulted in a 100 percent accuracy. The suggested methodology can be used to diagnose lung cancer in a consistent and long- term manner.

Keywords: ImageNet models

I. INTRODUCTION

1.1 Background

There have been varies models used in today's world to predict lung cancer. This section contains a summary of the literature and an analysis of lung cancer. The tests were undertaken performed using a publicly available collection of computed tomography (CT) images. For feature extraction and classification, CNNs were utilised. The most common kinds of lung cancer are listed, along with their features. The inputs are collected online from patients for online detections using the reinforcement algorithm. The disturbances in the data are eliminated by applying weighted mean histogram equalization approach. The IPCT is used to improve the image quality.

A probabilistic model is used to predict lung cancer by combining a low-dose CT scan dataset with full-mode iterative recombination and a CNN-based approach for deep feature extraction (IMR). Experiments are conducted on 1 year, 2 year and 3 years affected person CT images. After the second screening CT scan, 94 percent, 85 percent, and 71 percent of those designated high risk by Deep Learning were diagnosed with incident and interval lung cancer within one year, two years, and three years, respectively. In this test, two supervised machine learning classifiers are used to classify CT images as normal, malignant, or benign. This stage's working rule is based on the standard backpropagation algorithm. In this a new model is created by using the SVM and CNN together called SVM-CNN hybrid model for the pattern recognition. The last output layer of CNN is substituted with SVM in this case. Deep CNN model is utilised to create a classification scheme for microscopic lung cancer pictures, and CNN model is employed for classification. It has two fully linked layers, three convolutional layers, and three pooling layers. To avoid overfitting during the training phase, each picture is subjected to a series of techniques including reflection, rotation, and filtering. In the categorization procedure, it was successful 71% of the time

1.2 Lung Cancer

Cancer is a disease that results from abnormal cell/tissue development in the human body. Cancer has been one of the most devastating illnesses to endanger human health for many years. According to the numbers, there are 1.8 million new cancer cases identified each year, with more than 606,520 deaths in united states as of 2020 and estimated new cases of whopping 247,270 lung cancer cases. According to the research done in 2018, an additional 18.1 million cancer cases will be added to the world's available cancer cases, with about 9.6 million deaths. Lung cancer is the most frequent cancer in the world, with a 13 percent mortality rate.

The analysis and diagnosis of a lung CT scan is a delicate process that demands time and good qualifications in the medical profession. Variability among observers arises as a result of the subjective examination. Computer-based solutions are in high demand as a result of these factors. Existing technology can be used to diagnose the problem. Thus, the cost will be drastically reduced. For that purpose, one of the common models preferred for lung cancer diagnosis is deep learning. In our study, the CT image dataset is classified as Cancer and Non-cancer dataset. The Convolutional Neural Network is used to train the suggested model (CNN). During the training, picture enlargement techniques such as zooming, cutting, horizontal filling, and twisting are used on the dataset to improve the classification success rate.

The suggested one is a custom CNN model architecture using TensorFlow's Keras API. It does not correspond to a specific pre-defined CNN architecture like LeNet, AlexNet, VGG-16, etc. Instead, it is a custom architecture designed specifically for the task at hand, which appears to be binary image classification (potentially for diagnosing lung cancer based on CT scan images). This custom CNN architecture is a fairly standard design for image classification tasks, consisting of alternating convolutional layers followed by max-pooling layers for feature extraction, followed by fully connected layers for classification. The use of ReLU (Rectified Linear Unit) activation functions, dropout regularization, and SoftMax activation in the output layer is common in CNN architectures for classification tasks. The Dropout layer is used for regularization to prevent overfitting. The features are extracted from the network's final fully connected layer were used as distinct inputs to the SoftMax classifier. While it doesn't have a specific name like those pre-defined architectures, it is a flexible and effective approach tailored to the specific requirements of the problem being addressed.

II. LITERATURE SURVEY (RELATED WORK)

Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method- 2022

We developed and validated a deep learning (DL)-based model using the segmentation method and assessed its ability to detect lung cancer on chest radiographs. Chest radiographs for use as a training dataset and a test dataset were collected separately from January 2006 to June 2018 at our hospital. The training dataset was used to train and validate the DL-based model with five-fold cross-validation. The model sensitivity and mean false positive indications per image (MFPI) were assessed with the independent test dataset. The training dataset included 629 radiographs with 652 nodules/masses and the test dataset included 151 radiographs with 159 nodules/masses. The DL-based model had a sensitivity of 0.73 with 0.13 MFPI in the test dataset. Sensitivity was lower in lung cancers that overlapped with blind spots such as pulmonary apices, pulmonary hila, chest wall, heart, and sub-diaphragmatic space (0.50–0.64) compared with those in non-overlapped locations (0.87). The dice coefficient for the 159 malignant lesions was on average 0.52. The DL-based model was able to detect lung cancers on chest radiographs, with low MFPI..

A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms- 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)

The growth of cancerous cells in lungs is called lung cancer. The mortality rate of both men and women has expanded due to the increasing rate of incidence of cancer. Lung cancer is a disease where cells in the lungs multiply uncontrollably. Lung cancer cannot be prevented but its risk can be reduced. So, detection of lung cancer at the earliest is crucial for the survival rate of patients. The number of chain- smokers is directly proportional to the number of people affected with lung cancer. The lung cancer prediction was analysed using classification algorithms such as Naive Bayes, SVM, Decision tree and Logistic Regression. The key objective of this paper is the early diagnosis of lung cancer by examining the performance of classification algorithms..

Lung Cancer Detection Using Deep Learning- 2019

Radhika P.R.; Rakhi A.S. Nair; Veena G. Lung cancer keeps on dynamical on varied medical factors counting on topographical areas. The identification of respiratory organ cancer at initial stages is of maximum importance if it's meant to degrade high morbidity. The worldwide respiratory organ screening program focuses to imagine PET/CT examinations amongst most matured gatherings at danger to upgrade the first location rate. In spite of the very fact that utilization of obtrusive procedures, facet effects scarcely show up till infection is propelled creating it difficult for radiotherapist to acknowledge sores. Every year, the yank Cancer Society appraises the quantities of recent growth cases and spending which will happen within the world within the gift year and aggregates the most recent info on growth frequency, mortality, and survival. Real and precise info is that the basis of sickness management initiatives. Quite 3/4th of the sickness is known with tobacco utilization. What are more, hereditary parts, presentation to ecological poisons, second user smoking expands sickness quickly. Cures together with therapy, radiation, surgery, dermal open medications raise survival rate and private satisfaction. This strategy is a lot of regarding designation at earlier than schedule and demanding stages with keen procedure procedures with completely different noise elimination by segmentation ways and calculations that is that the root plan of digital image process. Location of CT photos received from cancer analysis organizations is investigated utilizing MATLAB. The accuracy was not up to 80 percent..

Lung Cancer Classification and Prediction Using Machine Learning and Image Processing-2022

G. Arunkumar, Anil Kumar Bisht, Shivlal Mewada, J. N. V. R. Swarup Kumar, Malik Jawarneh, and Evans Asenso Lung cancer is a potentially lethal illness. Cancer detection continues to be a challenge for medical professionals. The true cause of cancer and its complete treatment have still not been discovered. Cancer that is caught early enough can be treated. Image processing methods such as noise reduction, feature extraction, identification of damaged regions, and maybe a comparison with data on the medical history of lung cancer are used to locate portions of the lung that have been impacted by cancer. This research shows an accurate classification and prediction of lung cancer using technology that is enabled by machine learning and image processing. To begin, photos need to be gathered. In the experimental investigation, 83 CT scans from 70 distinct patients were utilized as the dataset. The geometric mean filter is used during picture preprocessing. As a consequence, image quality is enhanced. The K means technique is then used to segment the images. The part of the image may be found using this segmentation. Then, classification methods using machine learning are used. For the classification, ANN, KNN, and RF are some of the machine learning techniques that were used. It is found that the ANN model is producing more accurate results for predicting lung cancer.

Lung CT Image Segmentation Using Deep Neural Networks- 2018

Brahim Ait Skourt, Abdelhamid El Hassani, Aicha Majda Lung CT Image Segmentation Using Deep Neural Networks Image segmentation is a necessary initial step for lung image analysis, it is a prerequisite step to provide an accurate lung CT image analysis such as lung cancer detection. In this work, we propose a lung CT image segmentation using the U-net architecture, one of the most used architectures in deep learning for image segmentation. The architecture consists of a contracting path to extract high-level information and a symmetric expanding path that recovers the information needed. This network can be trained end-to-end from very few images and outperforms many methods

Lung Cancer Tumor Region Segmentation Using Recurrent 3D DenseUNet-2018

Uday Kamal, Abdul Muntakim Rafi, Rakibul Hoque, Jonathan Wu, Md. Kamrul Hasan 2018 The performance of a computer-aided automated diagnosis system of lung cancer from Computed Tomography (CT) volumetric images greatly depends on the accurate detection and segmentation of tumor regions. In this paper, we present Recurrent 3D-DenseUNet, a novel deep learning-based architecture for volumetric lung tumor segmentation from CT scans. The proposed architecture consists of a 3D encoder block that learns to extract fine-grained spatial and coarse-grained temporal features, a recurrent block of multiple Convolutional Long Short-Term Memory (Conv LSTM) layers to extract fine-grained spatio temporal information, and finally a 3D decoder block to reconstruct the desired volume segmentation masks from the latent feature space. The encoder and decoder blocks consist of several 3D-convolutional layers that are densely connected among themselves so that necessary feature aggregation can occur throughout the network. During prediction, we apply selective thresholding followed by morphological operation, on top of the network prediction, to better

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

differentiate between tumorous and non-tumorous image-slices, which shows more promise than only thresholding-based approaches. We train and test our network on the NSCLC-Radiomics dataset of 300 patients, provided by The Cancer Imaging Archive (TCIA) for the 2018 IEEE VIP Cup. Moreover, we perform an extensive ablation study of different loss functions in practice for this task. The proposed network outperforms other state-of-the-art 3D segmentation architectures with an average dice score of 0.7228

III. METHODOLOGY

The main objective of this project is to predict whether a person is having lung cancer or not. The project was built using Image Processing and Deep Neural Networks. This deep learning model has been built on PyCharm using Python. The project was built in 4 phases.

First being collecting and pre-processing the dataset. Second being creating the model and training the model. Third being the application layer of the model to obtain the desired results. Fourth being, creating the graphical user interface for the created model.

Phase 1: The first phase involves the pre-processing of collected data. Like gray scaling all the images in the dataset and then resizing the images in the desired shape.

Phase 2: This phase involves dividing the processed dataset into training data and testing data, then building a sequential model using Convolutional Neural Networks from the Keras neural network library which runs on top of TensorFlow. Then, will feed all this to a model class, identify the best optimizer, run it through multiple epochs to reach the best accuracy possible and save the model. Also, run this model on a benchmark dataset to check if the model is working well.

Phase 3: This is the application phase where the input is taken at runtime and the captured hand gesture is passed to the model to get it classified and mapped to the lung of its class as output.

Phase 4: I have saved the above model and imported the model in graphical user interface. I have created a graphical user interface where the user can upload the image. After uploading the image we can see a classify button and after classifying we get the desired output.Here's a breakdown of the methodology into steps:

Development of Experimental Dataset

The dataset is gathered from various sources such as open source applications and from the hospitals. I've gathered 100 images with 50 as cancer images and 50 as normal images. The format is DCIOM and later it is converted to jpg format for training the data due to the limitation of the format support in python. The picture has a resolution of 504 × 504 pixels. Because the dataset is limited, we are relying on data augmentation approaches to improve accuracy. The classification accuracy of CNN and traditional machine learning algorithms may be improved using various data augmentation strategies. The picture augmentation algorithms in this work are implemented in Python using the Keras package. Cutting, rotation, horizontal turning, width and height changes, and filling procedures are all used to enhance each image in the collection. AlexNet is the most well-known and well-classified image classification system. CNN's most successful model is this one. Three pooling layers, five convolutional layers, and three fully linked layers make up AlexNet's architecture. Convolutional layers have 3x3 and 2x2 filters. SoftMax is the last layer utilised in this. To avoid overfitting, the test set was employed without any augmentation approaches.

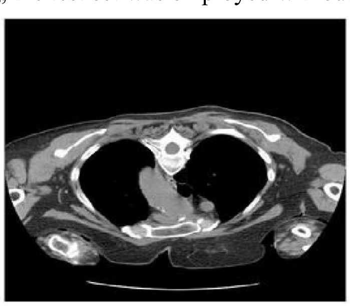


Fig 2.1 (i)Normal lung CT image

(ii)Lung Cancer CT image

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

Pre-processing

Gray Scaling

The images in the dataset are converted to gray scale images. The process of transforming a picture from various colour spaces to shades of grey is known as grey scaling. It might be completely black or completely white. The reason for the distinction between such pictures and any other type of colour image is that each pixel requires less information. Grayscale representations are frequently employed for extracting descriptors rather than operating directly on colour pictures since they simplify the technique and minimise computing requirements.

Grayscale images

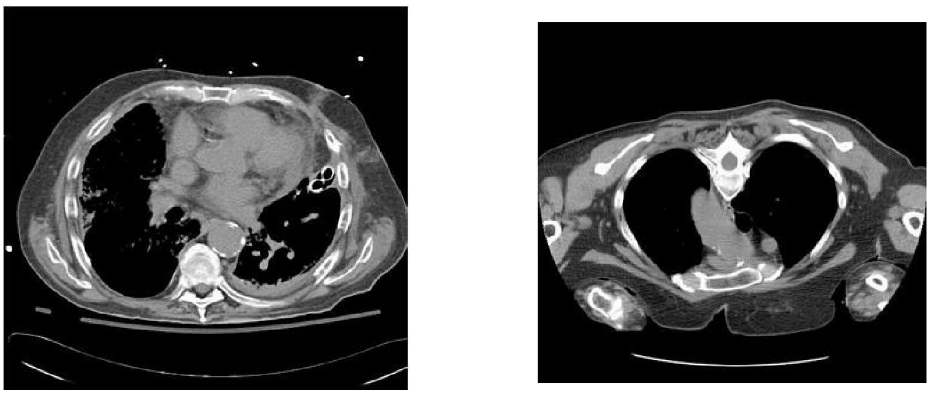


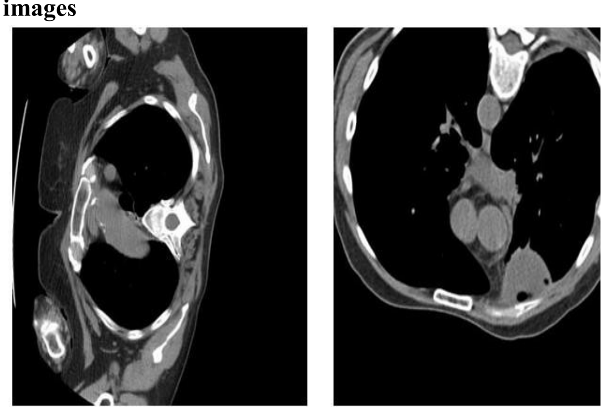
Image Resizing

The image resizing is done for my dataset because the images in the dataset are not of same size. So as to bring the images into the same size I have done resizing the images. While resizing the images the pixels in an image are updated. This usually results in a pixilated or blurry image. Hence, downsizing a picture is much safer than enlarging an image.

Data Augmentation

The next step involves expansion of the created image dataset. The created image dataset had images of complete similarity for each gesture class. This similarity will not allow us to build a generalized model and will go linear with any simple CNN model to give 100% accuracy. So, data augmentation was done to make the image dataset more challenging and sophisticated. For this, a data generator class was created to rotate, zoom, shift width, shift height and perform horizontal flips randomly on different images in a varied range. Data augmentation was also used to increase the dataset size by almost 3 times to make the data more complex and increase generalizability. Now, this image data requires a complex CNN architecture to give good accuracy.

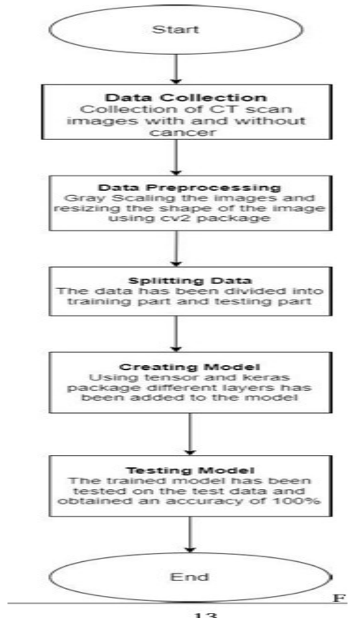
Augmentation done for images



DOI: 10.48175/IJARSCT-17535

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

Workflow of the Project



Developing the CNN Model

Building a complex dataset was done and the next phase of the project was to build a model with promising accuracy. Figure 3.2.1 shows the base architecture followed for this project. The flow or the sequence of the layers would be the same but the number of layers of batch normalization, dropout layers and kernel sizes, number of filters were twitched and tested for optimization.

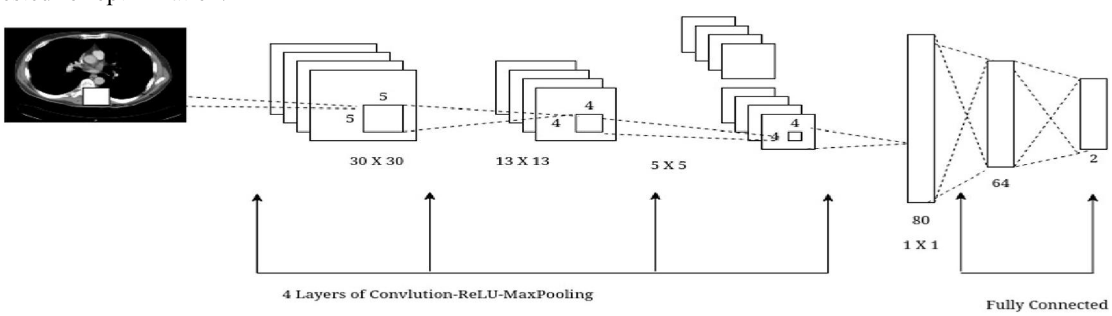


Fig CNN Model

Base Architecture The problem with the base architecture was, it did not have enough and appropriate bath normalization, dropout layers which either lead to overfitting or very poor performance of the model. So, on this architecture, more layers were built according to the data and the performance got better. After lots of changes to the base architecture, the architecture in Figure 5 was the first one to cross 90% accuracy. It consists of 2 layers of convolution and max pooling combination with batch normalization layers after every complex operation layer. Dropout layers were also added after flatten operation. Batch normalization helps in increasing the stability of the neural network. It takes in the output of the previous activation layer and normalizes it by subtracting the mean of batch and dividing by the standard deviation of batch. Dropout helps in preventing the model from overfitting. It works by taking random neurons and killing it that is by setting the output of neurons to 0 at every update that happens during

Evaluating the Model

The final model performed very well on our developed dataset. Now we can choose any image of CT scan and predict easily whether a person is having lung cancer or not with our proposed/developed algorithm. The final output consists of

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

2 classes i.e. true or false. If the output is true that means the given CT scan image consists of cancer. If the output is false that means the given CT scan image doesn't consist of cancer.

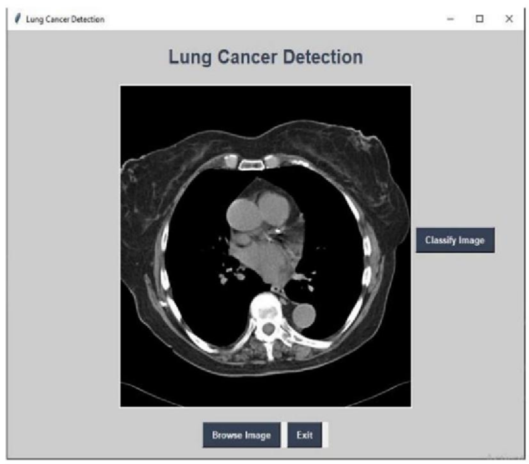
Graphical User Interface

The next and the final phase of the project is to build a graphical user interface. Home Screen: The user can either upload an image or close the application.

Browse Image: It directs you to the page where you can choose to select an image.

Classify Image: After uploading the image, we can classify the image weather the selected CT scan image consists of lung cancer or not.

Exit: To close the application.



IV. FUTURE WORK

In the current study we have done only predicting the lung cancer future we can improve our project by locating the place of the cancer i.e. to detect the place where the cancer is present/located in the lung.

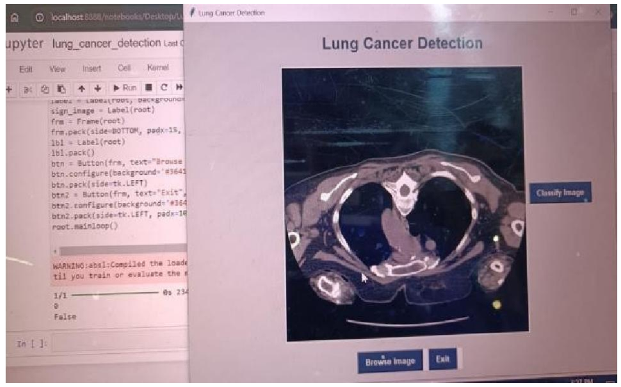
V. CONCLUSION

In conclusion, The project focuses on leveraging machine learning algorithms to identify the patient whether having lung cancer or not. The user provides the input lung CT image. The trained dataset checks the input such as image and shows the output that whether he/she has lung cancer or not. This predictive capability can be valuable for health care department. Unlike the existing, it can detect cancer at earlier stages.intrusions with greater accuracy and adaptability.

VI. RESULT

To identify the patient whether having lung cancer or not, we designed a graphical user interface (GUI). In this, the user easily gets to know the details of the CT image. The user provides the input lung CT image. The trained dataset checks the input such as image and shows the output. Whether he/she has lung cancer or not. By training the dataset of the lung computed tomography (CT) images, we are getting an accuracy of 100%. To detect lung cancer in people, a hybrid model comprising several machine learning classifiers and CNN models is utilised. AlexNet is combined to the SoftMax layer to get the most efficient results.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024



REFERENCES

[1]. C. Ridge, A. McErlean, and M. Ginsberg, “Epidemiology of Lung Cancer,” Semin. Intervent. Radiol., vol. 30, no. 2, pp. 093–098, May 2013 https://scholar.google.com/scholar?q=C.%20Ridge%2C%20A.%20McErlean%2C%20and%20M.%20Ginsberg%2C%20%E2%80%9CEpidemiology%20of%20Lung%20Cancer%2C%E2%80%9D%20Semin.%20Intervent.%20Radiol.%2C%20vol.%2030%2C%20no.%202%2C%20pp.%20093%E2%80%93098%2C%20May%20 2013
http://dx.doi.org/10.1109/CITDS54976.2022.9914128 .

[2]. K. Nakagomi, A. Shimizu, H. Kobatake, M. Yakami, K. Fujimoto, K. Togashi “Multi-shape graph cuts with neighbor prior constraints and its application to lung segmentation from a chest CT volume https://www.sciencedirect.com/science/article/abs/pii/S1361841512001090

[3]. A. Kalinovsky and V. Kovalev, “Lung Image Segmentation Using Deep Learning Methods and Convolutional Neural Networks,” 2016. https://scholar.google.com/scholar?q=A.%20Kalinovsky%20and%20V.%20Kovalev%2C%20%E2%80%9CLung%20Image%20Segmentation%20Using%20Deep%20Learning%20Methods%20and%20Convolutional%2 0Neural%20Networks%2C%E2%80%9D%202016.

[4]. Lung ct image classification- lBrahim Ait Skourt, Abdelhamid El Hassani, Aicha Majda https://www.google.com/search?q=Brahim+AIT+SKOURT%2C+Abdelhamid+EL+HASSANI%2C+Aicha+MAJDA+“Lung+CT+Image+Segmentation+Using+Deep+Neural+Networks+“Volume+127%2C+2018%2C+ Pages+109- 113.&oq=Brahim+AIT+SKOURT%2C+Abdelhamid+EL+HASSANI%2C+Aicha+MAJDA+“Lung+CT+Ima ge+Segmentation+Using+Deep+Neural+Networks+“Volume+127%2C+2018%2C+Pages+109- 113.&aqs=chrome..69i57.978j0j7&sourceid=chrome&ie=UTF-8

[5]. Kamal U, Rafi AM, Hoque R, Hasan M. Lung cancer tumor region segmentation using recurrent 3D-DenseUNet. https://www.google.com/search?q=Kamal+U%2C+Rafi+AM%2C+Hoque+R%2C+Hasan+M.+Lung+cancer+ tumor+region+segmentation+using+recurrent+3D- DenseUNet.+arXiv+preprint+arXiv%3A1812.01951.+2018+Dec+5&oq=Kamal+U%2C+Rafi+AM%2C+Hoq ue+R%2C+Hasan+M.+Lung+cancer+tumor+region+segmentation+using+recurrent+3D- DenseUNet.+arXiv+preprint+arXiv%3A1812.01951.+2018+Dec+5&aqs=chrome..69i57.930j0j7&sourceid=c hrome&ie=UTF-8iv:1812.01951. 2018 Dec 5 - Google Search
https://academic.oup.com/bjr/article-abstract/80/960/996/7443920?redirectedFrom=fulltext&login=false

[6]. U-Net: Convolutional Networks for Biomedical Image Segmentation | SpringerLink https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28?utm_source=getftr&utm_medium=getftr&utm_campaign=getftr_pilot