

# Machine Learning-Driven Lung Cancer Detection

Priyanshu Sahay<sup>1</sup>, Deepanshu Gupta<sup>2</sup>, Piyush Kumar<sup>3</sup>, Medhavi Pandey<sup>4</sup>

<sup>1,2,3</sup>Department of Computer Science and Engineering (Data Science), Galgotias College of Engineering and Technology

<sup>4</sup>Assistant Professor, Department of Computer Science and Engineering, Galgotias College of Engineering and Technology

## Abstract

Lung cancer remains one of the most prevalent and deadly forms of cancer worldwide, highlighting the critical need for early detection methods to improve patient outcomes. In this study, we present a novel approach to lung cancer detection utilizing machine learning techniques, specifically leveraging the Random Forest algorithm. Our research focuses on developing a robust and accurate model capable of detecting lung cancer from medical imaging data with high precision and sensitivity. We collected a comprehensive dataset consisting of radiographic images from patients with confirmed lung cancer diagnoses, as well as from healthy individuals for comparison. The Random Forest algorithm, known for its versatility and ability to handle complex datasets, was employed as the primary machine learning framework in our study. Through a process of feature extraction and selection, our model was trained to identify patterns and subtle abnormalities indicative of lung cancer within the imaging data. We conducted extensive experimentation and evaluation to assess the performance of our proposed approach. Our results demonstrate promising outcomes, with the Random Forest model achieving notable accuracy rates in distinguishing between cancerous and non-cancerous lung tissues. Additionally, the model exhibited favorable sensitivity and specificity metrics, indicating its potential as a reliable tool for early lung cancer detection. Furthermore, we conducted comparative analyses with other machine learning algorithms commonly utilized in medical image analysis tasks. The findings underscore the efficacy of the Random Forest algorithm in this particular domain, showcasing its superiority in terms of both performance and interpretability. In conclusion, our research contributes to the ongoing efforts aimed at improving lung cancer diagnosis through the integration of machine learning technologies. By harnessing the power of the Random Forest algorithm, we offer a promising solution for early detection, thereby facilitating timely interventions and ultimately enhancing patient outcomes in the fight against lung cancer.

**Keywords:** Lung cancer detection, machine learning, Random Forest algorithm, medical imaging data, feature extraction, model evaluation, sensitivity, specificity, comparative analysis, interpretability, early detection, patient outcomes.

## 1. Introduction

Lung cancer remains a formidable global health challenge, with its prevalence and lethality underscoring the urgent need for innovative diagnostic approaches. Traditional detection methods, reliant on manual interpretation of medical images, are often time-consuming and susceptible to human error, impeding

timely intervention crucial for improving patient outcomes. However, recent advancements in machine learning (ML) and digital image processing offer promising avenues for enhancing the accuracy and efficiency of lung cancer detection.

This research endeavors to harness the synergy between digital image processing techniques and machine learning algorithms, specifically focusing on the Random Forest (RF) algorithm, to automate the detection of lung tumors. Drawing insights from seminal studies in the field, this paper sets out to develop a robust system capable of analyzing lung images converted from Digital Imaging and Communications in Medicine (DICOM) format, thereby facilitating swift identification of abnormalities indicative of lung cancer. The objectives of this research are multifaceted. First and foremost, the aim is to implement a comprehensive system adept at reading DICOM images of lung scans and discerning abnormalities through a sequence of image processing steps, including grayscale conversion, noise reduction, and feature extraction. Subsequently, by harnessing the discriminative power of the RF algorithm, the system endeavors to differentiate between malignant and benign tumors, thus enabling early intervention and personalized treatment strategies. Moreover, the evaluation of the proposed system constitutes a critical facet of this research. By assessing its accuracy in detecting both malignant and benign tumors in lung images, this study seeks to ascertain the efficacy and reliability of the developed framework. Through rigorous experimentation and comparative analysis, insights into the performance of the RF algorithm in lung cancer detection can be gleaned, paving the way for its potential integration into clinical practice.

In tandem with the broader landscape of ML applications in healthcare, this research underscores the transformative potential of leveraging advanced technologies to confront the challenges posed by lung cancer. By amalgamating state-of-the-art image processing techniques with powerful ML algorithms, this study endeavors to not only mitigate the limitations of traditional detection methods but also catalyze advancements in early diagnosis and treatment efficacy. In essence, this research represents a concerted effort to propel the frontier of lung cancer detection through the innovative fusion of digital image processing and machine learning methodologies. By striving towards the realization of accurate, efficient, and accessible diagnostic tools, this endeavor ultimately aspires to make meaningful strides in combatting one of the most formidable adversaries in modern oncology.

## 2. Literature Review

In a recent publication by Dakhaz Mustafa Abdullah & Nawzat Sadiq Ahmed in the International Journal of Science and Business (2021) [1], a comprehensive review of contemporary techniques for detecting lung cancer using machine learning is presented. Lung cancer, known for its high mortality rates and diagnostic challenges, has spurred the development of various methodologies, predominantly relying on CT scan and, to a lesser extent, X-ray images. These techniques often involve coupling multiple classifier methods with diverse segmentation algorithms to enable image recognition for identifying lung cancer nodules. Notably, the study highlights that CT scan images yield more accurate results, thus being preferred for cancer detection.

Stephen Neal Joshua, Midhun Chakkravarthy, and Debnath [2] conducted a systematic study on lung cancer detection using machine learning techniques, as documented in their paper published in the *Revue d'Intelligence Artificielle* in 2020. They specifically focused on comparing the detection accuracy of different classifiers using CT scans, which they found to be significantly higher compared to traditional analogue radiography, with a detection velocity approximately 2.6 times faster. To address the challenges associated with workload and accuracy, they investigated the efficacy of Computer-Aided Detection

(CAD) methods. Their study involved testing a model using a dataset of 453 CT images, with 217 images allocated for training purposes. Through validation, they achieved an overall accuracy of 82.9%, demonstrating promising results for the utilization of machine learning in improving lung cancer detection. Raut et al. (2021) [3] present a method for lung cancer detection using machine learning. They enhance input CT images to improve quality, segmenting them into pixels or superpixels. Otsu's method automates image thresholding, separating foreground and background pixels. Sobel filtering detects edges by calculating gradients. The Grey-Level Co-Occurrence Matrix (GLCM) analyzes texture features by pixel relationships, enhancing diagnostic capabilities (Raut, Patil, & Shelke, 2021).

In their 2018 study, Juan Lyu and Sai Ho Ling [4] emphasized the importance of early lung cancer detection using CT imaging. They introduced a multi-level convolutional neural network (ML-CNN) to classify lung nodules by malignancy. The ML-CNN includes two convolutional layers with batch normalization (BN) and pooling, aiming to address internal covariate shift. Their approach consists of three levels, each sharing similar structures and feature map counts in the final stage but differing in convolutional kernels. This innovative method shows promise in improving lung nodule classification accuracy and advancing early cancer detection and treatment effectiveness.

Dodia, Annappa, and Mahesh (2022) [5] reviewed recent advancements in deep learning-based lung cancer detection. They highlighted the importance of early detection and discussed techniques for image analysis, including feature extraction and classification. The paper also addressed global lung cancer statistics, datasets, and challenges in the field, proposing future research directions.

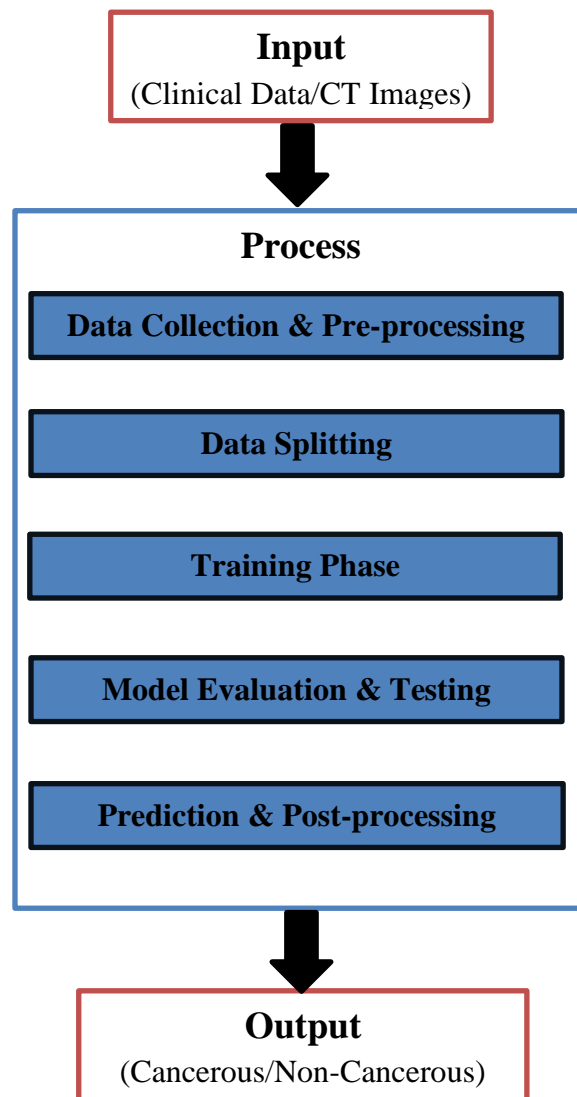
Manasee Kurkure and Anuradha Thakare (2016) [6] presented the method for early lung cancer detection using CT, PET, and X-ray images has gained attention. A genetic algorithm aids in early identification of lung cancer nodules. Naive Bayes and a genetic algorithm quickly classify cancer images with up to 80% accuracy. Contrast injection improves CT imaging quality, revealing various organs and issues. CT scans also detect kidney or gallstones, fluid buildup, and enlarged lymph nodes. They indirectly diagnose nearby soft tissue abnormalities, providing valuable insights.

### 3. Motivation and Objective

The motivation behind the project "Machine Learning-driven Lung Cancer Detection" stems from the urgent need to revolutionize current diagnostic methodologies for lung cancer, aiming to enhance early detection rates and consequently improve patient outcomes. Lung cancer remains a formidable global health challenge, with its prevalence and lethality underscoring the critical importance of timely intervention. Traditional detection methods often rely on manual interpretation of medical images, which can be time-consuming and prone to human error. By leveraging the power of machine learning, specifically the Random Forest algorithm, and analyzing clinical data of patients, this project seeks to develop a more efficient and accurate approach to lung cancer detection.

The primary objective of this project is to develop a robust machine learning-driven system capable of automatically detecting lung cancer from medical imaging data with high accuracy. By harnessing the discriminative power of the Random Forest algorithm, which excels in handling complex datasets and minimizing overfitting, the aim is to create a reliable framework for identifying suspicious patterns indicative of lung cancer. Furthermore, the project aims to validate the efficacy of this system by analyzing clinical data from a diverse patient population, ensuring its generalizability and real-world applicability. Ultimately, the overarching goal is to contribute to the advancement of early lung cancer detection methods, thereby facilitating prompt intervention and personalized treatment strategies, and ultimately

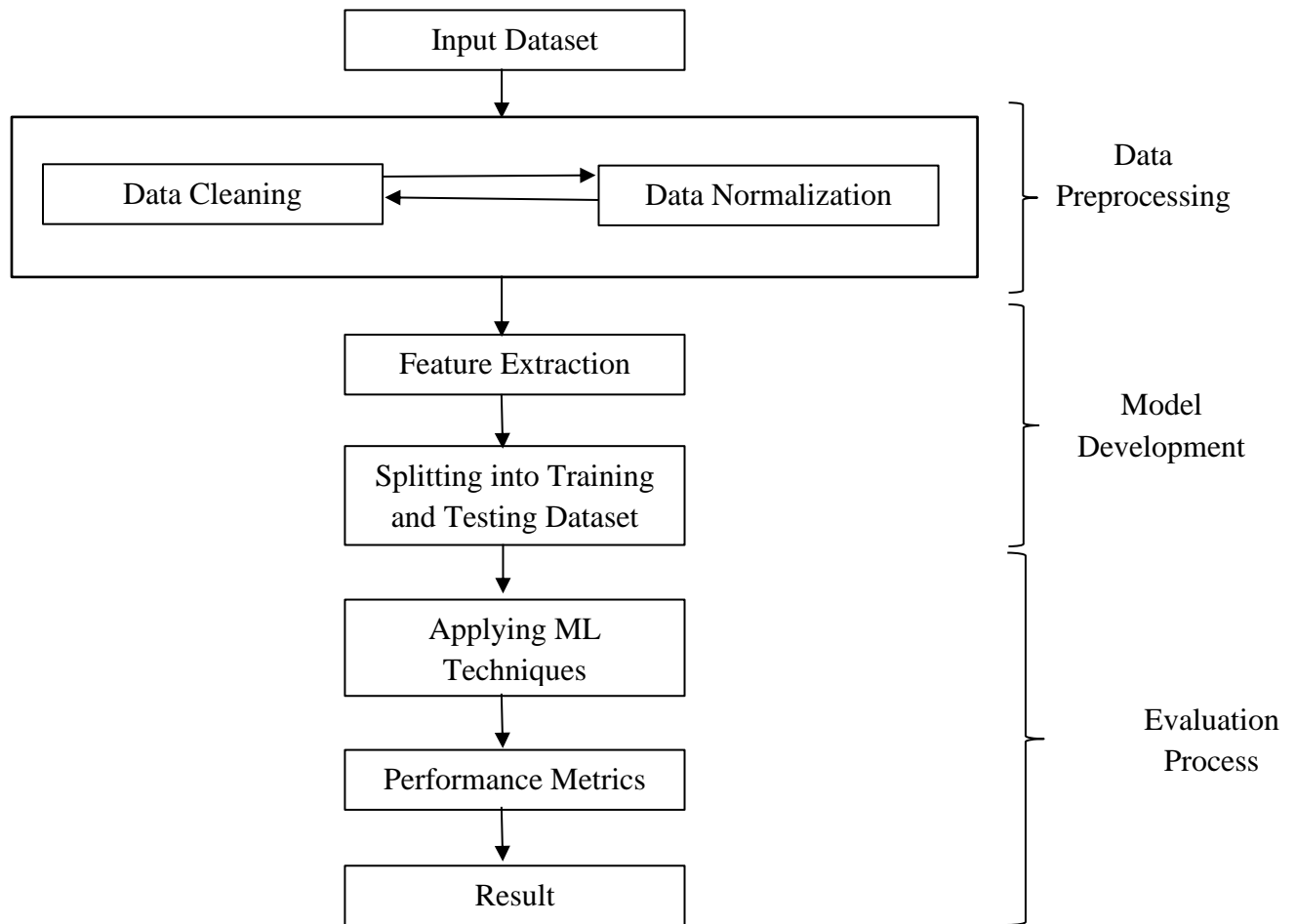
improving patient outcomes in the fight against this devastating disease. The proposed work introduces an innovative project aimed at revolutionizing the early detection of lung cancer through the integration of machine learning methodologies, specifically employing the Random Forest algorithm. Lung cancer is a formidable global health challenge, and its timely identification remains pivotal in improving patient prognosis and treatment outcomes. This project aims to leverage advanced computational techniques alongside medical imaging data to create a dependable predictive model for the early detection of lung cancer.



**Figure 1: Block Diagram of Target Methods**

#### 4. Methodology

To address the challenge, our research aim to develop a machine-learning-driven system for automated lung cancer detection. The focus is on enhancing the accuracy, speed, and reliability of lung cancer diagnosis through the integration of advanced machine learning techniques.



**Figure 2: Flowchart of Methodology**

The proposed methodology for the project is structured into three integral phases, each contributing to the development, evaluation, and refinement of the machine-learning-driven system for automated lung cancer detection.

#### 4.1 Phase 1: Data Collection and Pre-process

##### Data Collection:

In this initial phase, a diverse and comprehensive dataset of high-resolution CT scan images and several clinical data will be collected. The dataset will include instances of both normal lung scans and those exhibiting lung cancer nodules. Collaboration with medical institutions and leveraging publicly available datasets will ensure the acquisition of a representative sample.

##### Data Pre-processing:

To enhance the quality and consistency of the collected data, a series of pre-processing steps will be employed. This includes denoising techniques, normalization, and geometric standardization. Additionally, the data will undergo image segmentation to identify regions of interest (ROIs) relevant to potential lung cancer nodules.

#### 4.2 Phase 2: Model Development and Evaluation

##### Feature Extraction:

In this phase, the pre-processed CT scan images will undergo feature extraction to capture relevant

characteristics. Given the complexity of medical imaging, feature extraction involves identifying distinctive patterns and relevant information within the images and patient data. These extracted features serve as input variables for the Random Forest model, enabling it to discern intricate relationships and patterns associated with lung cancer.

### Ensemble Learning with Random Forests:

Simultaneously, an ensemble learning approach using Random Forests will be employed to complement the deep learning capabilities of CNNs. Random Forests, known for their adaptability and robustness, will contribute to the overall model by providing a structured decision-making mechanism and aiding in overcoming overfitting.

### Training and Evaluation:

The model will undergo training on a carefully selected subset of the dataset, followed by a meticulous evaluation of its performance. This evaluation will be comprehensive, incorporating key metrics including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC-ROC). Cross-validation techniques will be utilized to ensure the reliability and generalizability of the model.

## 4.3 Phase 3: Model Refinement and Feedback

### Refinement based on Evaluation:

Adjustments to hyperparameters and model architecture may be made to address any identified limitations or enhance overall performance. This step is crucial for iteratively improving the model's ability to accurately detect lung cancer across diverse cases.

### Incorporating Clinical Feedback:

Clinical feedback from healthcare professionals will be sought to ensure the model's clinical relevance and usability. Interpretability features, such as saliency maps, will be incorporated to facilitate transparency in the decision-making process.

### Model Deployment and Feedback Loop:

The refined model will be deployed in a simulated clinical environment, where it will process new data. Feedback from real-world scenarios will be continuously gathered, and the model will undergo further refinement as needed. This iterative feedback loop ensures the adaptability and responsiveness of the model to diverse clinical conditions.

## 5. Result Analysis

### 5.1 Source of Dataset

Developing a machine learning model for lung cancer prediction requires two types of datasets: clinical data and medical imaging data. Both play crucial roles in providing the model with the necessary information for learning and making accurate predictions.

| ID | GENDER | AGE | SMOKING | YELLOW_FINGERS | ANXIETY | PEER_PRESSURE | CHRONIC | FATIGUE | ALLERGY | WHEEZING | ALCOHOL | COUGHING | SHORTNESS | SWALLOWING | CHEST PAIN | LUNG_CANCER |
|----|--------|-----|---------|----------------|---------|---------------|---------|---------|---------|----------|---------|----------|-----------|------------|------------|-------------|
| 1  | M      | 69  | 1       | 2              | 2       | 1             | 1       | 2       | 1       | 2        | 2       | 2        | 2         | 2          | 2          | 2 YES       |
| 2  | M      | 74  | 2       | 1              | 1       | 1             | 2       | 2       | 2       | 1        | 1       | 1        | 2         | 2          | 2          | 2 YES       |
| 3  | M      | 59  | 1       | 1              | 1       | 2             | 1       | 2       | 1       | 2        | 1       | 2        | 2         | 1          | 2          | 2 NO        |
| 4  | F      | 63  | 2       | 2              | 2       | 1             | 1       | 1       | 1       | 1        | 2       | 1        | 1         | 2          | 2          | 2 NO        |
| 5  | M      | 63  | 1       | 2              | 1       | 1             | 1       | 1       | 1       | 2        | 1       | 2        | 2         | 1          | 1          | 1 NO        |
| 6  | F      | 75  | 1       | 2              | 1       | 1             | 2       | 2       | 2       | 2        | 1       | 2        | 2         | 2          | 1          | 1 YES       |
| 7  | M      | 52  | 2       | 1              | 1       | 1             | 1       | 2       | 1       | 2        | 2       | 2        | 2         | 1          | 2          | 2 YES       |
| 8  | F      | 51  | 2       | 2              | 2       | 2             | 1       | 2       | 2       | 1        | 1       | 1        | 2         | 2          | 2          | 1 YES       |
| 9  | F      | 68  | 2       | 1              | 2       | 1             | 1       | 2       | 1       | 1        | 1       | 1        | 1         | 1          | 1          | 1 NO        |
| 10 | M      | 53  | 2       | 2              | 2       | 2             | 2       | 1       | 2       | 1        | 2       | 1        | 1         | 2          | 2          | 2 YES       |
| 11 | M      | 61  | 2       | 2              | 2       | 2             | 2       | 2       | 1       | 2        | 1       | 2        | 2         | 2          | 2          | 1 YES       |
| 12 | M      | 72  | 1       | 1              | 1       | 1             | 2       | 2       | 2       | 2        | 2       | 2        | 2         | 2          | 1          | 2 YES       |
| 13 | F      | 60  | 2       | 1              | 1       | 1             | 1       | 2       | 1       | 1        | 1       | 1        | 2         | 1          | 1          | 1 NO        |
| 14 | M      | 58  | 2       | 1              | 1       | 1             | 1       | 2       | 2       | 2        | 2       | 2        | 2         | 1          | 2          | 2 YES       |
| 15 | M      | 69  | 2       | 1              | 1       | 1             | 1       | 1       | 2       | 2        | 2       | 2        | 1         | 1          | 2          | 2 NO        |
| 16 | F      | 48  | 1       | 2              | 2       | 2             | 2       | 2       | 2       | 2        | 1       | 2        | 2         | 2          | 2          | 2 YES       |

**Table 1: Survey for Lung Cancer**

Chest CT scans and X-rays provide visual representations of the patient's lungs. These images contain valuable information about the presence of tumors, nodules, or other abnormalities that may be indicative of lung cancer.

## 5.2 Size (No. of Samples) and description of attributes

### Clinical Data

The CSV file containing clinical information provides valuable insights into the patient's background and potential risk factors for lung cancer. The relevant features include:

- **Demographic information:** Gender, age
- **Lifestyle factors:** Smoking, alcohol consumption
- **Medical history:** Chronic diseases, allergies
- **Symptoms:** Wheezing, coughing, shortness of breath, chest pain
- **Psychological factors:** Anxiety, peer pressure

Each feature contributes to the model's understanding of the individual patient and their potential risk for developing lung cancer. Analysing relationships between these features can help identify potential patterns and correlations that the model can learn from.

## 5.3 Performance Measures

Evaluating the performance of a machine learning model is crucial for determining its effectiveness and identifying areas for improvement. In the case of lung cancer prediction, several performance measures can be used to assess the model's generalizability and accuracy.

### Accuracy:

Accuracy is the most basic measure, representing the proportion of correctly predicted cases (True Positives and True Negatives) over all cases.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

### Precision:

Precision measures the proportion of positive predictions that are actually correct (True Positives).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

### Recall:

Recall, in the context of evaluating a model's performance, quantifies the ratio of correctly identified positive cases (True Positives) out of all actual positive cases.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

### F1 Score:

The F1 score combines precision and recall into a single metric, providing a balanced view of the model's performance.

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

### Confusion Matrix:

The confusion matrix provides a detailed breakdown of the model's predictions, allowing for a more nuanced understanding of its performance. It displays the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

| Predicted | True Positive | False Positive |
|-----------|---------------|----------------|
| Positive  | TP            | FP             |
| Negative  | FN            | TN             |

**Table 2: Confusion Matrix**

**Error Metrics:**

Error metrics quantify the difference between the predicted and actual values. Common error metrics for regression tasks include:

- **Root Mean Square Error (RMSE):** Measures the average squared difference between predicted and actual values.
- **Mean Absolute Error (MAE):** Measures the average absolute difference between predicted and actual values.
- **R-squared:** Represents the proportion of variance in the actual values explained by the model.

**Choosing Appropriate Measures:**

The choice of performance measures depends on the specific task and the model's purpose. In lung cancer prediction, where accurate identification of positive cases is critical, prioritizing metrics like recall and F1 score may be more relevant than accuracy alone. Additionally, error metrics can be useful for regression tasks like predicting tumor size or survival rate.

**Interpreting Performance:**

Analyzing the chosen performance measures in conjunction with the confusion matrix provides valuable insights into the model's strengths and weaknesses. This information can be used to identify potential biases, imbalances in the dataset, or areas where the model struggles to make accurate predictions.

**Additional Considerations:**

- **Area Under the Curve (AUC):** For binary classification tasks, AUC can be used to evaluate the model's ability to distinguish between positive and negative cases.
- **Calibration curves:** These curves visualize the relationship between the predicted probability of a case being positive and the actual probability.
- **Sensitivity analysis:** This technique helps identify which features have the most significant impact on the model's predictions.

By utilizing various performance measures and analyzing them thoroughly, researchers and developers can gain valuable insights into the effectiveness of their machine learning models for lung cancer prediction. This information is crucial for improving model performance, identifying potential biases, and ensuring reliable and accurate predictions for patient diagnosis and treatment decisions.

**5.4 Output**

```
In [33]: prediction=model.predict(X_test)
         prediction

Out[33]: array([[1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
                1, 1, 1, 1,
                1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
                1, 1, 0, 1,
                1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0,
                1, 1, 0, 1,
                1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1])
```

**Figure 3: Prediction Output**

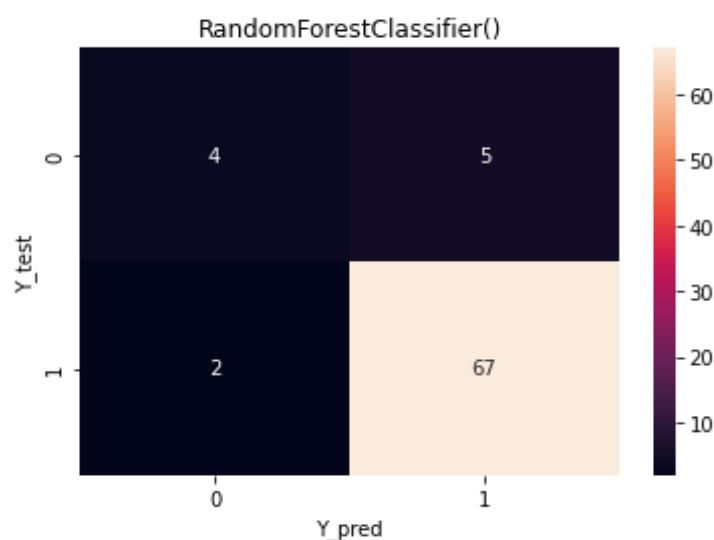


The provided result analysis output displays the predictions generated by the model for a set of test data ( $X_{test}$ ). The predictions are binary, denoted as 0 or 1, likely corresponding to the absence (0) or presence (1) of a particular condition, such as lung cancer in the context of the project. Understanding and interpreting these predictions is essential for assessing the model's performance and its practical implications. Upon generating predictions using the trained Random Forest classifier, the output array represents the model's classification decisions for each corresponding instance in the test set. In a binary classification task, the model assigns a label of 0 or 1 to each test instance based on its learned patterns from the training data. This array of predictions provides a snapshot of the model's output for the given set of test samples. To further analyze the results, it is crucial to compare the predicted labels with the actual ground truth labels in the test set. This comparison allows the computation of various performance metrics, such as accuracy, precision, recall, and F1 score, which provide insights into different aspects of the model's behavior.

For a qualitative assessment, visualizing the confusion matrix can be valuable. The confusion matrix breaks down the number of true positives, true negatives, false positives, and false negatives, offering a detailed overview of the model's strengths and weaknesses. It helps in understanding where the model excels and where it may need improvement, especially in terms of minimizing false positives or false negatives, depending on the application's requirements.

Moreover, examining the distribution of predictions can offer insights into potential imbalances in the dataset or the model's bias. For instance, if there is a substantial number of instances predicted as 1 (presence of the condition), it is crucial to evaluate the model's ability to correctly identify true positives while avoiding an excessive number of false positives.

In conclusion, the analysis of the prediction results involves a thorough examination of the model's output, comparison with the ground truth, and the computation of relevant performance metrics. This iterative process provides a comprehensive understanding of the model's efficacy, its limitations, and guides potential refinements for optimizing its performance in the specific context of lung cancer detection.



**Figure 4: Confusion Matrix Output**

The confusion matrix is a powerful tool in assessing the performance of a classification model, such as the Random Forest Classifier used in your lung cancer detection project. It provides a detailed breakdown

of the model's predictions, allowing for a thorough analysis of true positives, true negatives, false positives, and false negatives.

In the context of your Jupyter Notebook, after importing the confusion matrix, it appears that the accuracy of the Random Forest Classifier is 0.910, or 91.03%. This accuracy metric represents the ratio of correctly predicted instances to the total number of instances in the test set. While accuracy is an important overall measure, it might not provide a complete picture, especially in the presence of imbalanced datasets.

## 6. Discussions

In the discussion section, we can navigate on analyzing the prediction results from the Random Forest classifier is crucial for evaluating the model's performance in detecting lung cancer. By comparing the model's predictions with the actual labels in the test set, we can compute metrics like accuracy, precision, recall, and F1 score to gauge its effectiveness. Visualizing the confusion matrix provides insights into the model's strengths and weaknesses, helping identify areas for improvement, especially in reducing false positives or false negatives. Examining the distribution of predictions can reveal dataset imbalances or model biases. Overall, this analysis guides future refinements and contributes to advancing lung cancer detection methodologies.

Looking ahead, future research could explore ensemble techniques or deep learning models to enhance predictive accuracy. Integrating additional clinical features or multimodal imaging data may further improve diagnostic capabilities. Validating the model across diverse patient populations and clinical settings is essential for ensuring its real-world applicability. By pursuing these avenues, we can make significant progress in combating lung cancer through innovative machine learning approaches.

## 7. References

1. Abdullah, D. M., & Ahmed, N. S. (2021). A review of most recent lung cancer detection techniques using machine learning. *International Journal of Science and Business*, 5(3), 159-173.
2. Joshua, E. S. N., Chakkravarthy, M., & Bhattacharyya, D. (2020). An Extensive Review on Lung Cancer Detection Using Machine Learning Techniques: A Systematic Study. *Revue d'Intelligence Artificielle*, 34(3).
3. Raut, S., Patil, S., & Shelke, G. (2021). Lung cancer detection using machine learning approach. *International Journal of Advance Scientific Research and Engineering Trends (IJASRET)*.
4. Lyu, J., & Ling, S. H. (2018). Using multi-level convolutional neural network for classification of lung nodules on CT images. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 686-689). IEEE.
5. Dodia, S., Annappa, B., & Mahesh, P. A. (2022). Recent advancements in deep learning based lung cancer detection: A systematic review. *Engineering Applications of Artificial Intelligence*, 116, 105490.
6. Nageswaran, Sharmila, G. Arunkumar, Anil Kumar Bisht, Shivrul Mewada, J. N. V. R. Kumar, Malik Jawarneh, and Evans Asenso (2022). "Lung cancer classification and prediction using machine learning and image processing." *BioMed Research International*.
7. Pradhan, K., & Chawla, P. (2020). Medical Internet of things using machine learning algorithms for lung cancer detection. *Journal of Management Analytics*, 7(4), 591-623.
8. Abdullah, D. M., & Ahmed, N. S. (2021). A review of most recent lung cancer detection techniques using machine learning. *International Journal of Science and Business*, 5(3), 159-173.

9. Bhatia, S., Sinha, Y., & Goel, L. (2019). Lung cancer detection: a deep learning approach. In *Soft Computing for Problem Solving: SocProS 2017, Volume 2* (pp. 699-705). Springer Singapore.
10. Tekade, R., & Rajeswari, K. (2018). Lung cancer detection and classification using deep learning. In *2018 fourth international conference on computing communication control and automation (IC-CUBEA)* (pp. 1-5). IEEE.
11. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ, Jr, Wu YL, et al. (2017) Lung cancer: current therapies and new targeted treatments. *Lancet*;389:299–311.
12. Faisal, M. I., Bashir, S., Khan, Z. S., & Khan, F. H. (2018). An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer. In *2018 3rd international conference on emerging trends in engineering, sciences and technology (ICEEST)* (pp. 1-4). IEEE.
13. Shakeel, P. M., Burhanuddin, M. A., & Desa, M. I. (2019). Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. *Measurement*, 145, 702-712.
14. Lakshmi, S. V., Greeshma, B., Thanooj, M. J., Reddy, K. R., & Rakesh, K. R. (2021). Lung cancer detection and stage classification using supervised algorithms. *Turkish Journal of Physiotherapy and Rehabilitation*, 32, 3.



Licensed under [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)  
[Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)