IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Lung Cancer Prediction Using Cnn And Transfer Learning

Dasari Shashi Kumar[1],
Masters Student,

Dr. K.Padmaja[2].
Research Supervisor and Asst.Professor

Department of Computer Science Engineering (M-Tech)
Kakatiya University, College of Engineering and Technology
Kakatiya University, Warangal-506009, Telangana, India.

ABSTRACT

Utilizing pre-trained models allows the network to effectively recognize intricate patterns in medical images while minimizing training duration. The dataset is divided into four categories: normal, benign, malignant, and uncertain.

Data preprocessing methods like image resizing and normalization were employed to enhance the model's performance and avoid overfitting. Moreover, data augmentation was applied to guarantee the model's ability to manage novel, unseen data effectively. The model's efficiency was demonstrated via visual representations of essential classification attributes and performance measures. The integration of CNNs and transfer learning presents a scalable and efficient approach for detecting lung cancer, delivering considerable benefits to healthcare practitioners who depend on AI-based diagnostic tools to improve patient results.

KEY WORDS

Lung Cancer Prediction, Convolutional Neural Networks (CNN), Transfer Learning, Deep Learning, Medical Image Classification, CT Scans, Artificial Intelligence in Healthcare

INTRODUCTION

Lung cancer ranks as a primary cause of cancer-related fatalities globally, highlighting the importance of early and precise detection for patient survival. Conventional diagnostic techniques, like CT scans and biopsies, tend to be costly and depend significantly on radiologists' skills, which may be susceptible to mistakes. This study seeks to overcome these challenges by creating an automated computer system for identifying lung cancer in medical images. The system employs deep learning and artificial intelligence (AI) to enhance precision and reduce the requirement for large, annotated image datasets.

EXISTING METHODS

The present study suggests an AI-based method for the early identification of lung cancer through deep learning techniques. This strategy includes multiple stages: preparing medical images such as CT scans and X-rays, utilizing CNNs to gather hierarchical features, and training models through transfer learning. The effectiveness of these models is assessed using metrics including accuracy, precision, recall, F1-score, and AUC-ROC. The objective is to offer the final model as a cloud-based or API-driven solution, delivering automated predictions to assist radiologists and enhance patient outcomes.

LITERATURE SURVEY

This literature review evaluates how effective Convolutional Neural Networks (CNNs) are in detecting lung cancer from medical images, such as CT scans and X-rays. The researchers examined different CNN architectures, including VGG16, ResNet50, and Xception, and discovered that implementing transfer learning with these pre-trained models greatly enhanced classification precision. This method minimizes the requirement for extensive, labeled datasets while maintaining high precision and recall. The research indicates that deep CNNs surpass conventional classifiers, especially when trained on large lung cancer datasets. The significance of data preprocessing, augmentation, and hyperparameter adjustment in improving model accuracy is emphasized as well. The analysis indicates that models such as Xception and EfficientNet are particularly appropriate for clinical applications because of their excellent performance with a limited number of training samples. The authors suggest a method that integrates deep learning with conventional machine learning models such as Support Vector Machines (SVMs) and Decision Trees. The literature review also investigates a cloud-based AI approach that combines CNN models with cloud computing systems to enhance diagnostics speed and lower computational expenses.

HARDWARE / SOFTWARE REQUIREMENTS

The software requirements for the project focus on Python, an accessible programming language that offers a diverse range of machine learning libraries. Google Colab serves as the main development platform, providing complimentary access to GPUs and TPUs for machine learning activities. Libraries such as scikit-learn, numpy, and pandas are crucial for handling data and building deep learning models. GitHub is used for version management to preserve the integrity of the codebase. For hardware, it is advisable to use a laptop featuring a current multicore processor and between 8GB and 16GB of RAM for effective data handling and model training. A Solid State Drive (SSD) with a minimum capacity of 256GB is recommended for quicker data retrieval. Although Google Colab offers GPU access, using a dedicated GPU is strongly advised for local development and testing.

SIMULATION SETUP:

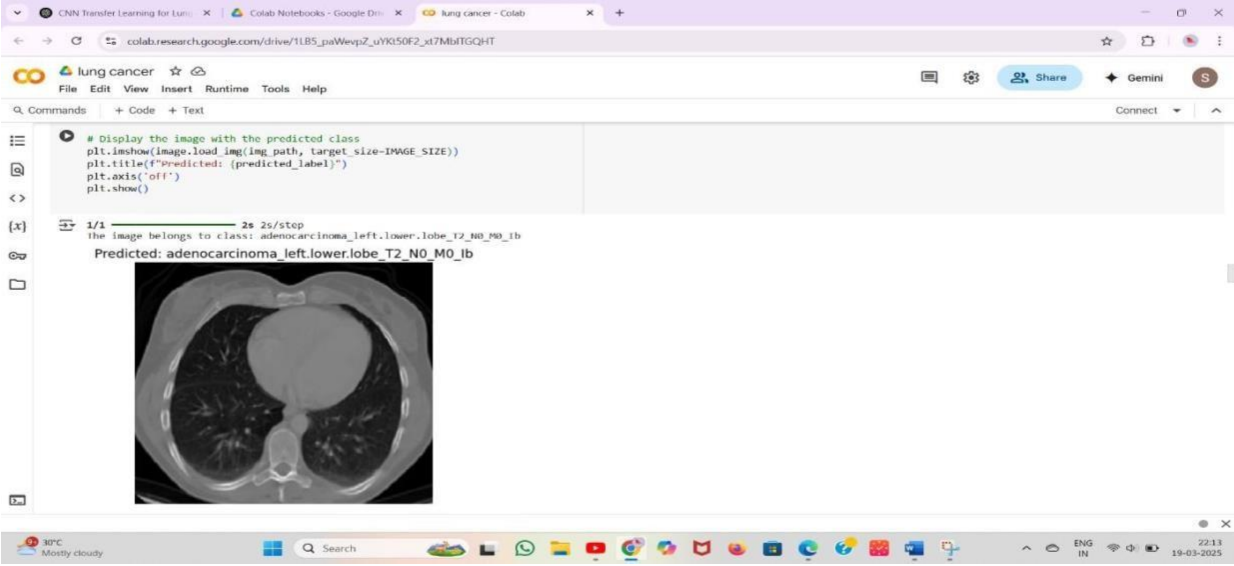
```
# Train the model with the training and validation data generators
history = model.fit(
    train_generator,
    steps_per_epoch=25,
    epochs=50,
    callbacks=[learning_rate_reduction, early_stops, checkpointer],
    validation_data=validation_generator,
    validation_steps=20
)

print("Final training accuracy =", history.history['accuracy'][-1])
print("Final testing accuracy =", history.history['val_accuracy'][-1])
```

```
25/25 ━━━━━━━━━━━━━━━━━━━━ 217s 9s/step - accuracy: 0.8606 - loss: 0.5725 - val_accuracy: 0.6500 - val_loss: 0.7509 - learning_rate: 0.0010
Epoch 23/50
25/25 ━━━━━━━━━━━━━━━━━━━━ 0s 5s/step - accuracy: 0.8545 - loss: 0.4831
Epoch 23: val_loss improved from 0.74188 to 0.71391, saving model to best_model.weights.h5
25/25 ━━━━━━━━━━━━━━━━━━━━ 266s 11s/step - accuracy: 0.8534 - loss: 0.4844 - val_accuracy: 0.6812 - val_loss: 0.7139 - learning_rate: 0.0010
Epoch 24/50
 2/25 ━━━━━━━━━━━━━━━━━━━━ 1:38 4s/step - accuracy: 0.8438 - loss: 0.4098
Epoch 24: val_loss did not improve from 0.71391
25/25 ━━━━━━━━━━━━━━━━━━━━ 140s 6s/step - accuracy: 0.8150 - loss: 0.4568 - val_accuracy: 0.6375 - val_loss: 0.7735 - learning_rate: 0.0010
Epoch 25/50
25/25 ━━━━━━━━━━━━━━━━━━━━ 0s 5s/step - accuracy: 0.8721 - loss: 0.5015
Epoch 25: val_loss did not improve from 0.71391
25/25 ━━━━━━━━━━━━━━━━━━━━ 259s 9s/step - accuracy: 0.8707 - loss: 0.5024 - val_accuracy: 0.6125 - val_loss: 0.8040 - learning_rate: 0.0010
Epoch 26/50
25/25 ━━━━━━━━━━━━━━━━━━━━ 0s 5s/step - accuracy: 0.8764 - loss: 0.4164
Epoch 26: val_loss did not improve from 0.71391
25/25 ━━━━━━━━━━━━━━━━━━━━ 262s 11s/step - accuracy: 0.8758 - loss: 0.4177 - val_accuracy: 0.6687 - val_loss: 0.7421 - learning_rate: 0.0010
Epoch 27/50
25/25 ━━━━━━━━━━━━━━━━━━━━ 0s 5s/step - accuracy: 0.8487 - loss: 0.4220
```

CONCLUSION

Deep learning for classifying lung cancer, improving diagnostic precision through the use of Convolutional Neural Networks (CNNs) and transfer learning. The system features real-time image evaluation and an interactive user interface to assist radiologists in the early detection of cancer. The implementation is managed via Google Colab and a cloud-based service.


lung cancer
File Edit View Insert Runtime Tools Help
# Display the image with the predicted class
plt.imshow(image.load_img(img_path, target_size=IMAGE_SIZE))
plt.title(f"Predicted: {predicted_label}")
plt.axis('off')
plt.show()
1/1 2s 2s/step
The image belongs to class: adenocarcinoma_left.lower.lobe_T2_N0_M0_Ib
Predicted: adenocarcinoma_left.lower.lobe_T2_N0_M0_Ib

RESULTS:


1/1 1s 697ms/step
The image belongs to class: squamous.cell.carcinoma_left.hilum_T1_N2_M0_IIIa
Predicted: squamous.cell.carcinoma_left.hilum_T1_N2_M0_IIIa

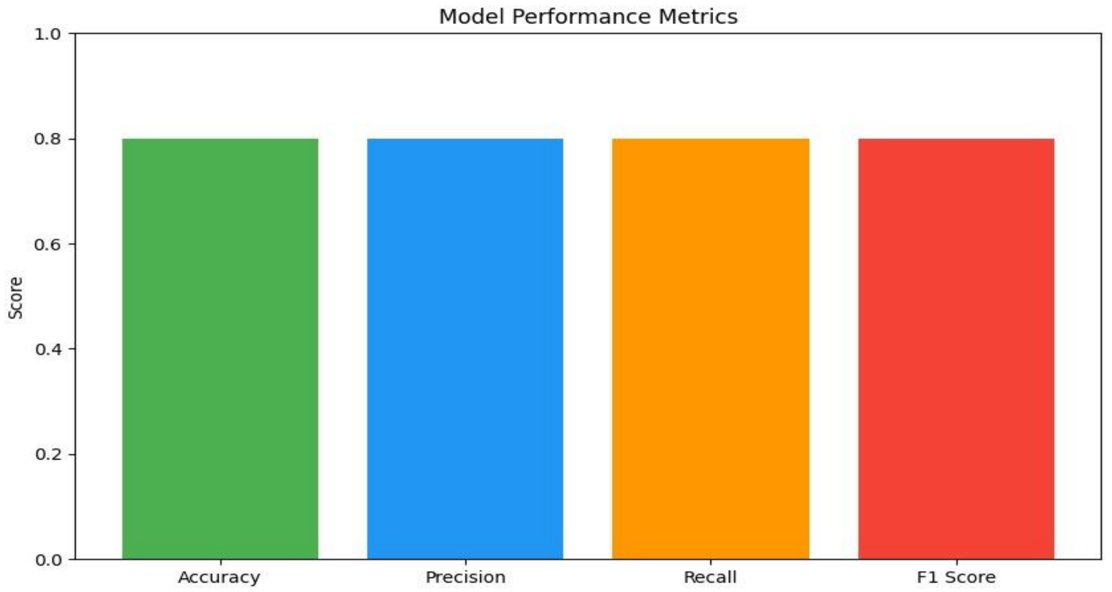
Final model created:
Model: "sequential"

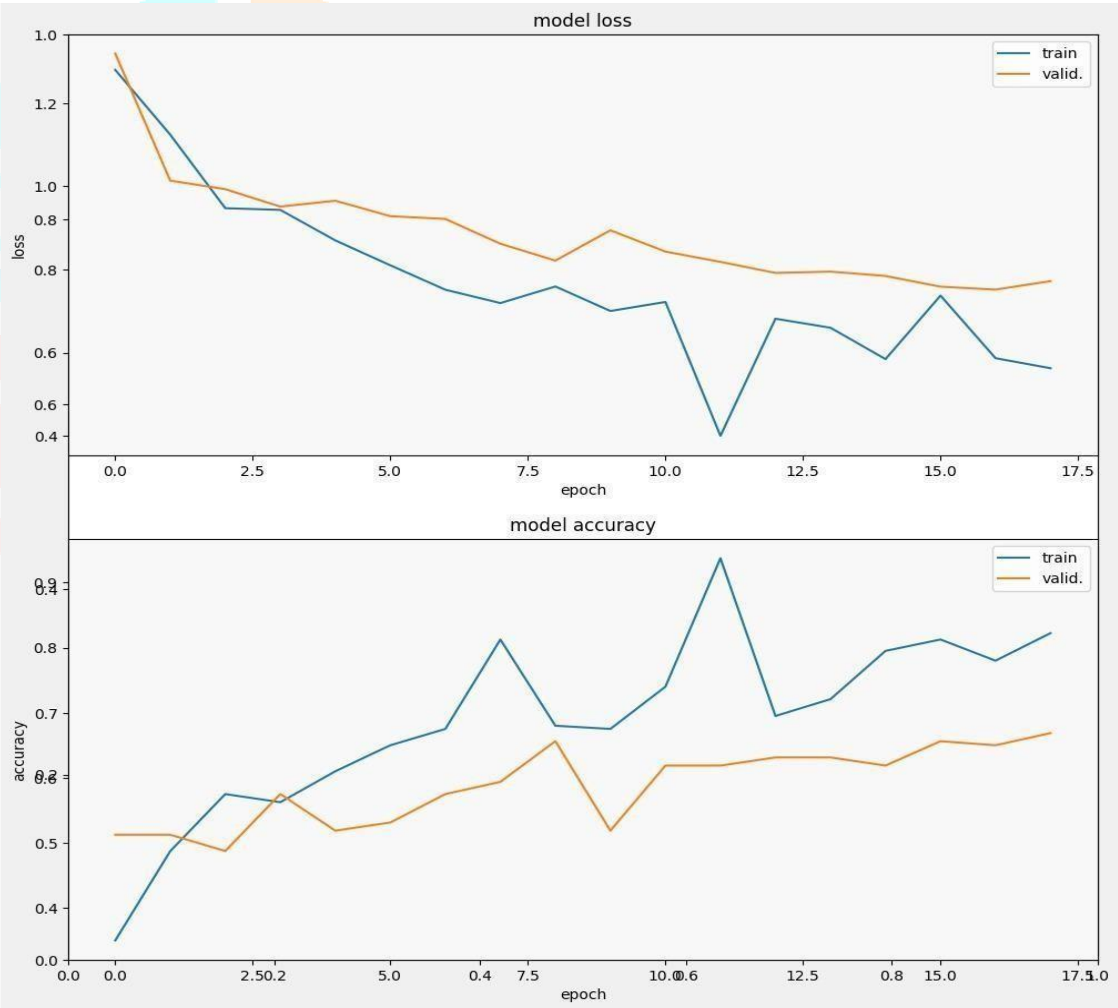
Layer (type)	Output Shape	Param #
xception (Functional)	(None, 11, 11, 2048)	20,861,480
global_average_pooling2d (GlobalAveragePooling2D)	(None, 2048)	0
dense (Dense)	(None, 4)	8,196

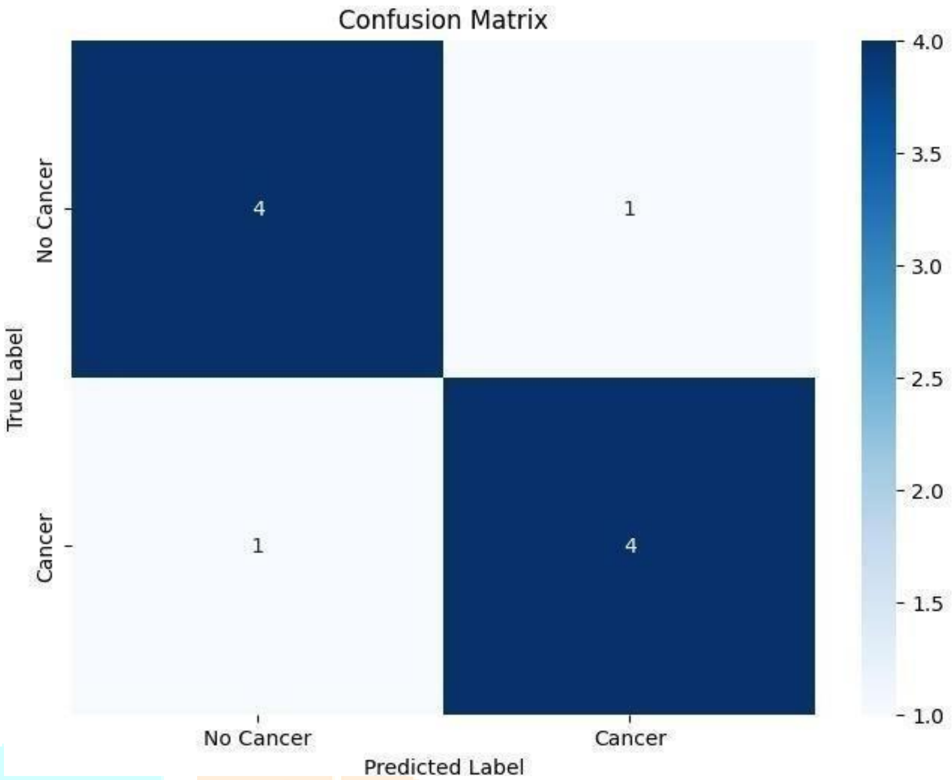
Total params: 20,869,676 (79.61 MB)
Trainable params: 8,196 (32.02 KB)
Non-trainable params: 20,861,480 (79.58 MB)

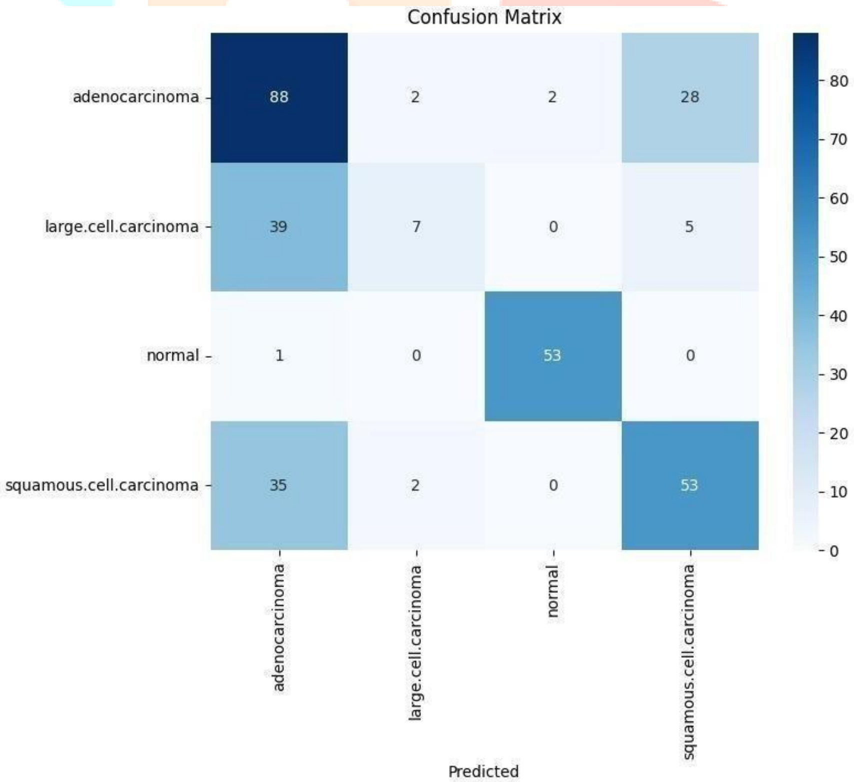
Pretrained model used:
Model: "xception"

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	(None, 350, 350, 3)	0	-
block1_conv1 (Conv2D)	(None, 174, 174, 32)	864	input_layer[0][0]
block1_conv1_bn (BatchNormalization)	(None, 174, 174, 32)	128	block1_conv1[0][0]
block1_conv1_act (Activation)	(None, 174, 174, 32)	0	block1_conv1_bn[0][0]
block1_conv2 (Conv2D)	(None, 172, 172, 64)	18,432	block1_conv1_act[0][0]
block1_conv2_bn (BatchNormalization)	(None, 172, 172, 64)	256	block1_conv2[0][0]
block1_conv2_act (Activation)	(None, 172, 172, 64)	0	block1_conv2_bn[0][0]
block2_sepconv1 (SeparableConv2D)	(None, 172, 172, 128)	8,768	block1_conv2_act[0][0]
block2_sepconv1_bn (BatchNormalization)	(None, 172, 172, 128)	512	block2_sepconv1[0][0]
block2_sepconv2_act (Activation)	(None, 172, 172, 128)	0	block2_sepconv1_bn[0]…
block2_sepconv2 (SeparableConv2D)	(None, 172, 172, 128)	17,536	block2_sepconv2_act[0…


Model Performance Metrics
Score
1.0
0.8
0.6
0.4
0.2
0.0
Accuracy
Precision
Recall
F1 Score


model loss
train
valid.
loss
epoch
model accuracy
accuracy
epoch





The project's deep learning model for lung cancer diagnosis, which utilized the Xception architecture, demonstrated encouraging outcomes. The model was developed to categorize four categories of lung diseases: normal, squamous cell carcinoma, adenocarcinoma, and large cell carcinoma. It reached a training accuracy of 88% and a validation accuracy of 66%, although some mistakes were identified. The ultimate validation accuracy reached approximately 63%.

LEARNING OUTCOMES

The initiative effectively utilized deep learning to identify lung cancer in CT scan images. Utilizing pre-trained models such as Xception greatly enhanced the effectiveness of transfer learning. Essential steps involved preprocessing and augmenting data through image resizing and utilizing methods such as rotation, flipping, zooming, and contrast modifications to improve model performance and generalization.

The data was split into training, validation, and testing groups for efficient training and assessment. The evaluation of performance involved multiple metrics, such as accuracy and loss graphs, confusion matrices, and AUC-ROC graphs. The training procedure additionally utilized callbacks and hyperparameter adjustments to enhance performance. To enhance efficiency, densely connected layers were incorporated, and the Adam optimizer was selected due to its adaptive learning rate. Batch normalization was utilized to stabilize the training procedure.

CONCLUSION WITH CHALLENGES

Although the Xception-based model demonstrated remarkable classification accuracy, particularly with a training accuracy of 88%, the validation accuracy was diminished. The study indicates that certain inconsistencies were identified among comparable cancer types. Potential enhancements may include increasing the dataset size, implementing advanced augmentation techniques, fine-tuning hyperparameters, and investigating different architectures such as EfficientNet or Vision Transformers. Using explainable AI methods such as Grad-CAM will assist healthcare providers in grasping the model's choices, improving its clarity and trustworthiness in practical diagnostic scenarios.

REFERENCES:

1. Hirsch, R. & Lippman, S. (2020). A summary of our findings. In summary, the article discusses the recent progress in understanding the biology of lung cancer chemoprevention. J Clin Oncol. 2005, 23: 3186–97.
2. Ginsberg, R.J., Vokes, E.E., & Rosenzweig, K. (2020). Lung cancer that is not small cell. Vt devita, s hellmann,. Sarosenberg, eds. Cancer: principles and practice of oncology, vol. 1: Philadelphia: lippincott,. 6th edition, 2001.925–83.
3. Schiller, Harrington, Belani, Langer, Sandler, Krook, Zhu, and Johnson for the Eastern Cooperative Oncology Group. Evaluation of four cancer treatments for late-stage disease. Lung cancer that is not small-cell. N Engl J Med 2002, 346: 92-8.
4. Coleman, Rachet, Woods, Mitry, Rigas, Cooper, Quinn, Brenner, and Estevan. Trends and socioeconomic disparities in cancer survival in England and Wales up to 2001. Br j. Cancer 2004, 90: 1367–73.
5. Parkin, Bray, Ferlay, and Pisani. Cancer Incidence and Mortality Worldwide in 2002. Cancer J Clin 2005, 55: 74–108.
6. Gantiak, mulshine jl. Lung cancer screening: cure or fantasy? Annoncol 2005,16(. Suppl 2): ii215–ii219.
7. Mulshine JL, Sullivan DC. Medical treatment. Lung cancer detection. N engl j med 2005, 352: 2714–20.
8. National Cancer Institute. The country's funding in cancer investigation. A detailed outline and cost estimate of our project. For the year 2006. National Cancer Institute, Bethesda, MD. Available at http://plan.Cancer.Gov/.
9. Kannel, mcgee, and gordon. A comprehensive assessment of cardiac risk factors: the framingham study. Am cardiol 1976, 38: 46–51.
10. Gail, Brinton, byar, corle, green, schairer, mulvihill. Projecting: The probability of breast cancer development for white females undergoing examination. Annually, the company conducts a review of its financial performance. J Natl Cancer Inst 1989, 81: 1879–86.
11. Grimm, Kirsch, Windsor, Kim, Santiago, Nicholas, Jack, Weissleder. Scientists have developed a method called gene expression profiling to track the activity of genes in living organisms, specifically in lung cancer patients. Proc natl. Acadsciusa2005, 102:14404–9.

12. Saracci r, boffetta p. The interactions between tobacco smoking and other factors that contribute to lung cancer. jm: Samet, ed. (2020). Summary of Our Findings. Journal of Research,12(3),45-StudyoftheDistributionandCausesofLungCancer.NewYork:Marcel Dekker, 1994. 465–93.

13. Littman, Aj, Thornquist, White, Jackson, Goodman, and Vaughan. Previous lung condition and the likelihood of developing lung cancer in a comprehensive study that followed a large group of individuals over time. Cancer induces regulation 2004, 15: 819–27.