Abstract:

1. lung cancer detection using two distinct datasets
⇨ lung cancer detection using two distinct datasets: the IQ-OTHNCCD lung cancer dataset and the LC25000 Lung Dataset.

2. For the IQ-OTHNCCD dataset, the CNN (ResNet50)
⇨ For the IQ-OTHNCCD dataset, the CNN (ResNet50) achieved an accuracy of 96.67%, F1-score of 96.67%, precision of 96.97%, and recall of 96.67%, while the Random Forest model yielded a perfect 100% across all metrics.

3. To enhance the performance
⇨ To enhance the performance of the Random Forest model, we implemented Genetic Algorithm (GA) optimization and K-fold cross validation, resulting in consistent 100% accuracy.

4. These findings emphasize the robustness of the Random Forest model
⇨ These findings emphasize the robustness of the Random Forest model and its superior performance in this domain while highlighting the potential of CNNs for lung cancer imaging analysis.

Introduction

5. Traditionally, lung cancer diagnosis

⇨ Traditionally, lung cancer diagnosis relies on chest X-rays and low-dose CT scans, interpreted by radiologists.

6. Recent advancements in imaging technologies, coupled with the rapid growth of artificial intelligence (AI)

⇨ Have greatly increased lung cancer diagnosis efficiency and accuracy.

7. Computed tomography (CT), and ultrasound (UI)—

⇨ Presents challenges for efficient and accurate analysis.

8. Computer-aided detection (CAD) systems, powered by neural networks and machine learning algorithms

⇨ Are designed to mimic human decision-making processes in identifying target sites within medical images

9. Transformers, with their capacity for parallel processing and capturing long-range dependencies,

⇨ Represent the next frontier in image analysis.

Literature review

10. Lung cancer is one of the deadliest malignancies globally,

⇨ Largely due to the late-stage diagnosis in many patients.

11. The advent of artificial intelligence (AI) and deep learning (DL)

⇨ Has presented new possibilities for overcoming these limitations, particularly through Convolutional Neural Networks (CNNs).

12. Random Forest (RF) models,

⇨ A machine learning alternative to CNNs, have also shown strong potential in lung cancer diagnosis.

13. Techniques such as Genetic Algorithms (GAs)

⇨ And K-fold cross-validation have been employed to further improve RF performance by optimizing hyper parameters and validating robustness across multiple data subsets

14. This offers a promising avenue for improving both

⇨ Lung nodule classification accuracy and model explain ability.

Methodology

15. The methodology comprises three distinct phases:

⇨ Defining research questions, establishing criteria for article selection, and conducting the review.

16. PRISMA

⇨

Inclusion Criteria	**Exclusion Criteria**
Articles published in peer-reviewed journals or presented at reputable conferences. Articles published in peer-reviewed journals or presented at reputable conferences. - Articles written in English or with English translations available.	Articles published in peer-reviewed journals or presented at reputable conferences. Studies unrelated to deep learning or lung cancer imaging analysis. - Articles written in languages other than English without available translations.

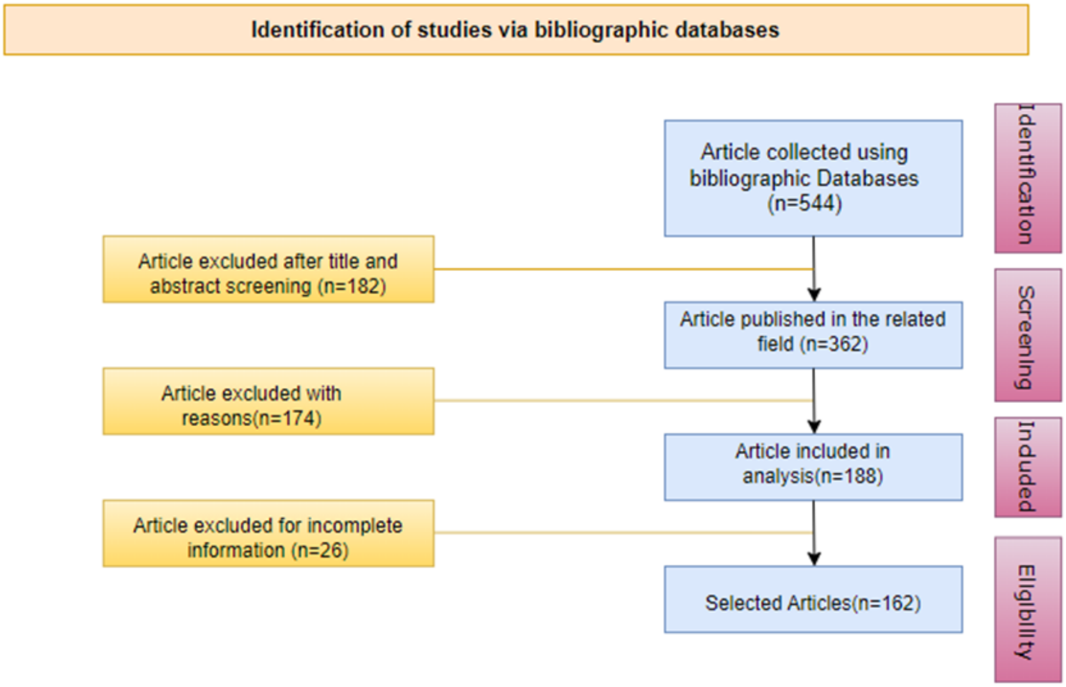


Fig. 1. The PRISMA diagram illustrates the process of selecting articles for applications and highlights the most recent developments in the field.

Traditional Methods for Lung Cancer Diagnosis

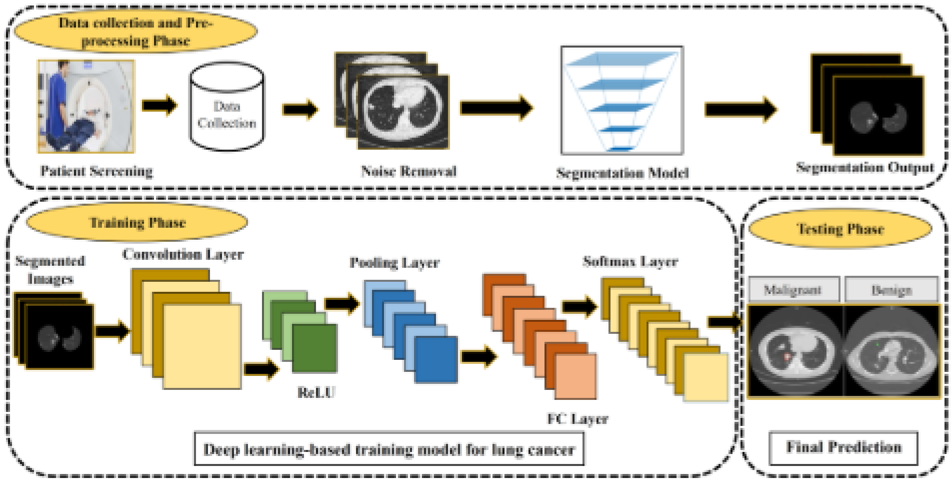
17. While chest X-rays are readily available and relatively inexpensive,

⇨ Their sensitivity for detecting small lung nodules, a hallmark of early-stage lung cancer, is limited

18. Additionally, inter-reader variability exists,

⇨ Meaning different radiologists may interpret the same scan differently, potentially leading to discrepancies in diagnosis.

⇨



19. DCNN designs typically consist of numerous

⇨ Layers that incorporate intricate nonlinear interactions.

20. Random Forest, a robust machine learning ensemble method,

⇨ Has gained significant traction in various data analysis tasks, including medical imaging.

21. Random Forest has proven effective in tasks such as

⇨ Image classification, anomaly detection, and feature selection.

= Lung Nodule Detection

= Lung Nodule Classification

= Data Availability and Quality

= Interpretability and Explain ability

= Generalization and Robustness

= Disease Description

Algorithms

22. **Deep Learning Algorithms**

⇨Reinforcement Learning (RL) involves agents interacting with environments to learn optimal policies for sequential decision making.

23. CNN Convolutional neural networks (CNN)

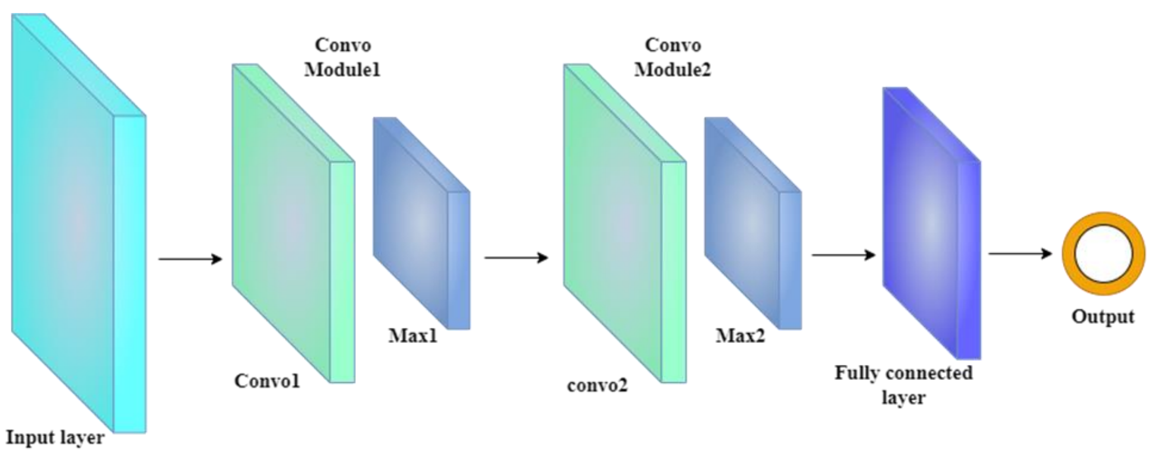


Fig. 3: Convolutional neural network

24.

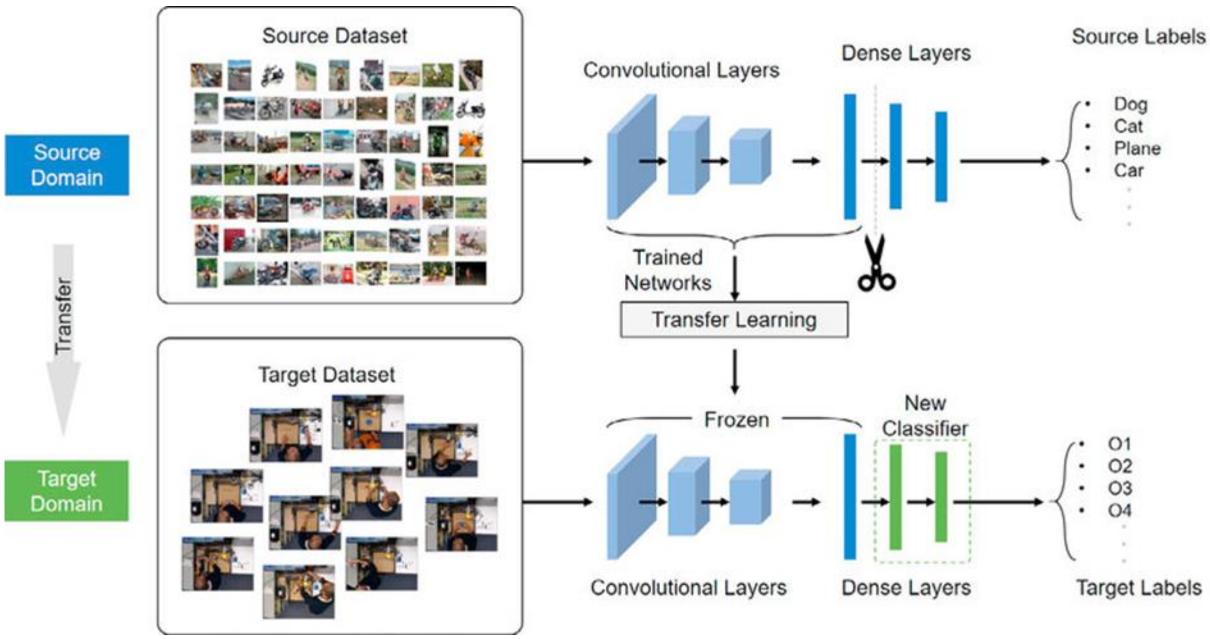
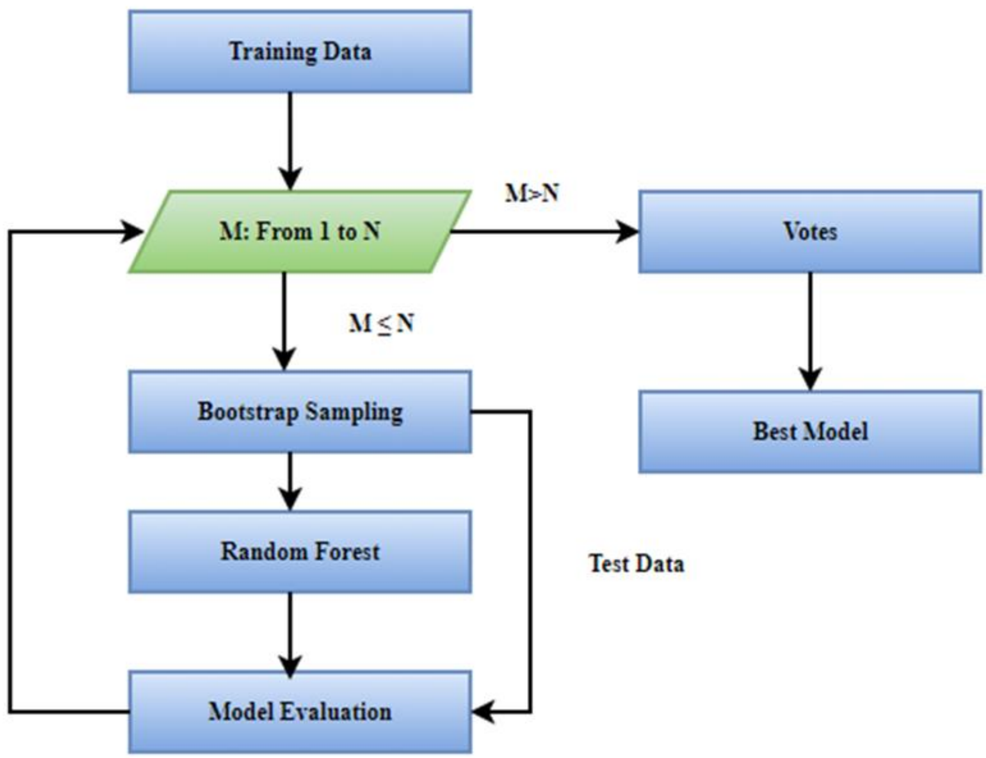


Fig. 4. Transfer Model Architecture (How it works).

5.2.1. Random Forest (RF)



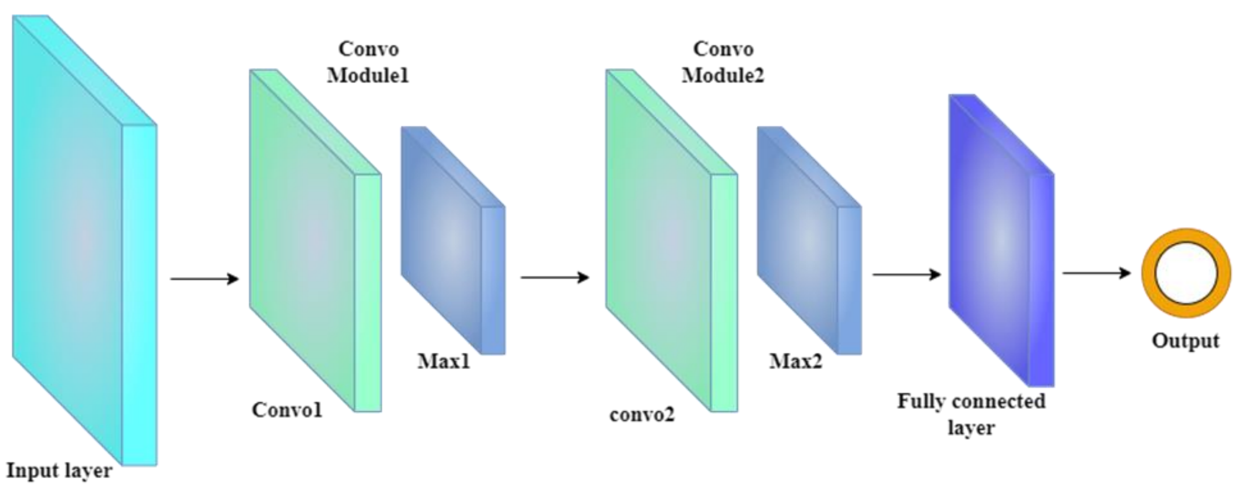


Fig. 3. Convolutional neural network

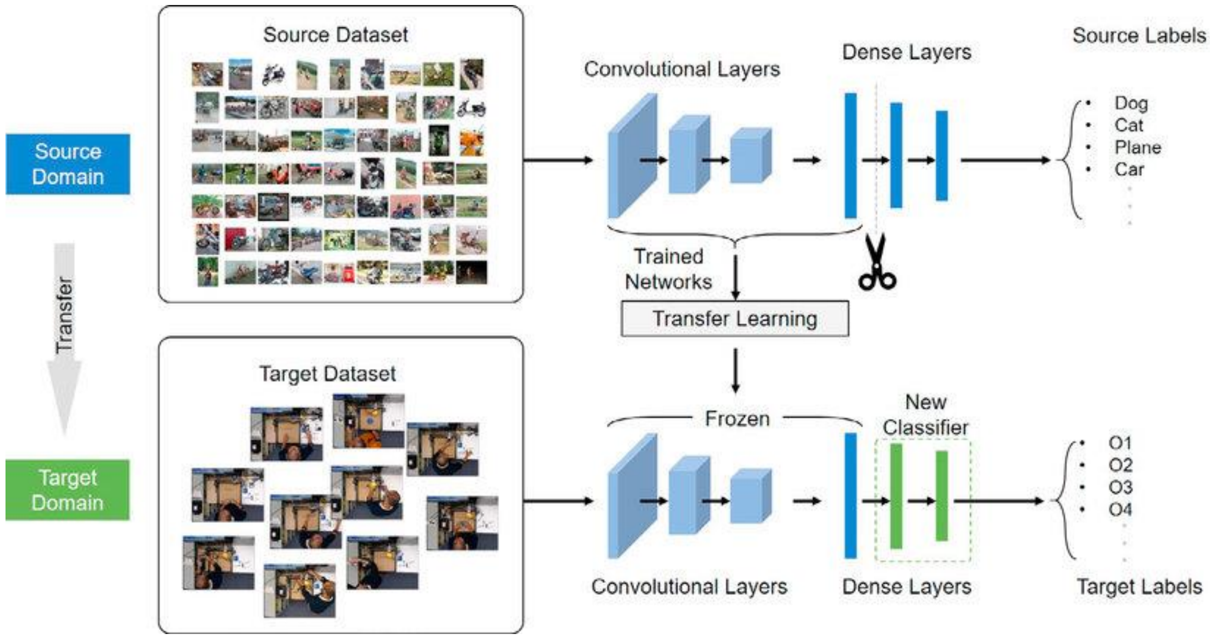


Fig. 4. Transfer Model Architecture (How it works).

5.2.1. Random Forest (RF)

The Genetic Algorithm (GA) is an evolutionary-based optimization technique that is designed to find the optimal solution by simulating the process of natural selection. In this context, GA is applied to optimize the Random Forest model by adjusting important hyper parameters such as partition number (s), decision tree count (c), and sampling threshold (λ).

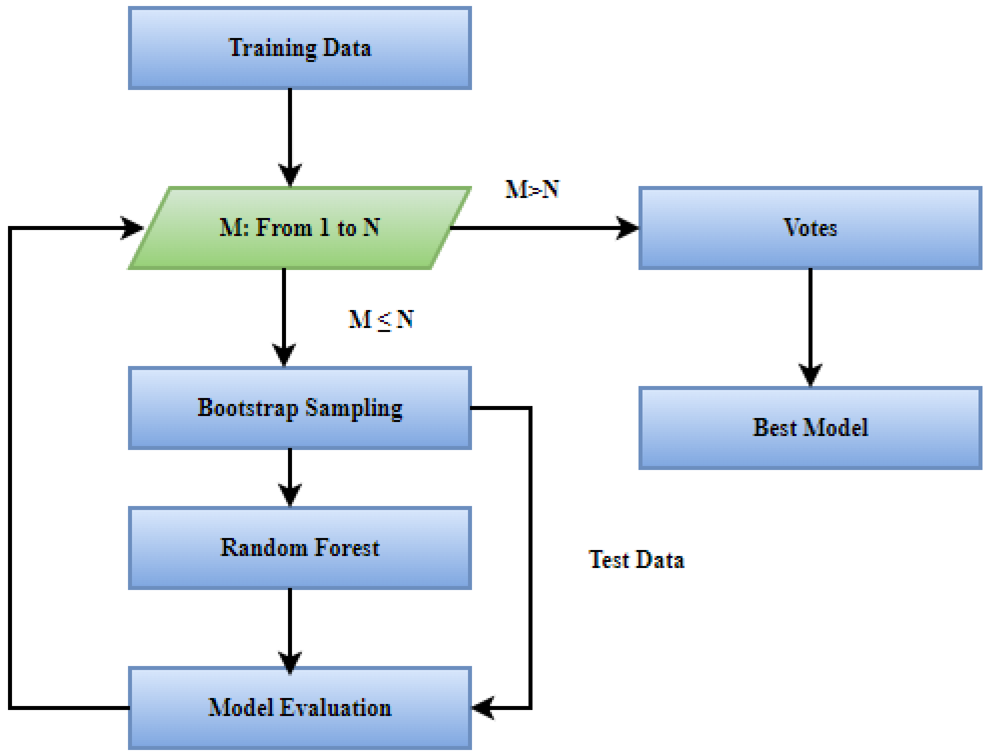


Fig. 5. Genetic Algorithm using Random Forest

Datasets

25. **Datasets** This work is primarily concerned with a detailed analysis of the IQ-OTH/NCCD

⇨ Lung cancer dataset, which is used to identify and classify lung cancers using deep learning techniques.

26. This dataset comprises a significant collection of preprocessed lung cancer images

⇨ Allowing for in-depth examination and enhancement of models designed for accurate disease detection.

27. Results Analysis

In this section, we present a detailed analysis of our experimental results using Convolutional Neural Networks (CNNs) and Random Forest models on the IQ-OTHNCCD lung cancer dataset and the LC25000 lung cancer dataset. Numerous studies have been conducted using these datasets, with the primary goal being to develop highly accurate models for early detection of the disease. In our research, we successfully achieved this goal with the Random Forest algorithm, which outperformed CNNs in terms of accuracy. While the CNN provided competitive results, it fell short in certain areas where Random Forest demonstrated superior performance, achieving more accurate results.

- ⇨ **Data Augmentation**
- ⇨ **Random Forest Classifier**
- ⇨ Hyper parameter Tuning

CNN Model		Result
	Accuracy	96.67%
	F1-Score	96.67%
	Recall	96.67%
	Precision	96.97%
Random Forest		
	Accuracy	100%
	F1-Score	100%
	Recall	100%
	Precision	100%

Fig. 6. Lung Cancer image: Benign cases.

Fig. 7. Lung Cancer image: Malignant cases.

Fig. 8. Lung Cancer image: Normal cases

Discussion

This study advances the application of deep learning in lung cancer imaging analysis, comparing the performance of CNNs and Random Forest models on two distinct datasets: the IQ-OTHNCCD and LC25000 Lung Datasets. Our findings suggest that Random Forest, particularly when enhanced with optimization techniques like Genetic Algorithms (GA) and K-fold cross-validation, offers exceptional accuracy in classifying lung cancer subtypes.

Conclusion

Lung cancer continues to be a leading cause of mortality worldwide, necessitating the development of accurate and reliable diagnostic tools. In this study, we applied both Convolutional Neural Networks (CNNs) and Random Forest models to classify lung cancer subtypes using two distinct datasets: the IQ-OTHNCCD lung cancer dataset and the LC25000 Lung Dataset. Our findings revealed that the Random Forest model consistently achieved perfect performance, with 100% accuracy, F1-score, precision, and recall across both datasets.

- **Results Analysis}**
- **Discussion} Are my part**
- **Conclusion}**