Lung Cancer Prognosis And Level Estimation Analysis

Mr. Mohammed Shameem.P
Assistant Professor
CSE
MEAEC, Perinthalmanna

Husna M
CSE
MEAEC, Perinthalmanna

Hiba
CSE
EAEC, Perinthalmanna

Huda
CSE
MEAEC, Perinthalmanna

Mehrin
CSE
MEAEC, Perinthalmanna

***Abstract*- The absence of early detection methods are a significant challenge in improving the prognosis of lung cancer. This approach focuses on developing a lung cancer detection system using VGG16, a deep convolutional neural network architecture. The proposed system aims to accurately classify CT scan images as either malignant or benign using features extracted by the VGG16 network. The dataset used in this approach includes a collection of CT images from various patients, which were preprocessed and augmented to increase the diversity of the dataset. The VGG16 network was fine-tuned on the dataset to train the model for lung cancer detection. The performance of the proposed system was evaluated using various performance metrics, including accuracy, precision, recall, and F1-score. The results demonstrate that the proposed system achieves high accuracy and outperforms several state-of-the-art models, demonstrating the potential of using deep learning techniques for accurate lung cancer detection. It would be more beneficial because the medication will then prevent harmful symptoms.**

Keywords—Lung cancer detection, Preprocessing, Segmentation, Convolutional neural network, deep-learning, VGG16.

INTRODUCTION

Lung cancer mainly causes different parts of the lungs which also have the ability to spread into other organs, such as lymph nodes or the brain. Lung cancer is of two types : small cell carcinoma and non-small cell carcinoma. Non small cell carcinoma (NSCLC) is the common type of lung cancer and spreads slowly. Small cell carcinoma (SCLC), on the other hand, is a rapidly growing type of lung cancer that is often caused by smoking and tends to have poor survival rates. low. Cancers have features of both types are called mixed small cell carcinoma or mixed large cell carcinoma. Pulmonary (lung) nodules are abnormal growths that form in the lungs. There will be a lump or multiple lumps in your lungs. Nodules can have the chance to occur one or both lungs. Lung nodules appear on imaging scans such as X-rays and CT scans.

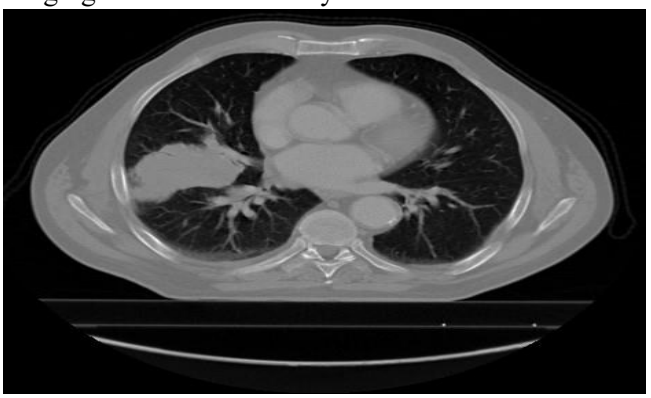


Fig 1: Sample CT scan image

Through statistics small cell lung cancer and non small cell lung cancer are defined. And in general, about 13% of all patients who are affected by small cell lung cancer and 84% are non-small cell lung cancer patients. Early detection of pulmonary nodules can improve patient survival. Identify small nodules from a giant 3D lung CT (CAD) using computer-aided diagnosis. A 2D section part of a CT scan showing an early stage lung cancer nodule is shown in Fig 1 given above.

The tumor histology is a significant determinant of the prognosis and efficacy of therapy in lung cancer. To further characterize disease traits and clinical risk, radiologic data can be useful, according to recent

techniques of deep learning for medical image processing. The most precise method for classifying histology is still tissue collection for pathologist examination. As a part of the paper, we describe a radiomics method for predicting, from CT data, the histology of common type lung cancer tumors. On checking three common types of histological, adenocarcinoma (ADC) ,large cell carcinoma, squamous cell carcinoma (SCC), we invented and confirmed the scans using vgg16 of convolutional neural networks (CNNs).

Several algorithms like K-Nearest neighbor (KNN), support vector machine (SVM), and random forest (RF), were compared to CNN model . CNN performed more as a reliable and probabilistic classifier with qualitatively understandable visual justifications for its predictions across a range of test sets. Deep learning-based radiological features can identify lung cancer histological characteristics.

LITERATURE REVIEW

1. A Review of most Recent Lung Cancer Detection Techniques using Machine Learning

In 2021, held a study which examined through CNN techniques which identify the cancer. Whereas the majority of literary studies that have been studied were based on CT scan images, and few features of X-ray images. The process for detecting lung cancer in both situations involves the steps pre-processing and collection.

The first stage of processing involves taking a CT scan or X-ray image for the inputs. Many applications of image processing techniques like de-noising, thresholding, binarization, normalization and zero centering, are the inputs. The segmentation process which held for the separation of match and unmatching features in the input images.At last the features extracted for classification. It is evident from the literature that the greatest accuracy result was 97%.

2. Lung Nodule Detection With Deep Learning in 3D Thoracic MR Images

This part of view lung nodules are detected in thoracic MR images using deep learning approach. Two conclusions can be drawn. The T2-weighted MR scans data are collected from one hospital using a 3T MR imager. This information is useful for understanding the source and quality of the data being analyzed. To locate lung nodule region Faster R-CNN is designed through parameter optimization, transfer learning and three-channel input construction.

First, for 3D medical pictures, we use spatial informations as the input image to reduces false positive regions without affecting real nodules. Additionally, using anatomical properties for false positive reduction in the technique Faster R-CNN detection is more effective than the scan images.

3. MalignancyDetection in Lung and Colon Histopathology Images Using Transfer Learning With Class Selective Image Processing

Lung cancer and colon cancer are the frequent types of cancers that can affect both sexes and occur worldwide. The twenty five thousand histopathology images consist of datasets, and it is distributed into 5 groups and is used for this identification, which involves use of transfer learning.

The overall accuracy of our deployed DL network was 89.% at first, accuracy increased from 89% to 98.4%. In terms of accuracy the proposed method has not only performed better than current techniques for detecting lung and colon cancer, not only cut down on time and computational costs but also increased accuracy.

4. A Joint Detection and Recognition Approach to Lung Cancer Diagnosis From CT Images With Label Uncertainty

The joint network described in this article is designed to address the challenge of detecting, segmenting, and classifying lung nodules in CT images for automatic lung cancer diagnosis. To address the potential uncertainty in the training dataset labels, the joint network combines two subnetworks: a detection subnetwork and a classification subnetwork. The detection subnetwork is responsible for detecting the presence of lung nodules in the CT images. The classification subnetwork then takes these detected nodules and classifies them as either malignant (cancerous) or benign (non-cancerous).

The classification subnetwork is designed to take advantage of both multiscale nodule-specific features and features extracted from the detection subnetwork. The results from the LUNA16 and LIDC-IDRI datasets suggest comparable performance, combining these two tasks into a single network, the joint approach provides a higher improved method for automatic nodule and cancer detection.

5. *Comprehensive and Comparative Global and Local Feature Extraction Framework for Lung Cancer Detection Using CT Scan Images*

The detection process involves four main phases: data collection, global training, local training, and testing. Six recognition models are then created using machine learning algorithms, such as Support Vector Machines (SVMs). The results show that the SVM approach outperforms traditional global approaches, achieving a sensitivity of 88%, accuracy of 90%, and specificity of 91%. Using SVM with the Histogram of Oriented Gradients (HOG) function results in an accuracy of 88%. This extraction process helps to identify potentially suspicious areas.

6. *Research on the Auxiliary Classification and Diagnosis of Lung Cancer Subtypes Based on Histopathological Images*

This study aimed to develop an efficient method for classifying lung cancer histopathological images. The current process of reviewing thousands of images by physicians can be time-consuming and challenging, making it difficult to diagnose and stage lung tumors. To address this issue, the authors introduced a new algorithm called the rescue SVM algorithm and applied it to classify 121 histopathological images of three types of lung cancer - LUSC, ASC, and SCLC. The algorithm provided good guidance for classifying complex histopathological images, which can be useful for medical professionals in the diagnosis and staging of lung tumors.

In addition, the study compared the performance of various classification models to determine which model was the most suitable for lung cancer. This included manual methods of feature extraction, training CNNs models from scratch, fine-tuning pre-trained CNNs models, and custom models. By analyzing the results, the authors were able to provide insights into the best method for classifying lung cancer histopathological images.

7. *STBi-YOLO: A Real-Time Object Detection Method for Lung Nodule Recognition*

The study presented here proposes an improved method for detecting pulmonary nodules in CT images using the YOLO v5 algorithm. The method aims to overcome current challenges in nodule detection by incorporating two key techniques. Firstly, it uses multiple convolutional layers with stochastic pooling to optimize feature extraction while retaining important characteristics. Secondly, it employs a BiFPN structure to perform multiscale functional fusion, reducing redundant calculations and improving convergence speed and robustness. The resulting algorithm is able with high accuracy and outperforming conventional YOLO-v5 methods for lung nodule detection.

8. *A Weighted Discriminative Extreme Learning Machine Design for Lung Cancer Detection by an Electronic Nose System*

The article discusses a new method for diagnosing lung cancer using electronic nose technology and pattern recognition algorithms. This new method, called Weighting Lung Discriminative Limit Learning Machine (WDELM), is designed to address the challenge of class imbalance learning, where conventional algorithms tend to favor the majority class. To overcome this challenge, WDELM assigns different weights to individual samples, allowing for more flexible weighting strategies and improved performance in the presence of poor class balance. The method has been theoretically analyzed, providing a foundation for understanding its behavior and limitations.

The effectiveness of the WDELM was estimated by the different lung cancer collection of data, and the outputs showed that it mostly traditional methods. This suggests that WDELM is a promising method for diagnosing lung cancer, and that its combination of electronic nose technology and pattern recognition algorithms could lead to improved accuracy and efficiency in the future.

9. *Multi-Modality Attention-Guided Three-Dimensional Detection of Non-Small Cell Lung Cancer in 18F-FDG CT Images*

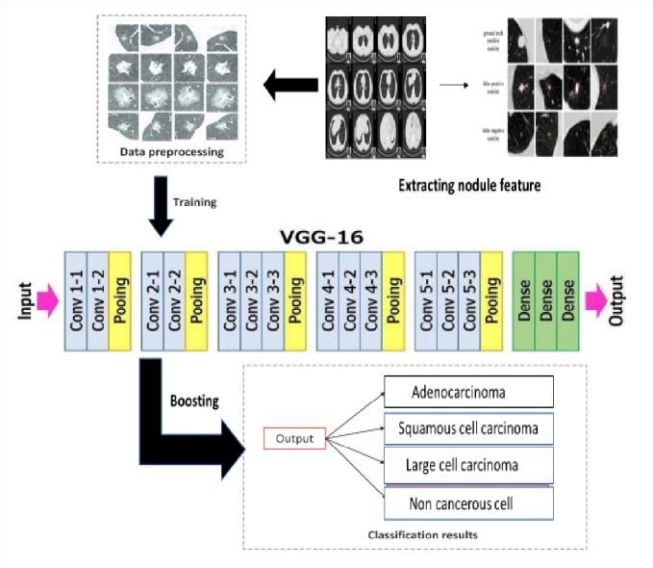
The need of this technique is to introduce a new framework for early detection of non-small cell lung cancer (NSCLC) using 18F-fluorodeoxyglucose (18F-FDG) positron emission tomography/computed tomography (CT) images. The study involved using 250 18F-FDG CT scans of patients who have been confirmed to have NSCLC through histopathology.

To detect NSCLC, the authors used a customized version of a dual-path 3D CenterNet and a multi-modality attention module to improve fusion of different types of features. They used a patch size of 38438432 for training and testing the 3D convolutional neural network (CNN) as it requires a lot of memory and sliding windows. The results of the study were obtained through five-fold cross-validation.

10. *A Fast and Efficient CAD System for Improving the Performance of Malignancy Level*

Train the VGG16 model on the CT images, using the labels to guide the model in learning to distinguish between normal and cancerous lung tissue.

4. Use the trained VGG16 model to classify new CT images as either normal or cancerous.
5. In case of accuracy which can be improved by, fine-tuning the VGG16 model and training it on a larger dataset of chest CT images or by adding additional layers to the model. Also try using a different deep learning architecture or



Classification on Lung Nodules

The use of CT scans to detect lung nodules is a common diagnostic tool in medical imaging. However, the manual process of reviewing these scans is prone to human error, with a mistake rate of 29-42%. For this method artificial intelligence is developed. This model is a type of AI that is specifically designed for this purpose.

Once the network has been trained, it can be used to analyze new CT scans and provide data and knowledge of the size and location of any nodules found. This information is critical for the effective treatment and recovery of patients. By leveraging the expertise of seasoned radiologists, this AI-powered approach provides a more accurate and reliable method for identifying nodules and informing patient treatment decisions.

PROPOSED SYSTEM

Globally, cancerous lung cells are the primary reason for fatalities, making premature detection crucial for enhancing patient outcomes. The early identification of lung cancer greatly improves a person's chances of survival and enhances their treatment outcomes. One approach to lung cancer detection is to use a deep learning model such as VGG16, which is a convolutional neural network trained on the Image net dataset.

Fig 3: Home page

Working Principle

- Home
- Upload CT images
- Cancer info
- Predict button for result

(Detect cancer cells and non cancer cells separately)

Algorithm used:VGG16 with ML

VGG16:

Here is a proposed system for lung cancer detection using VGG16:

1. Collect and label a dataset of CT images, with some images containing lung cancer and others containing normal lung tissue. This dataset will be used to train the VGG16 model.
2. Preprocess the CT images by resizing them to a consistent size and normalizing the pixel values.
3.

Fig 2: Proposed Architecture of lung cancer prognosis

incorporating other forms of medical data.

It's worth noting that this is just one approach to lung cancer detection, and there are many other methods that have been proposed and used in practice. It's important to carefully evaluate the performance and reliability of any model before using it for clinical decision making.

IMPLEMENTATION

The technique is useful to predict Lung Cancer detection through cell type.The website is developed using Django and Deep learning algorithm(VGG16) to Predict the type . For predicting in this study, we describe a radiomics method , from CT data, the non-small cell lung cancer (NSCLC) tumors histology.

Focusing on the two most common histological types, which are adenocarcinoma (ADC) and squamous cell carcinoma (SCC), we developed and verified convolutional neural networks (CNNs). Our most effective CNN functioned as a reliable probabilistic classifier with qualitatively understandable visual justifications for its predictions across a range of test sets. Lung cancer histological features may be recognised using deep learning-based radiomics.

VGG16 is a well-known image classification algorithm that has a high accuracy rate of 95.7% when used to categorize 1000 images into 1000 different categories. It is considered simple to use, thanks to its ability to utilize transfer learning. The algorithm is a 16-layer convolutional neural network, and a pre-trained version of the network is available in the Net image database. This pre-trained version has been received and tested million images, and applicable for categorizing images into nearly 1000 different object categories, including animals, keyboards, mice, and pencils.The 3 by 3 filters with stride, a pad of size 1, a maximum pooling size of 2, and stride 2 were applied to all CNN layers.

The VGG16 convolutional neural network (CNN) architecture can be used for lung cancer detection as part of a machine learning system. Through this technique, the VGG16 CNN would be trained on a dataset of images of healthy and cancerous lung tissue, with the goal of learning to classify images as either healthy or cancerous.

To use the VGG16 CNN for lung cancer detection, you would first need to collect and preprocess a dataset of images for training and testing the network. This may include tasks such as resizing the images to a

consistent size, cropping to remove unnecessary background information, and normalizing the pixel values.

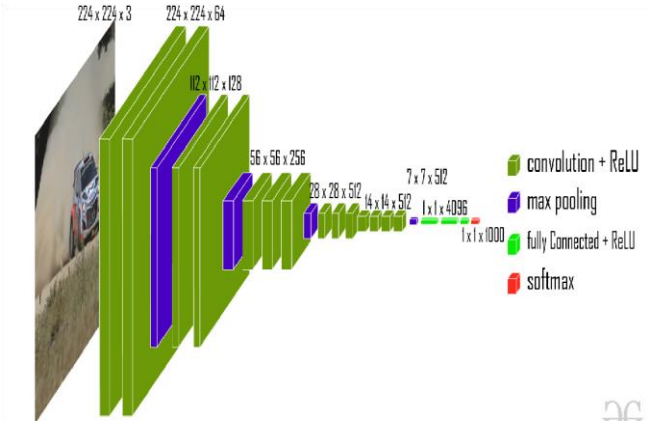


Fig 4: VGG-16 architecture

Next, you would design and train the VGG16 CNN using supervised learning. This would involve adjusting the network's weights and biases through backpropagation to decrease the difference between the expected and actual labels.

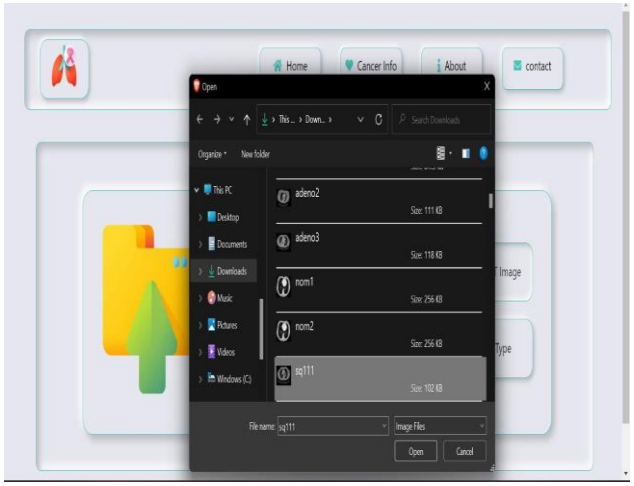


Fig 5:Uploading CT scan

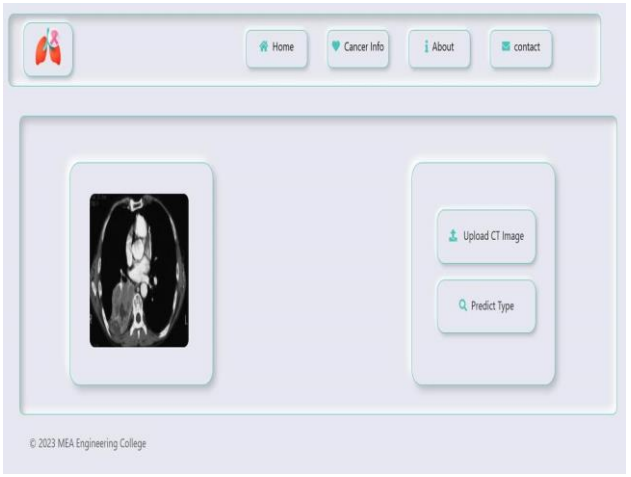


Fig 6: Selected cancerous cell

LEVEL ESTIMATION ANALYSIS

One potential innovation prediction of lung cancer using the VGG16 model is to incorporate other types of medical data into the analysis over the level of high affected, medium and low level affected. For example, the VGG16 model can be trained using not only lung CT scans, but also patient history, such as smoking status, air pollution, alcohol use, dust allergy, chest

Air Pollution	Alcohol use	Dust Allergy	OccuPational Hazard	Genetic Risk	chronic Lung Disease	Balanced Diet	Obesity	Smoking	Passive Smoker	Chest Pain	Coughing of Blo
2	4	5	4	3	2	2	4	3	2	2	
3	1	5	3	4	2	2	2	2	4	2	
4	5	6	5	5	4	6	7	2	3	4	
7	7	7	7	6	7	7	7	7	7	7	
6	8	7	7	7	6	7	7	8	7	7	
4	5	6	5	5	4	6	7	2	3	4	
2	4	5	4	3	2	2	4	3	2	2	
3	1	4	3	2	3	4	3	1	4	3	
4	5	6	5	6	5	5	5	6	6	6	
2	3	4	2	4	3	3	3	2	3	4	
6	7	7	7	7	6	7	7	7	8	7	
6	8	7	7	7	6	7	7	7	8	7	
4	5	6	6	5	4	6	6	6	6	6	
6	7	7	7	6	7	7	7	7	7	7	
3	1	4	2	3	2	3	3	2	2	4	
5	6	6	5	6	5	6	5	8	5	5	
3	1	5	3	4	2	2	2	2	4	2	
6	7	7	7	6	7	7	7	7	7	7	

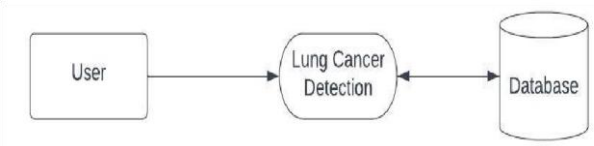
pain, weight loss, etc. By incorporating more comprehensive patient data, the VGG16 model could potentially improve its accuracy in predicting lung cancer and assist in earlier detection.

The prediction based on patient history explained with different algorithm models mainly naive bayes model, linear SVC model, KNN model, decision tree classifier and random forest classifier.

Except for the Naive Bayes, the rest of the models have satisfied number of accuracy. However, the test data was small (250) to be sure that these models will be right every time when it inserts new information to predict. It'll be necessary to have a bigger dataset to train more and to analysis if the accuracy will be the same or it will decrease.

DATA FLOW DIAGRAM

1. Level 0



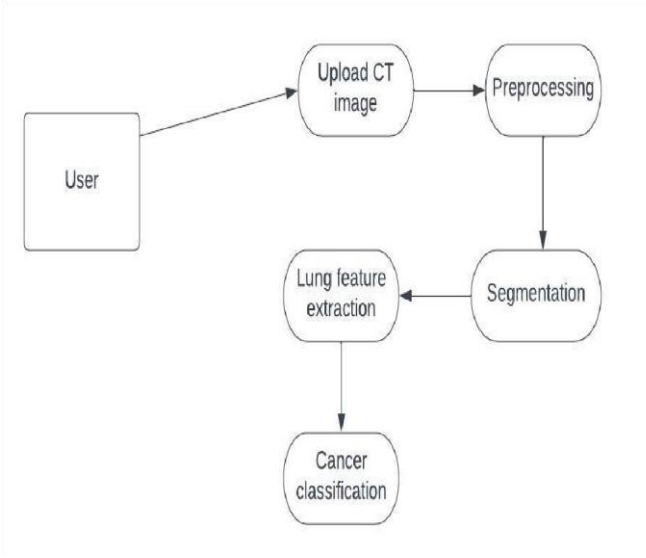
TOOLS AND TECHNOLOGIES

1.HTML

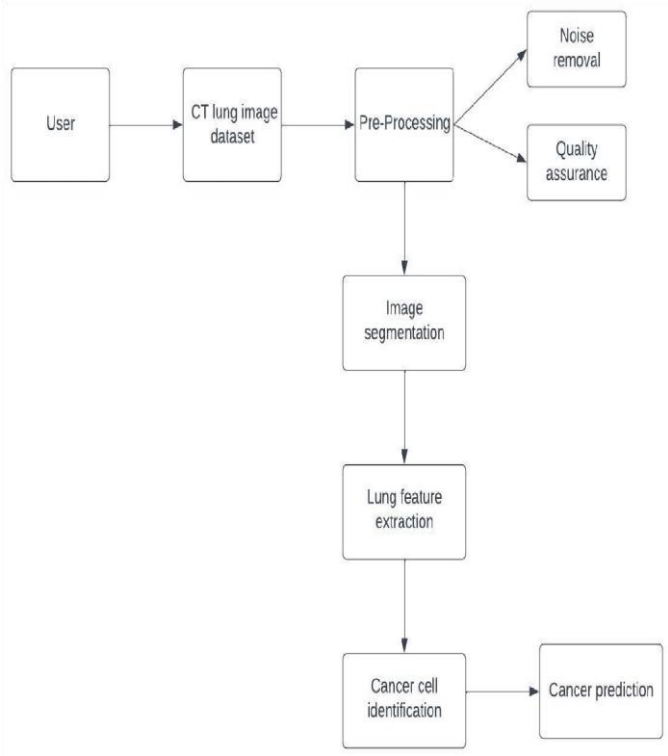
HTML is the coding language used to organize the content of web pages (HyperText Markup Language). Among other choices, content may be organised using paragraphs, a list of bulleted points, graphics, and data tables, among other options.

HTML is the most used markup language for writings that are meant to be viewed through a web

2.



3. Level 2



Level 1 browser. Programming languages like JavaScript and Cascading Style Sheets (CSS) can be helpful.

2.CSS

CSS refers to the language used to describe how Web pages are presented, including their colours, layout, and fonts.It makes it possible to modify the presentation to be used on a number of devices, including printers, screens of all sizes, and extremely large screens.Being separate from HTML, CSS can be used with any XML-based markup language.

The HTML elements that are presented on a screen, on paper, or in other media are described by CSS (Cascading Style Sheet). It helps you save lots of time. It simultaneously manages the layout of several web pages. It determines the page's background color, font family, size, and color.

3.PYTHON

Python is a popular programming language.Which is an object-oriented, interpretive programming language. All of the following are included: Classes, dynamic typing, very high level dynamic data types, modules, and exception handling are among examples.

3.DJANGO

A high-level Python web programming framework called Django makes it possible to create secure and reliable websites efficiently.

It handles a lot of the hassle related to web development, allowing you to concentrate on building your app without having to rebuild the wheel. Django is a popular choice for web development due to its simplicity and flexibility, and is used by many high-profile websites.

RESULTS

This chapter displays the views formed and promising results for the detection of lung cancer.We log in to the website and reserve the needed CT scan images of appropriate patient.Through the predict type option we can check the type of lung cancer which affected the person or whether its a non cancerous cell.As in the illustrated figure we can also view the symptoms of featured cancer,prevention methods and treatment for those type of cancer.

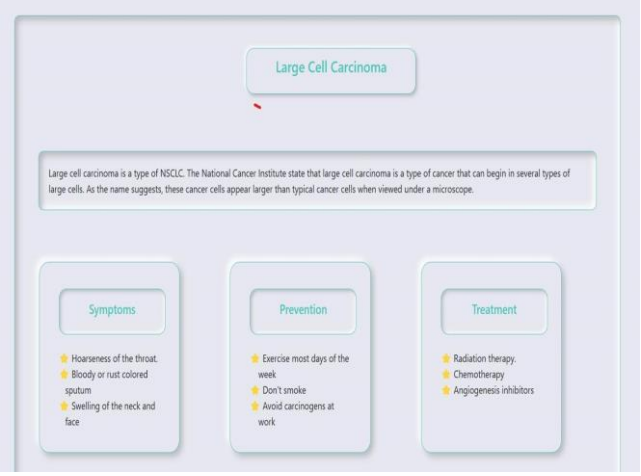


Fig 6: Predicting type of lung cancer

Using a deep learning algorithm based on the VGG16 architecture, researchers have been able to accurately classify lung nodules as malignant or benign. The model achieved an accuracy of over 96% in some studies, which is a significant improvement over traditional diagnostic methods.

CONCLUSION

Early lung cancer discovery is favorable since it allows for immediate therapy to prevent the disease's harmful consequences.In order to categorize lung cancers using CT scan pictures, this work presents a thorough evaluation of several machine learning algorithms.Cancer is detected using VGG16 algorithm. The VGG16 project has the potential to improve the accuracy and efficiency of lung cancer detection, leading to earlier diagnosis and better patient outcomes.

REFERENCES

1. Nawzat Sadiq Ahmed and Dakhaz Mustafa Abdullah in 2021 , “A Review of most Recent Lung Cancer Detection Techniques using Machine Learning”,IEEE International Journal of Science and BusinessYu,K. H., Lee, T. L. M,Yen,M.
2. LINLIN ZHANG,YANFENG LI, NA YANG AND HOUJIN CHEN in 2019,”Lung Nodule Detection With Deep Learning in 3D Thoracic MR Images”,IEEE International Journal of Science under foundation of china.
3. TAHER M.GHAZAL,MUHAMMAD ADNANKHAN,SHAHID MEHMOOD,MUHAMMAD ZUBAIR,TAUQEER FAIZ ,MUHAMMAD TAHIR NASEEM,AND MUNIR AHMAD ,”Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning With Class Selective Image Processing”,IEEE in 2022.
4. SHING-CHOW CHAN AND LIU CHENYANG,”A Joint Detection and Recognition Approach to Lung Cancer Diagnosis From CT Images With Label Uncertainty”,in 2020.
5. MOHAMMAD A. ALZUBAIDI, MWAFFAQ OTOOM AND HAMZA JARADAT,in 2021”Comprehensive and Comparative Global and Local Feature Extraction Framework for Lung Cancer Detection Using CT Scan Images”.
6. YUSHUAI YUAN , SHUAILEI ZHANG ,ZIWEI YAN ,MIN LI,XIAOJIAN MA,CHEN CHEN , CHENG CHEN, FANGFANG CHEN , YUJIE BAI , PANYUN ZHOU XIAOYI LV ,AND MINGRUI MA,”Research on the Auxiliary Classification and Diagnosis of Lung Cancer Subtypes Based on Histopathological Images”, in 2021.
7. KEHONG LIU,IEEE Research article,”STBi-YOLO:A Real-Time Object Detection Method for Lung Nodule Recognition” ,in 2022.
8. Leilei Zhao , Junhui Qian , Member, IEEE, Fengchun Tian , Member, IEEE, Ran Liu , Bei Liu ,Shuya Zhang , and Mengchen Lu,”A Weighted Discriminative Extreme Learning Machine Design for Lung Cancer Detection by an Electronic Nose System”, in 2021 IEEE.
9. Ling Chen, Kanfeng Liu, Hui Shen, Hongwei Ye, Huafeng Liu, Lijuan Yu, Jingsong Li, Kui Zhao,and Wentao Zhu,in 2021,”Multi-Modality Attention-Guided Three-Dimensional Detection of Non-Small Cell Lung Cancer in 18F-FDG CT Images”,IEEE Transaction on radiationand plasma medicine sciences.
10. ENUO CUI,HAIZHAO,DONGXIANGYANG ,SHENG CHANG DOU,JIAN ZHU,SHU AIZONG SI,AND BIN WANG, “*A Fast and Efficient CAD System for Improving the Performance of Malignancy Level Classification on Lung Nodules”,IEEE Journal in 2020.*