International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Lung Cancer Detection Using Transfer and Deep Learning Techniques

T.Ramani Narasimha,K.Yasudha,U.Santosh Kumar

ABSTRACT

Lung cancer is also known as lung carcinoma, is the uncontrollable growth of abnormal cells in the lungs. Lung cancer is the leading cause of cancer deaths worldwide. Detecting the early stage of lung cancer helps to reduce mortality rates. This paper uses histopathological images for detecting lung cancer. Histopathological images help to identify cancer types very clearly because they are tissues sample images taken from patients through biopsy tests. These images allow us to detect cancer subtypes such as adenocarcinoma and squamous cell carcinoma, making identification easier. Using medical image analysis techniques, such as transfer learning based models Xception, ResNet50, MobileNetv2, and DenseNet121.The DenseNet121 model achieved an accuracy of 98%.

INTRODUCTION

Lung cancer is a type of cancer that starts when abnormal cells grow in an uncontrolled way in the lungs. The most common types of lung cancer are non-small cell carcinoma (NSCLC) and small cell carcinoma (SCLC).NSCLC is more common and gros slowly ,while SCLC is less common but often grows quickly The most common symptoms include cough that does not go away, chest pain, shortness of breath, fatigue.

Lung adenocarcinoma represents the most prevalent form of lung cancer and, similar to other lung cancer types, exhibits unique cellular and molecular characteristics. It falls under the category of non-small cell lung cancers (NSCLC), differentiating it from small cell lung cancer, which has distinct behavior and prognosis. Additionally, lung adenocarcinoma is subdivided into various subtypes and variants. The manifestations of this particular lung cancer type resemble those of other lung cancers, with patients frequently reporting a persistent cough and difficulty breathing.

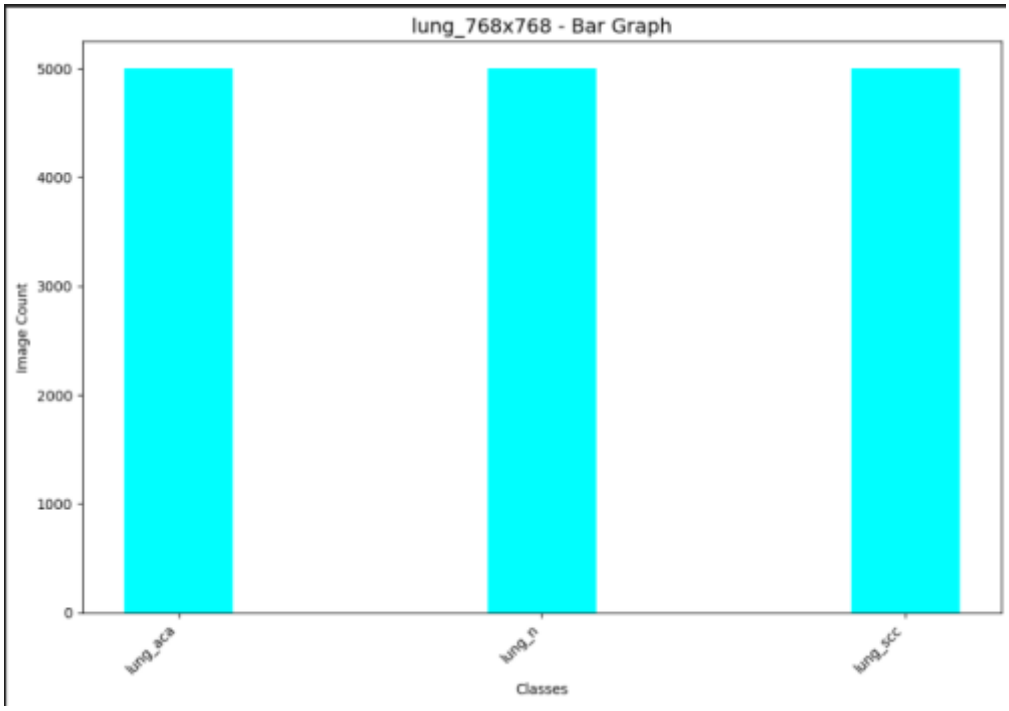
Small-cell lung carcinoma (SCLC) is traditionally categorized into two clinicopathological stages limited stage (LS) and extensive stage (ES). The determination of the stage primarily relies on the presence of metastases, whether the tumor is confined to the thoracic region, and if the total tumor burden within the chest can be treated with a single radiotherapy portal. Typically, when the tumor is restricted to one lung and the adjacent lymph nodes, it is classified as LS. Conversely, if the cancer has disseminated beyond this area, it is classified as ES.

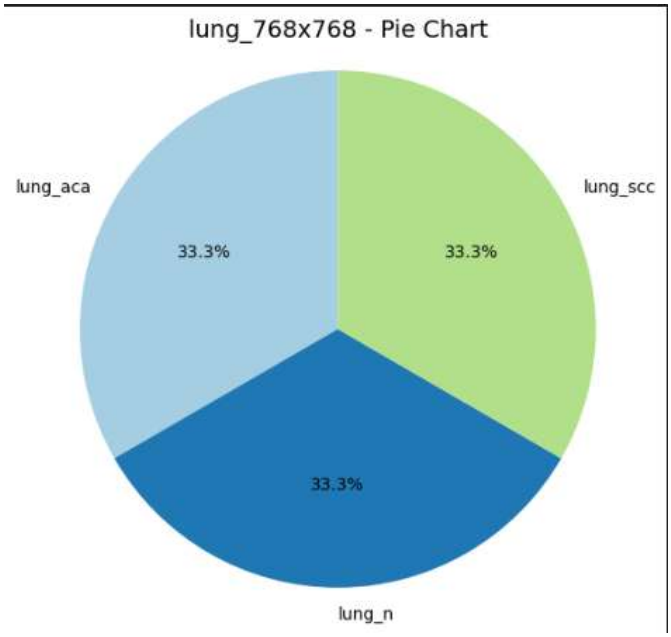
DATASET

This paper uses the dataset, lungcancer105k(LC105k),containing histopathological images.

There are 3 classes: adenocarcinoma(lung_aca), squamous cell carcinoma(lung_scc) and benign(lung_n).

A total of 15000 images are evenly distributed with each images being 768x768 pixels in jpeg format.

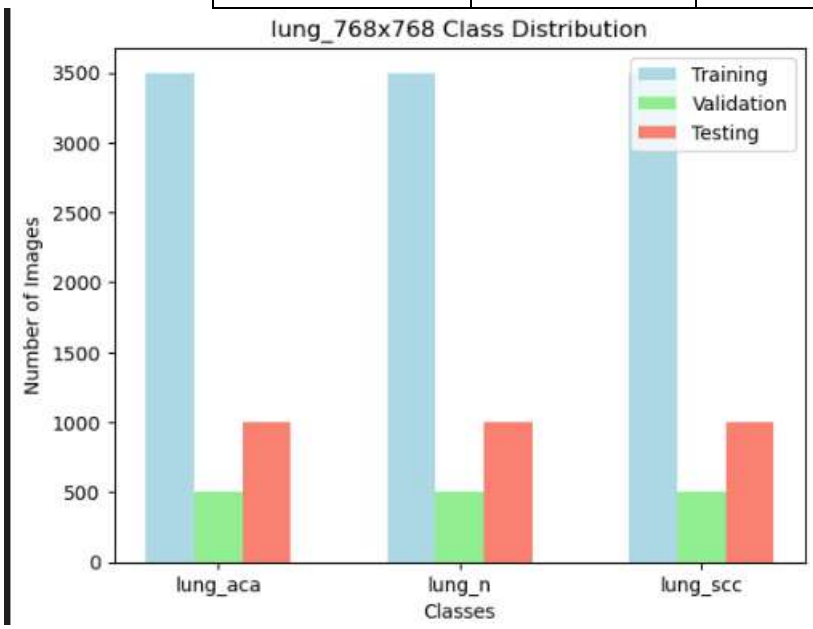




Data splitting

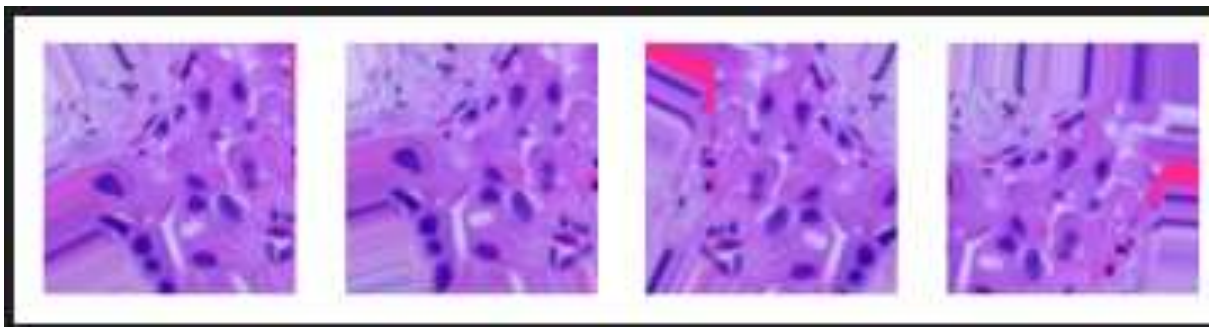
The dataset comprises a total of 1,500 histopathological images. Upon dividing the dataset, each class is allocated 3,500 images for training, 1,000 images for testing, and 500 images for validation.

S.No	Name	Parameter	Value
1	Random rotation	rotation_range	30
2	Horizontal shift	width_shift_range	0.2
3	Vertical shift	height_shift_range	0.2
4	Horizontal flip	horizontal_flip	True



DATA AUGUMENTATION

Data augumentation is a technique to increase dataset size It applies transformations to existing images, such as rotation, flipping and scaling. This technique reduces overfitting and improves model robustness , helping models learn invariant features. Data augumentation is essential for deep learning models, improving model performance and accuracy .



METHODOLOGY

PRE-TRAINED MODELS

Transfer learning is a deep learning technique that utilizes pre-trained convolutional neural network(cnn) models. Instead of training a model from scratch, this method uses pre-trained models have already been trained on large datasets. When applied to a new task, these models can be fine-tuned for improved performance. In lung cancer detection, transfer learning helps in identifying different types of cancer more accurately. Since the model has already learned important features from a large dataset, it can quickly recognize patterns in new medical images. This helps to reduce the need for large amounts of new training data and speeds up the detection process. The benefits of transfer learning include reduced training time and improved computational efficiency.

This study utilizes four pre-trained models MobileNetV2,Xception,ResNet50 and DenseNet121. These models trained on the ImageNet dataset.

MODEL ARCHITECTURES

1. **Mobilenetv2:** a lightweight, efficient architecture utilizing depthwise separable convolutions, which reduces computational cost while maintaining accuracy. It consists of 53 layers, including 17 depth wise separable convolutional layers.
2. **Xception:** A deep neural network architecture employing depthwise separable convolutions for high performance. It consists of 71 layers, including 36 depth wise separable convolutional layers, and uses a unique "extreme" inception module.
3. **Resnet50:** A residual learning-based architecture, composed of five stages of convolutional and identity blocks. It uses skip connections to facilitate training and improve accuracy, with a total of 50 layers.
4. **Densenet121:** A densely connected convolutional network architecture utilizing dense connections for improved feature extraction. It consists of 121 layers, including 4 dense blocks, which promote feature reuse and reduce the number of parameters.

EVALUATION METRICS

To evaluate the performance of a classification model, experts utilize various metrics specifically designed for classification tasks. These metrics assess the model's precision in forecasting the different categories present in the input data. Frequently employed performance measures in data classification include the F1-score, recall, accuracy, and precision.

1. **Confusion Matrix**

A confusion matrix is a detailed table that demonstrates a classification model's performance by comparing predicted labels to actual labels. Accuracy is the most commonly used metric, indicating the ratio of correct predictions made by the model.

Key Terms:

- **True Positive (TP):** The count of accurately identified positive cases.
- **True Negative (TN):** The count of accurately identified negative cases.
- **False Positive (FP):** Cases where the negative class is wrongly predicted as positive (Type I Error).

- **False Negative (FN):** Cases where the positive class is mistakenly predicted as negative (Type II Error).

Formulas:

- **Accuracy** = (TP + TN) / (TP + TN + FP + FN)
- **Precision** = TP / (TP + FP)
- **Recall (Sensitivity)** = TP / (TP + FN)
- **F1-Score** = 2 * (Precision * Recall) / (Precision + Recall)

2. **Classification Report**

This report encapsulates key classification metrics, typically generated using tools like **sklearn.metrics.classification_report.**

Metrics Included:

- **Precision:** Represents the accuracy of predicted positive results.
- **Recall (Sensitivity):** Indicates the percentage of actual positives that were predicted accurately.
- **F1-Score:** The harmonic mean of precision and recall, particularly useful for imbalanced datasets.
- **Support:** The total count of actual occurrences for each category within the dataset.
- **Accuracy:** The overall percentage of correct predictions.

RESULTS

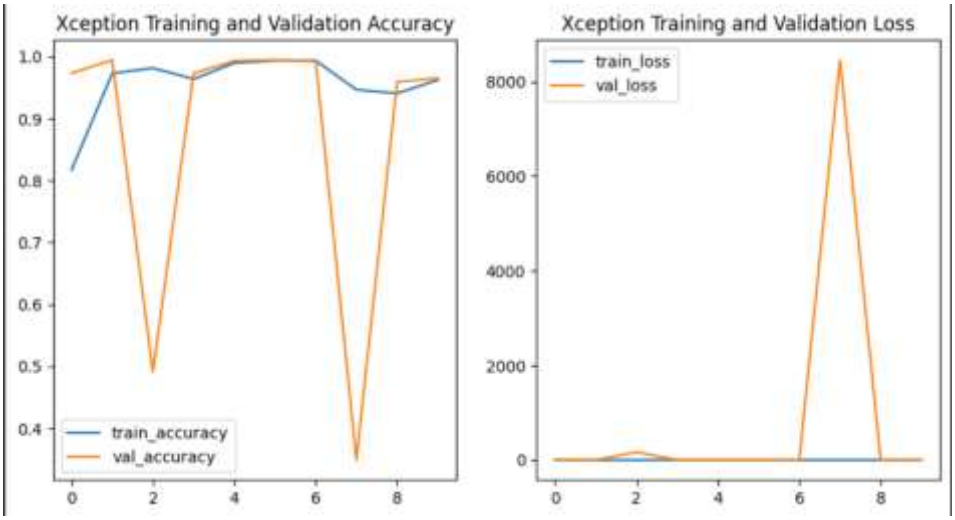
1)Xception

Accuracy – 96%

```
Confusion Matrix
[[ 883    5  112]
 [   0 1000    0]
 [   5    0  995]]
Classification Report
                      precision    recall  f1-score   support

        Adenocarcinoma       0.99      0.88      0.94      1000
                Benign       1.00      1.00      1.00      1000
Squamous Cell Carcinoma       0.90      0.99      0.94      1000

              accuracy                           0.96      3000
             macro avg       0.96      0.96      0.96      3000
          weighted avg       0.96      0.96      0.96      3000
```



2) ResNet50

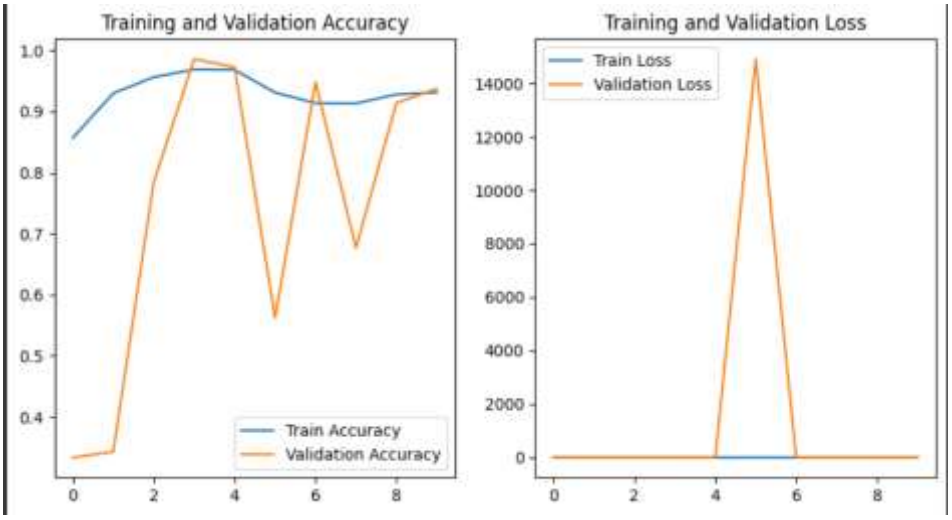
Accuracy – 92%

```
Confusion Matrix
[[799   1 200]
 [  8 992   0]
 [ 24   0 976]]
```

```
Classification Report
                    precision    recall  f1-score   support

     Adenocarcinoma       0.96      0.80      0.87      1000
             Benign       1.00      0.99      1.00      1000
Squamous Carcinoma        0.83      0.98      0.90      1000

           accuracy                           0.92      3000
          macro avg       0.93      0.92      0.92      3000
       weighted avg       0.93      0.92      0.92      3000
```



3) MobileNetV2

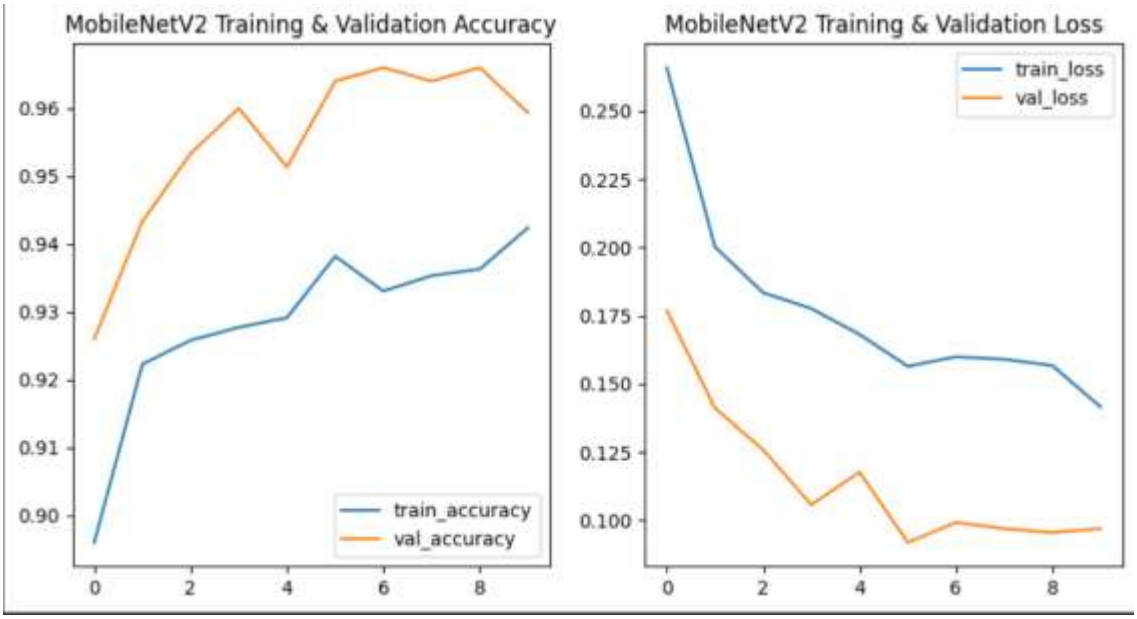
Accuracy – 94%

```
Confusion Matrix
[[954   6  40]
 [  2 998   0]
 [110   0 890]]
```

```
Classification Report
                  precision    recall  f1-score   support

    Adenocarcinoma       0.89      0.95      0.92      1000
            Benign       0.99      1.00      1.00      1000
Squamous Carcinoma       0.96      0.89      0.92      1000

          accuracy                           0.95      3000
         macro avg       0.95      0.95      0.95      3000
      weighted avg       0.95      0.95      0.95      3000
```



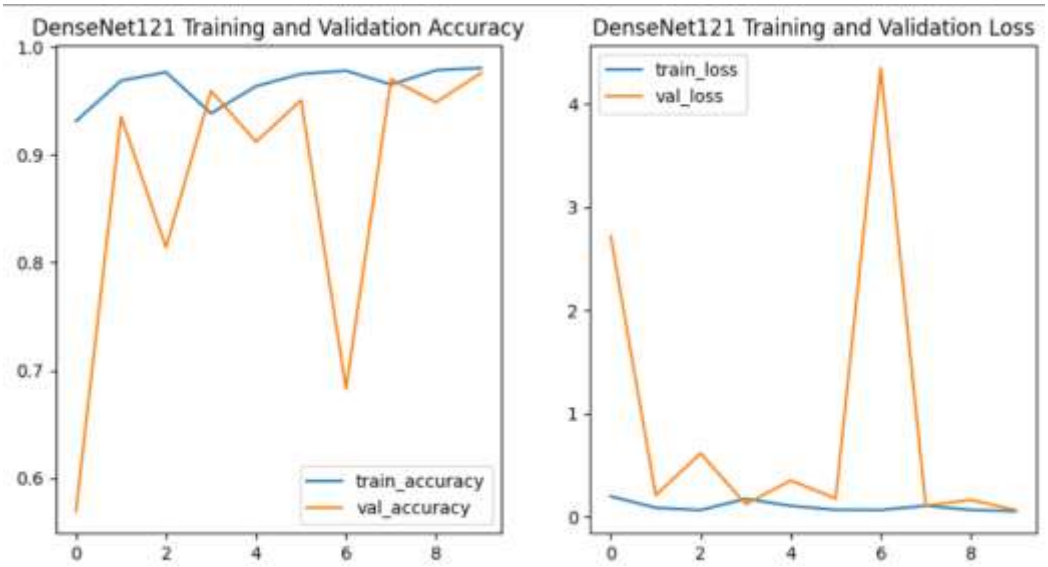
4)DenseNet121

Accuracy -98%

```
Confusion Matrix
[[ 974    6   20]
 [   0 1000    0]
 [  38    0  962]]
Classification Report
                      precision    recall  f1-score   support

        Adenocarcinoma       0.96      0.97      0.97      1000
                Benign       0.99      1.00      1.00      1000
Squamous Cell Carcinoma       0.98      0.96      0.97      1000

              accuracy                           0.98      3000
             macro avg       0.98      0.98      0.98      3000
          weighted avg       0.98      0.98      0.98      3000
```



CONCLUSION

The study evaluated the performance of four different pre-trained Convolutional Neural Network (CNN) models to detect lung cancer using histopathological samples. The four models examined were ResNet50, MobileNetV2, Xception, and DenseNet121. DenseNet121 achieved an accuracy of 98% in identifying lung cancer from histopathological images, the greatest of all models assessed.

DenseNet121's remarkable ability is attributable, in part, to its dense connectivity framework that enables better feature reuse and improved gradient flow allowing it to be very successful in evaluating images used in medicine. Comparatively, while other models such as MobileNetV2 and ResNet50 produced satisfactory outcomes, they were inferior to DenseNet121 in terms of accuracy and reliability, with DenseNet121 providing the best accuracy of 98% for lung cancer detection.

REFERENCES

1. Al-Shouka, T. T., & Alheeti, K. M. A. (2023, July). A transfer learning for intelligent prediction of lung cancer detection. In 2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT) (pp. 54-59). IEEE.
2. Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., ... & Tsirigos, A. (2018). Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature medicine, 24(10), 1559-1567.
3. Srinidhi, C. L., Ciga, O., & Martel, A. L. (2021). Deep neural network models for computational histopathology: A survey. Medical image analysis, 67, 101813.
4. Yektaei, H., & Manthouri, M. (2020). Diagnosis of lung cancer using multiscale convolutional neural network. Biomedical Engineering: Applications, Basis and Communications, 32(05), 2050030.
5. Al-Tarawneh, M. S. (2012). Lung cancer detection using image processing techniques. Leonardo electronic journal of practices and technologies, 11(21), 147-58.
6. Asuntha, A., & Srinivasan, A. (2020). Deep learning for lung Cancer detection and classification. Multimedia Tools and Applications, 79(11), 7731-7762.
7. Makaju, S., Prasad, P. W. C., Alsadoon, A., Singh, A. K., & Elchouemi, A. (2018). Lung cancer detection using CT scan images. Procedia Computer Science, 125, 107-114.
8. Midthun, D. E. (2016). Early detection of lung cancer. F1000Research, 5, F1000-Faculty.
9. Günaydin, Ö., Günay, M., & Şengel, Ö. (2019, April). Comparison of lung cancer detection algorithms. In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT) (pp. 1-4). IEEE.
10. Kubik, A., & Polak, J. (1986). Lung cancer detection results of a randomized prospective study in Czechoslovakia. Cancer, 57(12), 2427-2437.