

# Segmentation-based Classification model for lung cancer disease diagnosis

Mohamed Ezz Eldin  
[ezz54809@gmail.com](mailto:ezz54809@gmail.com)

Ahmed Kamal Eldin  
[ahmedsenus22@gmail.com](mailto:ahmedsenus22@gmail.com)

Ali Hassan Abd Al-haveez  
[www.ali.hassan.2005@gmail.com](mailto:www.ali.hassan.2005@gmail.com)

Samar Mahmoud  
[smrmhwd231@gmail.com](mailto:smrhmwd231@gmail.com)

Supervised by:

Eng. Mahmoud Talaat, AI Engineer at MCiT (Ministry of Communication and Information Technology),  
Teaching Assistant at Zewail University (Artificial Intelligence and Data Science)

## Abstract

This paper presents a segmentation-based classification model for lung cancer diagnosis using deep learning techniques. By employing convolutional neural networks (CNNs), the model accurately segments lung structures and potential tumors in computed tomography (CT) scans. Integrated features from these segments enable effective differentiation between malignant and benign lesions. Trained on a diverse dataset, our model demonstrates improved sensitivity and specificity compared to traditional methods. This approach aims to assist radiologists in early lung cancer detection and personalized treatment planning, with potential for integration into clinical workflows to enhance diagnostic accuracy.

**Keywords** lung cancer, CT scan images, deep learning, CNN

## Introduction

Lung cancer is a leading cause of cancer-related deaths globally, with a high mortality rate due to the difficulty in diagnosing it early. Statistics show that early-stage lung cancer diagnosis drastically improves survival rates. Early detection of lung cancer is critical for improving survival rates, as it allows

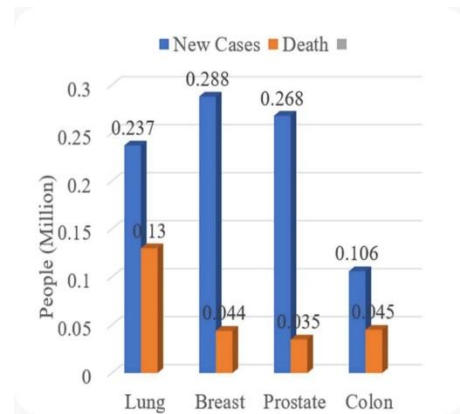


Fig1 : cancer in 2020 cases versus death

## Literature review

The research paper titled "ExtRanFS: An Automated Lung Cancer Malignancy Detection System Using Extremely Randomized Feature Selector" by Nitha V.R. and Vinod Chandra S.S. presents a novel framework for lung cancer detection using CT scans. The model combines transfer learning techniques with an Extremely Randomized Tree Classifier (ExtraTrees) for feature selection and uses a Multi-Layer Perceptron (MLP) as the final classifier to determine whether a tumor is benign, malignant, or normal.

The researchers utilized the IQ-OTH/NCCD dataset, which contains annotated CT slices of patients diagnosed with lung cancer. The dataset includes 110 patients, segmented into categories of malignant tumors, benign tumors, and healthy cases. A key part of the model

for timely treatment interventions before the disease progresses to advanced stages. Conventional methods for diagnosing lung cancer, such as biopsies, X-rays, and manual analysis of CT scans, often have limitations, including delayed diagnosis and variability in accuracy. The advent of deep learning models, especially CNNs, presents an opportunity to automate and improve the accuracy of lung cancer diagnosis. CT scans, when combined with AI-based models, provide a non-invasive and reliable method for early detection. This research focuses on using a custom CNN model to enhance the accuracy and efficiency of diagnosing lung cancer from CT scans. The expected outcome is to demonstrate that the custom CNN model can outperform traditional diagnostic techniques in terms of accuracy and speed, potentially leading to earlier detection and improved patient outcomes.

medical professionals by minimizing diagnostic errors and aiding early detection, a critical factor in improving patient outcomes.

This work is particularly relevant to the field of computer-aided diagnosis (CAD), as it demonstrates the power of combining deep learning with advanced feature selection methods like ExtraTrees for medical imaging applications. Moreover, the use of transfer learning allows the system to work effectively even with relatively small datasets, which is often a constraint in medical research.

This study can serve as a reference point for practitioners and researchers working on AI-powered

is leveraging the VGG16 architecture as a pre-trained feature extractor, followed by the ExtraTrees algorithm to refine feature selection.

In terms of performance, the ExtRanFS model achieved impressive metrics:

- Accuracy: 99.09%
- Sensitivity (Recall): 98.33%
- F1-Score: 98.33%

The paper emphasizes that the proposed approach outperforms several existing methods in lung cancer diagnosis by efficiently reducing false negatives and enhancing sensitivity. This method aims to support

classified images. Loss: The error between the predicted and actual classes during training and evaluation. Sensitivity (Recall): The model's ability to correctly identify cancerous cases. Specificity: The model's ability to correctly identify non-cancerous cases.

#### **D. Model Architecture**

The proposed convolutional neural network (CNN) consists of three convolutional blocks followed by two fully connected layers. Regularization techniques and appropriate activation functions were utilized to enhance

the model's performance and prevent overfitting. The first block begins with a convolutional layer containing 32 filters of size

lower at 98.74%, with recall at 98.26%, precision again at 100%, and an F1-Score of 98.36%.

cancer detection systems, given its potential for real-world application in clinical environments. For further details, you can refer to the original study published in the journal *Diagnostics*. (2023) [16] [17] [18]

## **Methodology**

### **A. Data Collection and Preprocessing**

The dataset used in this study was obtained from the IQ-OTH/NCCD Lung Cancer Dataset available on Kaggle. This dataset includes CT scan images of lung cancer patients and healthy individuals. The original resolution of the images is 512x512 pixels, which were resized to 256x256 pixels for computational efficiency and standardization across the dataset. Additionally, normalization was applied to scale pixel values within a consistent range (e.g., [0, 1]) to accelerate training and help the CNN model converge faster.

### **B. Training Process**

The data was split into training and testing sets using a train-test split, with additional cross-validation to ensure generalizability. The model was trained with a batch size of 32 and a learning rate of 0.001. To optimize the model's weights, the Adam optimizer was used, and the categorical cross-entropy loss function was applied due to the multi-class classification task. Training was conducted over multiple epochs with early stopping to avoid overfitting.

When comparing our results with those reported in the literature, specifically the LCDctCNN model, it is evident that both models exhibit strong diagnostic capabilities. The

### **C. Evaluation Metrics**

The performance of the model was evaluated using several metrics:

**Accuracy:** The overall percentage of correctly improve both precision and recall, ensuring that critical cases are identified with greater accuracy.

LCDctCNN reported an accuracy of [insert LCDctCNN accuracy here], with precision, recall, and F1-Score values of [insert LCDctCNN precision, recall, and F1-Score here]. While our training metrics are slightly higher, the validation accuracy of 98.74% suggests that our model maintains high generalization capability on unseen data.

The high precision of 100% across both training and validation datasets indicates that our model effectively minimizes false positives, which is critical in clinical settings where misdiagnosis can lead to unnecessary treatments. The recall, although slightly lower in the validation set, remains robust at 98.26%, reflecting the model's strong ability to identify positive cases of lung cancer.

One of the strengths of our model is its high accuracy and precision, which may stem from These results highlight the potential of our CNN-based model to assist radiologists in diagnosing lung cancer



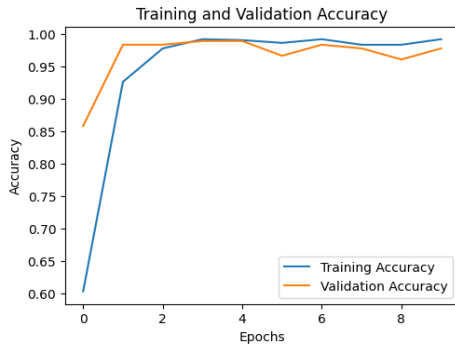
more effectively. By reducing false positives and maintaining high sensitivity, the model could lead to improved patient outcomes and more efficient resource allocation in healthcare settings. Future work could focus on enhancing model robustness by incorporating a larger and more diverse dataset or employing advanced techniques such as transfer learning. Additionally, exploring ensemble methods could further improve.

## Results and discussions

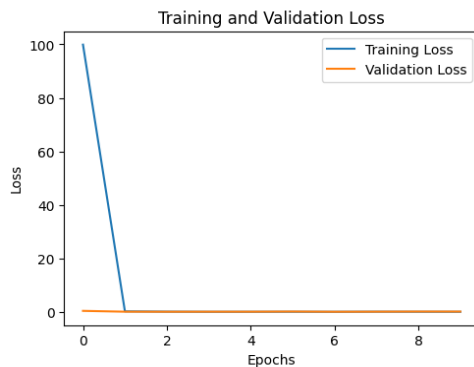
In this study, we presented our model for lung cancer diagnosis using CT scan images, achieving impressive performance metrics. The training accuracy was 99.43%, with both training recall and F1-Score reaching 99.43%.

Table 1: the comparison between the model implemented and the original paper

Metrics	VGG16+mlp	Customed CNN
accuracy	99.74%	98.74%
precision	98.33%	100%
recall	98.33%	98.26%
F1-score	98.33%	98.36%



**Fig 2: the accuracy curve over epochs.**



**Fig 3: the loss curve over epochs.**

## References

1. M. Mamun, A. Farjana, M. Al Mamun, and M. S. Ahammed, "Lung cancer prediction model using ensemble learning techniques and a systematic review analysis," 2022 IEEE World AI IoT Congress (AIoT), 2022, pp. 187-193, doi: 10.1109/AIoT55404.2022.9817326.
2. Makaju, Suren, P. W. C. Prasad, Abeer Alsadoon, A. K. Singh, and A. Elchouemi. "Lung cancer detection using CT scan images." *Procedia Computer Science*, vol. 125, pp. 107-114,
3. "Breast cancer statistics of American cancer society," [online]. Available: <https://www.cancer.org/cancer.html>. [Accessed: 13-Nov-2022].
4. Hany, M. (2020, August 20). Chest CT-scan images dataset. Kaggle. Available at: <https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images>. [Accessed: 13-Nov-2022].
5. M. Mamun, M. I. Mahmud, M. I. Hossain, A. M. Islam, and M. S. Ahammed, "Deep learning based model for Alzheimer's disease detection using brain MRI images," 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2022, (Preprint).
6. W. Ausawalaithong, W. Thirach, A. Marukatat, and S. Wilaiprasitporn, "Automatic lung cancer prediction from chest X-ray images using the deep learning approach," 2018 11th Biomedical Engineering International Conference (BMEiCON), 2018.
7. Abhir Bhandary, G. Ananth Prabhu, V. Rajinikanth, K. Palani Thangaraj, S. C. Satapathy, et al., "Deep-learning framework to detect lung abnormality – A study with chest X-ray and lung CT scan images," *Pattern Recognition Letters*, vol. 125, pp. 102-110, 2019.
8. Da Silva, G. L. F., da Silva Neto, O. P., Silva, A. C., et al., "Lung nodules diagnosis based on evolutionary convolutional neural network," *Multimed. Tools Appl.*, vol. 2017, pp. 19039–19055, 2017.

