



## LUNG CANCER PREDICTION WITH CT SCAN USING MACHINE LEARNING

<sup>1</sup>Mr. Siju V Soman, <sup>2</sup>Ms. Akanksha, <sup>3</sup>Ms. Akanksha U Saadhana, <sup>4</sup>Ms. Arva Fathima, <sup>5</sup>Ms. Nazila Hamad

<sup>1</sup>Assistant Professor, <sup>2</sup>Student, <sup>3</sup>Student, <sup>4</sup>Student, <sup>5</sup>Student

<sup>1</sup>Department of Computer Science and Engineering,

<sup>1</sup>A.J Institute of Engineering and Technology, Mangaluru, India

<sup>2</sup>A.J Institute of Engineering and Technology, Mangaluru, India

<sup>3</sup>A.J Institute of Engineering and Technology, Mangaluru, India

<sup>4</sup>A.J Institute of Engineering and Technology, Mangaluru, India

**Abstract:** Lung cancer is a serious disease that can be life-threatening if not diagnosed and treated in time. Early detection is crucial for improving patient survival rates. To achieve this, researchers are using advanced technologies such as image processing and machine learning to analyze medical images for the early detection and treatment of lung cancer. In this study, the researchers propose an automated approach that uses computed tomography (CT) images to identify lung cancer in its early stages. The main objective is to achieve high accuracy in lung cancer diagnosis. The proposed framework involves several steps, including image enhancement, filtering, segmentation, feature extraction, and support vector machine. The experiment results show the accuracy performance of the proposed method, indicating its potential to improve lung cancer diagnosis and treatment.

**Index Terms --** Lung Cancer, CT scan, Machine Learning, Convolution Neural Network, Image processing.

### I. INTRODUCTION

Over the past few decades, lung cancer has surpassed breast, prostate, and colon cancer combined as the top cause of death worldwide. Healthcare experts are continuously working to increase the effectiveness and efficiency of healthcare delivery, and the application of machine learning techniques in the understanding and treatment of lung cancer is progressing. Electronic health records offer a plethora of patient data that can be used to enhance lung cancer early diagnosis and improve patient prognosis. The grade of lung cancer depends on the kind of cells that make up the tumor and can be either malignant or benign. The 3D imaging technique known as computerized tomography (CT) is frequently employed for diagnosing and screening lung cancer. contemporary 3D reconstruction. Lung Cancer is a deadly disease that has become a leading cause of death worldwide over the last few decades, surpassing breast, prostate, and colon cancer combined. The use of machine learning techniques in the understanding and treatment of lung cancer is advancing, and healthcare professionals are constantly striving to improve the efficacy and efficiency of healthcare delivery. Electronic health records provide a wealth of patient information, which can be used to improve early detection of lung cancer and positively impact patient prognosis. Lung cancer can be either malignant or benign, and is graded based on the type of cells that form the tumor. Computerized Tomography (CT) is a 3D imaging modality that is widely used for lung cancer screening and diagnostics. Advanced 3D reconstruction techniques have been developed to improve image quality and diagnostic accuracy. Primary lung cancers originate from cells within the lung, while secondary cancers (metastatic) begin in another part of the body and spread to the lungs. Steps like feature extraction and dimensionality reduction are essential for lung cancer detection from CT images to increase accuracy. It's interesting to think about how the photos have unobtrusive personalities. To make the data more manageable, pertinent details are either chosen or filtered out during feature selection. The collected characteristics are then used to train and test a classifier model. Lung cancer classification models have been developed using a variety of techniques, including support vector machine (SVM), k-nearest neighbor (k-NN), and artificial neural network (ANN), but these techniques can be expensive and may not catch the disease in its early stages, when the survival rate is low. Hence, a technique that can identify malignant cells at an early stage is required to assure

To lower mortality rates, it is essential to find lung cancer early. Although many different lung cancer prediction techniques have been put forth over the past few decades, medical image analysis needs special consideration because of the potential influence it could have on people's lives. For feature selection, which removes superfluous information from the data to make it simpler before feeding it into a classifier model for training and testing, non-obtrusive features in CT scans are particularly intriguing. There is an urgent need for a technique that can identify malignant cells at an earlier stage because early detection of cancer is crucial for effective therapy. Thankfully, new learning algorithms and network topologies for medical image analysis have been presented with the development of deep learning techniques. Convolutional Neural Networks (CNNs) have shown to be the most effective among these.

Principal

A.J. Institute of Engineering & Technology  
Mangaluru - 575 006

## II. PURPOSE OF THE PROJECT

Lung cancer has repeatedly shown up as one of the most lethal diseases in all of human history. Additionally, it is one of the malignancies that causes the most deaths and is the most prevalent. Lung cancer cases are rising quickly. In India, there are roughly 70,000 instances per year. The disease is frequently asymptomatic in its early stages, making detection practically impossible. Because of this, early cancer identification is crucial to preserving lives. A patient may have a better chance of recovery and cure with an early diagnosis.

## III. NEED OF THE PROJECT

It highlights the significant impact of lung cancer on global health, as well as the economic burden associated with its diagnosis and treatment. Lung cancer is a serious disease that affects people of all ages and can be difficult to diagnose using existing clinical techniques such as X-ray and other imaging procedures. One of the biggest challenges in lung cancer diagnosis is the need for accurate and reliable prediction methods, which can be achieved using machine learning models. Machine learning models are more effective and cost-efficient than traditional clinical techniques, making them a promising tool for medical diagnosis. It also says that long-term tobacco smoking is the leading cause of lung cancer, accounting for 85 percent of cases. However, a significant percentage of cases also arise in people who have never smoked, which highlights the need for more advanced detection methods. Technology, including machine learning models, is playing an increasingly important role in detecting cancer efficiently. Many researchers have proposed different methods based on their studies to improve early detection rates and reduce the cost of treatment. In conclusion, it emphasizes the need for advanced and reliable prediction methods for lung cancer diagnosis. Machine learning models can offer a cost-effective and efficient way to achieve this goal, potentially improving outcomes for individuals with lung cancer.

## IV. EXISTING SYSTEM

Lung cancer, which has become a major contributor to mortality rates worldwide. Medical imaging, particularly Computer Tomography (CT) images, are commonly used to diagnose lung cancer. However, physicians face challenges in accurately interpreting and classifying cancerous lesions in CT scan images. To address this challenge, computer-based analysis tools have been developed using image processing approaches. These tools can assist physicians in the accurate interpretation and classification of cancerous lesions in CT scan images, improving the accuracy of diagnosis and treatment planning.

## V. PROPOSED SYSTEM

Machine learning algorithms and Deep learning algorithms are the two emerging techniques that have recently attracted many researchers. Deep learning techniques has also achieved tremendous success in computer vision. These techniques provide a uniform feature extraction classification framework to users and free them from troublesome handcrafted feature extraction. Deep learning techniques also provides the opportunity to increase the accuracy of the early detection of diseases. Proposed system uses the deep learning techniques, namely - convolutional neural network and recurrent neural network are utilized to propose a model for lung cancer diagnosis using CT scans and to obtain high accuracy.

## VI. SCOPE OF THE PROJECT

The paragraph discusses the characteristics of lung cancer, which is one of the most dangerous diseases in the world. Lung cancer is caused by abnormal cell growth in the lungs, which can be classified into two major types: non-small cell lung cancer and small cell lung cancer. Symptoms of lung cancer include difficulty breathing, weight loss, wheezing, and changes in the voice. Lung cancer can also spread to other parts of the body, which makes early detection important for successful treatment. Imaging techniques, such as computed tomography (CT) scans, can be used to detect lung tumors by providing detailed images of the lung structure. The CT scan can help identify tumors in the lungs by providing multiple slices of the lung structure, which allows physicians.

## VII. METHODOLOGY

A CT scan is a medical imaging technique that combines X-rays and computer technology to produce detailed images of the inside of the body. CT scans can provide visual information about bones, muscles, organs, blood vessels, and other internal structures. Image processing is a technique used to perform operations on images to extract useful information or enhance the quality of the image. Image processing involves a range of operations including image acquisition, image enhancement, image analysis, and image interpretation.

### BACKEND

#### Gather the dataset:

It is important to compile a dataset, which is essentially a collection of data items that can be processed as a single unit, in order to do analysis and prediction using a computer. In this instance, we are using CT scans to forecast lung cancer using photos of tumors from several patients and machine learning. The World Health Organization has given its blessing to several of the websites from where the datasets were gathered.

#### Import the modules:

Importing pertinent Python modules is crucial before moving on to the dataset analysis. The steps in the backend, as shown in Figure, include loading the required modules, preparing the data, creating the model, and determining prediction accuracy. Importing pertinent Python libraries is the first step before performing any analysis on the dataset. After that, the dataset is put into Jupyter Notebook and divided into 4:1 testing and training sets. It can be difficult to choose the best model to train the data. The 20% of the dataset designated for testing is used for testing after the training phase is complete.

#### Load the dataset and pre-processing:

The dataset must be divided into testing and training sets at a 4:1 ratio once it has been loaded into Jupyter Notebook. According to this, training takes up 80% of the dataset, and testing takes up the remaining 20%. The batch size hyperparameter is crucial for fine-tuning deep learning, and a size of 16 is typically suggested. In addition, data augmentation is a helpful method for expanding the training set. A data generator was developed to automate the collection of data from folders and its entry into Keras. For this purpose, Keras provides practical Python generator methods.

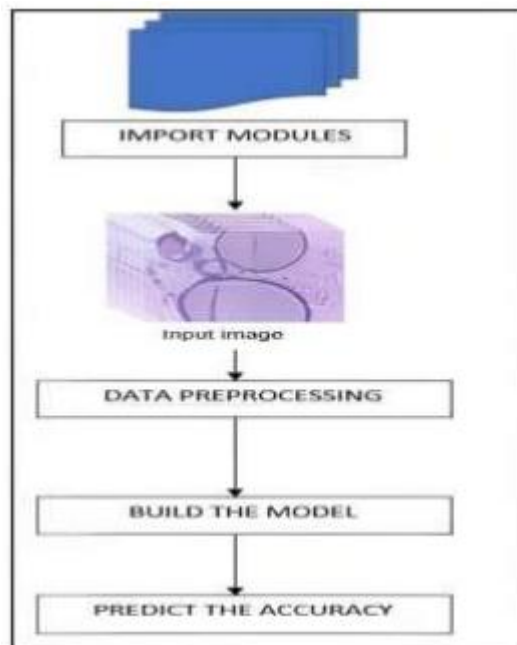


Figure: Stages of Backend

**Build the model:**

It can be difficult to choose the best model to train the dataset, which calls for exploratory data analysis. K-Nearest Neighbour, Support Vector Machine (SVM), and Convolutional Neural Network models are just a few examples of machine learning models that can be used to train the dataset.

**Predict the accuracy:**

After the training is finished, the trained models are tested using the 20% of the dataset that was designated for testing. Each model is assessed according to its accuracy score, which must be higher than 0.7 in order to pass the system's tests. Unit testing of the back end is carried out during this phase.

**FRONTEND**

The frontend consists of several stages, beginning with the importing of libraries such as Streamlit and TensorFlow. Next, a graphical user interface is created that is primarily aimed at non-technical users, such as medical students, oncologists, and physicians. The interface is developed using the Python programming language and includes images and graphics to enhance user experience. To design the front-end, the Streamlit and TensorFlow libraries are imported. In addition, connecting the frontend and backend is crucial to provide a better understanding of the system for non-technical users.

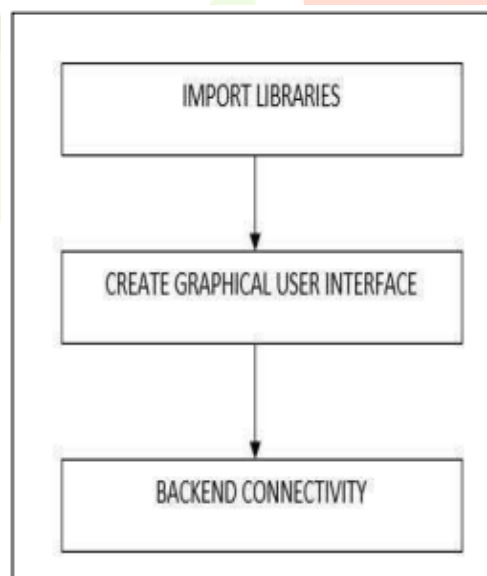


Figure: Stages of Frontend

**VIII. APPLICATION**

In order to increase the precision and effectiveness of lung cancer detection using CT scan pictures, machine learning has showed a lot of promise. In order to aid in the early diagnosis of lung cancer, several research have investigated the use of machine learning algorithms to analyze and categories CT scans of the lungs. By facilitating early identification and prompt care, these machine learning technologies for lung cancer detection seek to enhance patient outcomes. Researchers have created models that can precisely detect and categories lung nodules and lesions using vast datasets of CT images and cutting-edge algorithms, assisting in the detection and treatment of lung cancer.

## IX. ADVANTAGES

- A higher success rate in detecting cancer nodules than the top existing model.
- Determines whether the lung cancer discovered is malignant or benign.
- Eliminates the salt-pepper and irregular disturbances that lead to incorrect cancer detection.

## X. DISADVANTAGES

- End user should know the basic operations of a computers such as uploading an image.
- This system might take more time to predict depending on the target system.

## XI. CONCLUSION AND FUTURE WORK

In this effort, a novel automated computer-aided model for the detection of lung cancer on the used CT scans has been established. Several processes make up the model that is being shown. Following preprocessing of the input image, a number of significant features are recovered from the segmented image. Following that, photos will be categorized using the CNN classifier model, which will ultimately produce output as categorized images that can be categorized as "normal" or "abnormal." The primary goal of this work is to increase the prediction systems' accuracy. By assessing patients' symptoms and lifestyles using textual data, this research makes it simpler for doctors and other medical professionals to treat and diagnose patients with lung cancer. This analytical study's goal is to identify the main factors that either contribute to or are major signs of lung cancer. When it's necessary to find and comprehend the various patterns and principles concealed in large amounts of complex data, this approach can be quite helpful. Additionally, this has made it simpler for us to group patients according to the lifestyle decisions they need to maintain control over to prevent endangering their lives from lung cancer. There is scope for the future improvement of this undertaking. PC innovation continues to track down new strategies and advances on an everyday premise. It is dynamic and not static. In this way, it isn't finished up. However, it will improve with additional expansions. Increases should be possible successfully. Hence the venture is adaptable and can be improved whenever with additional ever-evolving highlights.

## REFERENCES

- [1] Atharva Banker et. Ed in "Symptom Analysis using a Machine Learning approach for Early-Stage Lung Cancer" 2019 International Conference on Opto- Electronics and Applied Optics (Optronics), Kolkata, India, 2019.
- [2] Qing Wu et. ed in "Small-Cell Lung Cancer Detection Using a Supervised Machine Learning Algorithm" 2017 International Symposium on Computer Science and Intelligent Controls (ISCSIC).
- [3] Anand Gupta et ed. in "Prediction of Lung Cancer from Low-Resolution Nodules in CT-Scan Images by using Deep Features" Published 1 September 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI).
- [4] Swati Mukherjee et. ed in "Lung Cancer Disease Diagnosis Using Machine Learning Approach" [https://www.researchgate.net/publication/350083397\\_Lung\\_Cancer\\_Disease\\_Diagnosis\\_Using\\_Machine\\_Learning\\_Approach](https://www.researchgate.net/publication/350083397_Lung_Cancer_Disease_Diagnosis_Using_Machine_Learning_Approach).
- [5] G. Niranjana Published 1 April 2017 Medicine 2017 International Conference on Technical Advancements in Computers and Communications (ICTACC).
- [6] Ahmed Shaffie in "On the Integration of CT-Derived Features for Accurate Detection of Lung Cancer"
- [7] Allison M Rossetto in "Deep Learning for Categorization of Lung Cancer CT Images.
- [8] James A. Bartholomai in "Lung Cancer Survival Prediction via Machine Learning Regression, Classification, and Statistical Techniques"
- [9] James A. Bartholomai in "Lung Cancer Survival Prediction via Machine Learning Regression, Classification, and Statistical Techniques" proposed a predictive model.
- [10] Dr. A. Nagarajan in "Machine Learning Approach to Predict Lung Cancer using CT scan Images"
- [11] Jithu Varghese in "REVIEW ON LUNG CANCER DETECTION USING DEEP LEARNING"
- [12] Samuel H. Hawkins et ed. in "Predicting Outcomes of Non-small Cell Lung Cancer Using CT Image Features"
- [13] Ge Zhang1 in "Lung Nodule Classification in CT Images Using 3D Dense Net"
- [14] Melike GUNAY et. ed in "Comparison of Lung Cancer Detection Algorithms"
- [15] Kannan Rajeswari ET ED. IN "Deep Learning Algorithm for Classification and Prediction of Lung Cancer using CT Scan Images" 2019 5th International Conference on Computing, Communication, Control and Automation (ICCUBEA)
- [16] Muhammad Imran Faisal et. ed in "An Evaluation of Machine Learning Classifiers and Ensembles for Early-Stage Prediction of Lung Cancer" 3rd International Conference on Emerging Trends in Engineering, Sciences and Technology – 2018.
- [17] S. Lakshmana Prabu, S. N. Mohanty, K. Shankar, N. Arunkumar, and G. Ramirez, "Optimal deep learning model for classification of lung cancer on CT images," Future Generation Computer Systems, vol. 92, pp. 374–382, 2019.
- [18] Convolutional Neural Networks"-Caroline Petitjean, and Laurent Heutte University of Rouen, LITIS Lab Saint Etienne du Rouvray, France Fabio Alexandre Spanhol, Luiz S. Oliveira Federal University of Technology.
- [19] Pooja R. Katre, Dr. Anuradha Thakare, "Detection of Lung Cancer Stages using Image Processing and Data Classification Techniques", 2nd International Conference for Convergence in Technology (I2CT) 2017.
- [20] S. Kalaivani, Primit Chatterjee, Shikhar Juyal, Rishi Gupta, "Lung Cancer Detection Using Digital Image Processing and Artificial Neural Network", International Conference on Electronics, Communication and Aerospace Technology ICECA 2017
- [21] Anita Chaudhary, Sonit Sukhraj Singh, "Lung Cancer Detection on CT Images using Image Processing", International conference computing sciences, IEEE, 2012.
- [22] 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.
- [23] 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.

- [24] Chaudhary and S. S. Singh, "Lung Cancer Detection on CT Images by Using Image Processing," in Computing Sciences (ICCS), International Conference on, pp. 142-146, 2012
- [25] N. Mesanovic, M. Grgic, H. Huseinagic, M. Males, E. Skejic, and M. Smajlovic, "Automatic CT Image Segmentation of the Lungs with Region Growing Algorithm," in 18th International Conference on Systems, Signals and Image Processing-IWSSIP, pp. 395-400, 2011.
- [26] C. Kavitha, "A review on computer aided detection and diagnosis of lung cancer nodules," International Journal of Computers & Technology, vol. 3, pp. 393-400, 2012.

