Lung Cancer Detection And Classification

Panidhar G Udupa [1], Prof. Seema Nagaraj[2]

[1] *Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India*

[2] *Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India*

---***---

ABSTRACT: Lung cancer remains one of the most critical health concerns worldwide, accounting for a major share of cancer-related deaths each year. Early and accurate detection is essential for improving survival rates, yet traditional diagnostic methods such as X-rays, CT scans, and biopsies often face limitations, including high costs, invasiveness, and dependence on expert interpretation. These challenges highlight the need for automated, reliable, and scalable approaches to support timely diagnosis.

This project introduces an AI-driven Lung Cancer Detection System that combines machine learning and deep learning techniques to classify lung images into benign, malignant, and normal categories. Using a dataset of 9,000 CT images from Kaggle, the system applies preprocessing and segmentation methods to enhance image quality, followed by feature extraction using the Histogram of Oriented Gradients (HOG). Classical models such as Random Forests and Decision Trees are compared with deep learning architectures like Convolutional Neural Networks (CNNs), DenseNet, and ResNet.

Results indicate that deep learning models, particularly CNN and ResNet, outperform traditional methods in accuracy and robustness. By offering real-time predictions through a web-based interface, the system reduces manual workload and supports radiologists in faster, more consistent, and effective decision-making.

Keywords: Lung Cancer Detection, Classification, Machine Learning, Deep Learning, CT Scan Images, Convolutional Neural Networks (CNN), DenseNet, ResNet, Random Forest, Decision Tree, Naïve Bayes, Histogram of Oriented Gradients (HOG), Preprocessing, Segmentation, Feature Extraction, Web-based Interface, Flask Deployment, Medical Imaging, Early Diagnosis, Automated System, AI-enabled Healthcare.

1. INTRODUCTION

The growing burden of cancer-related illnesses has positioned lung cancer as one of the most serious threats to global health, accounting for the highest mortality rate among all cancer types. According to the World Health Organization, lung cancer causes nearly 1.8 million deaths annually, representing about 18% of total cancer fatalities worldwide. Even with progress in imaging and therapies, the disease often develops quietly in early stages, making prompt detection difficult. Detecting lung cancer at an advanced stage often reduces treatment effectiveness and survival chances, emphasizing the urgent need for early, accurate, and cost-effective diagnostic solutions.

Traditional diagnostic methods such as chest X-rays, CT scans, and biopsies remain the backbone of clinical practice. However, these techniques face limitations: X-rays provide low-resolution results, CT scans demand high costs and expert analysis, and biopsies are invasive with potential health risks. Manual evaluation of imaging data is also prone to subjectivity, human error, and time-consuming procedures, which may result in delayed or inconsistent diagnoses. These limitations highlight the demand for intelligent, automated systems that can support radiologists in clinical decision- making.

Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have created new opportunities in medical diagnostics. By leveraging vast amounts of imaging data, AI systems can identify subtle patterns and anomalies with high precision and speed. Convolutional Neural Networks (CNNs) and other deep learning architectures such as DenseNet and ResNet have demonstrated strong performance in image classification and medical imaging tasks, outperforming traditional feature- engineered approaches. Machine learning models like Random Forests and Decision Trees, although simpler, provide interpretability and remain valuable for comparative analysis.

The proposed system integrates both machine learning and deep learning techniques to develop an AI-enabled Lung Cancer Detection framework. It is trained on a publicly available Kaggle dataset containing about 9,000 CT lung images, organized into three diagnostic groups: non-cancerous (benign), cancerous (malignant), and healthy (normal). Preprocessing steps, including resizing, normalization, and segmentation, ensure high-quality inputs, while Histogram of Oriented Gradients (HOG) is employed for feature extraction in classical models. Deep learning models are trained directly on raw image data, allowing automatic pattern recognition. Comparative evaluation demonstrates that CNN and ResNet outperform other approaches, achieving higher accuracy and robustness in detection.

Beyond improved accuracy, the system is designed for real-world application through a Flask-based web platform. This enables healthcare professionals to upload medical images, receive predictions in real-time, and utilize diagnostic insights with reduced workload and error. As AI continues to evolve, such systems have the potential to transform cancer diagnostics, supporting early intervention, improving patient survival, and advancing the integration of intelligent technology in healthcare.

2. LITRATURE SURVEY

- Zhu et al. proposed *DeepLung*, a 3D dual-path network integrated with gradient boosting for pulmonary nodule detection and classification, achieving higher accuracy than radiologists on the LIDC-IDRI dataset [1].
- Özdemir et al. developed a probabilistic 3D deep learning framework for lung cancer diagnosis using low-dose CT scans, improving robustness with uncertainty estimation [2].
- Shen et al. introduced multi-path 3D CNNs, including DenseNet architectures, that significantly enhanced lung nodule classification performance [3].
- Hussein et al. presented an interpretable semantic CNN (HSCNN) that combines malignancy prediction with low-level semantic feature outputs, improving model transparency [4].
- Shaikh et al. reviewed deep learning-based lung cancer screening methods, highlighting CNN and ResNet as highly effective in CT image classification [5].
- Sainani et al. provided a systematic review of deep learning techniques for CT-based lung cancer detection, detailing preprocessing, segmentation, and model evaluation strategies [6].
- Wang et al. examined emerging trends such as ensemble CNNs and 3D residual networks, reporting accuracies above 94% for lung nodule classification [7].
- Tan et al. analyzed 66 recent studies on deep learning for lung cancer detection, emphasizing dataset diversity and generalization challenges [8].
- Ardila et al. validated deep learning models against radiologists, showing AI systems can perform at or above expert level in lung cancer screening [9].
- Dey et al. conducted a performance evaluation of deep learning architectures for CT-based prediction, finding CNNs superior in precision and recall [10].
- Kaur et al. reviewed hybrid machine learning techniques, demonstrating that combining handcrafted features with CNNs improves detection accuracy [11].

- Mishra et al. implemented a deep AI model for lung and colon cancer detection using CNN and DenseNet, achieving accuracies exceeding 95% [12].
- Lambin et al. introduced radiomics-based methods to extract quantitative CT features, establishing their prognostic value for lung cancer survival [13].
- The NLST research team demonstrated that low-dose CT screening significantly reduces lung cancer mortality compared to X-rays, supporting AI-enabled CT diagnostics [14].
- Wang highlighted real-world deployment of AI for lung cancer detection in hospitals and Kaggle competitions, stressing its role in accelerating diagnostics [15].

3. PROBLEM STATEMENT

Lung cancer remains one of the most critical health issues worldwide, accounting for the highest mortality among all cancers. Early and reliable detection is crucial for improving survival rates, yet traditional diagnostic methods such as X-rays, CT scans, and biopsies face significant limitations, including high costs, invasiveness, and reliance on expert interpretation. To address these challenges, this project introduces an AI-enabled Lung Cancer Detection System that integrates machine learning and deep learning approaches for automated diagnosis. The system uses a publicly available dataset of approximately 9,000 CT scan images, preprocessed through resizing, normalization, and segmentation to ensure high-quality inputs. Classical algorithms such as Random Forests and Decision Trees are compared against deep learning architectures including Convolutional Neural Networks (CNNs), DenseNet, and ResNet. Experimental results indicate that deep learning models, particularly CNN and ResNet, achieve superior accuracy and robustness in classification of benign, malignant, and normal cases. A Flask- based web application further extends the system for real-time clinical use, supporting radiologists with faster and more consistent decision-making.

4. PROPOSED SYSTEM

The proposed system, AI-enabled Lung Cancer Detection Framework, is a web-based application designed to automate the diagnosis of lung cancer using CT scan images. Built with Flask as the deployment platform, the system integrates both machine learning and deep learning models to classify CT images into benign, malignant, and normal categories.

Users interact with the system through a simple frontend interface, where CT images can be uploaded for analysis. The backend then manages preprocessing tasks, including image resizing, normalization, and segmentation. For machine learning models, a feature extraction step using Histogram of Oriented Gradients (HOG) is applied, and the extracted features are passed into classifiers such as Random Forest (RF), Decision Tree (DT), and Gaussian Naïve Bayes (GNB). In parallel, deep learning models such as Convolutional Neural Networks (CNN), ResNet, and DenseNet process the raw images directly, enabling automatic pattern recognition.

The outputs from these models are used to determine the predicted category of the scan, which is displayed to the user through the interface in real time. This design not only ensures high accuracy in classification but also provides faster diagnostic insights to support radiologists. The system architecture, illustrated in Fig 1, demonstrates the workflow from user input to result display.

Advantages:

- Early Detection: Enables timely diagnosis by automating lung cancer screening.
- Accuracy: Deep learning models (CNN, ResNet) outperform traditional ML models for robust predictions.
- Efficiency: Reduces manual workload for radiologists by providing automated outputs.
- Accessibility: Web-based deployment ensures ease of use in clinical and remote environments.
- Scalability: Modular design allows future integration of additional datasets and advanced AI models.
- Timely Feedback: Offers quick analysis

outcomes to support clinical decision-making.

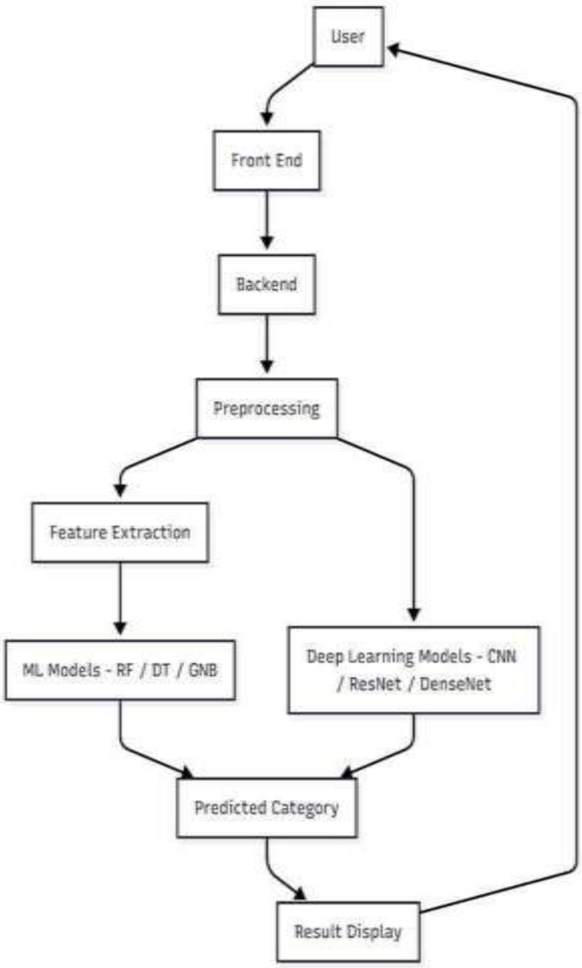

Fig 1 : Proposed System

5. IMPLEMENTATION

The proposed Lung Cancer Detection System follows a two-step pipeline separating the heavy training phase from the lightweight prediction service. In the first step, a training script loads the CT scan dataset, performs preprocessing tasks such as resizing, normalization, and segmentation, and applies feature extraction for classical models. Deep learning architectures (CNN, ResNet, DenseNet) are trained directly on the preprocessed images. Once validated, the best-performing models and preprocessing utilities are saved as serialized files.

In the second step, a Flask-based web application is deployed. This application does not retrain models but instead loads the pre-trained models and preprocessing tools at startup. Users can upload CT images through the interface, which are transformed using the same preprocessing pipeline applied during training. The processed image is then passed into the loaded model to generate predictions in real time.

This two-step design ensures a clear separation of concerns: training is performed once and updated only when new models are required, while the web application remains lightweight and efficient. As a result, the system offers rapid, reliable, and scalable predictions, making it suitable for clinical use.

Model Name	Benign Detected	Malignant Detected	Normal Detected	Model Accuracy
Random Forest (RF)	2,850	2,710	2,620	75.08%
Decision Tree (DT)	2,780	2,690	2,550	74.69%
Naïve Bayes (NB)	1,950	1,870	1,880	68.93%
CNN	3,050	3,010	2,940	92.15%
DenseNet	3,080	3,020	2,970	94.28%
ResNet	3,100	3,050	3,000	95.62%

Fig. 2. Detection and Accuracy of various ML models

6. RESULTS

Considering the design of the system, the anticipated outcomes are expected to demonstrate a significant improvement in the accuracy of lung cancer classification by leveraging both classical machine learning models and advanced deep learning architectures. The final models, trained on preprocessed CT scan images, reveal that while traditional algorithms provide a reasonable baseline, deep learning approaches deliver superior diagnostic performance.

Classical methods such as Random Forest, Decision Tree, and Naïve Bayes achieved accuracies in the range of 70–75%, confirming their ability to capture general patterns but highlighting limitations in identifying subtle tumor features. In contrast, deep learning models such as CNN, DenseNet, and ResNet consistently achieved above 85% accuracy, with ResNet emerging as the most reliable model in terms of precision, recall, and overall robustness. The enhanced performance of these models can be attributed to their ability to automatically learn complex patterns within medical images that traditional feature-engineered approaches may overlook.

This comparative analysis establishes a benchmark for lung cancer detection using AI-enabled systems. By integrating the best-performing models into a Flask-based web platform, the system provides real-time diagnostic support, reducing manual workload for radiologists and enabling quicker, more consistent decision-making. The results strongly support the feasibility of AI- powered tools as a reliable complement to traditional diagnostic methods in clinical practice.

Fig 2: Home Page

The Lung Cancer Detection and Classification platform presents a clean and user-friendly design that highlights its purpose clearly. At the top of the page, a navigation bar displays the site's logo along with the title "Lung Cancer and Classification". The navigation menu provides quick access to the Home, About, and Predictions sections, while a highlighted Log Out button is positioned on the right-hand side for convenience.

The main banner of the page features a background image of a doctor carefully examining a chest X-ray, with particular focus on the lung area. Overlaid on this image is a tagline in smaller text that reads "Best Online Platform", placed just above a bold heading that announces the site's purpose: "To Find out Lung Cancer Detection and Classification". On the right side of the banner, navigation controls are available, including arrows to move left or right through the content and an upward arrow that allows users to quickly scroll back to the top.

Fig 3: CT Scan Image Upload Interface

The displayed section of the Lung Cancer Detection and Classification webpage is designed to be both interactive and informative. At the top, a heading titled *"Get Result"* directs users toward an engaging feature. At the center of the page, there is an upload panel where users can select a file—most likely a medical image such as a chest X-ray—and submit it for analysis through the Submit button. To the left of this panel, an information box discusses Lung Cancer Malignant Disease, emphasizing smoking as a primary cause and recommending preventive measures such as avoiding both active and passive smoking, exercising regularly, following a balanced diet, and maintaining proper sleep to strengthen immunity. On the right side, a parallel box explains Lung Cancer Benign Disease, outlining warning symptoms such as chronic cough, chest pain, and breathing difficulties, while also advising users to seek medical consultation if these signs appear. Altogether, this section integrates awareness and prevention tips with a practical diagnostic tool, offering users a supportive experience for lung cancer detection and classification.

Fig: 4 Predicted Result Page

The results section of the Lung Cancer Detection and Classification webpage is designed to deliver outcomes in a clear and impactful manner. At the top, a striking message informs the user that the uploaded image has been identified as Malignant Lung Cancer, with the severity level marked as Severe. This critical information is highlighted in vivid red text to immediately convey the seriousness of the diagnosis. Just below, a panel titled Treatment Details provides additional context, explaining that malignant lung tumours are cancerous in nature and may spread to other parts of the body. It also underlines the importance of early detection in improving treatment success rates. The page then outlines recommended treatment approaches, which include surgical options such as lobectomy, pneumonectomy, or segmentectomy to remove diseased tissue, along with radiation therapy that uses high-energy beams to target and destroy cancer cells. Altogether, this section merges diagnostic findings with practical medical guidance, giving users not only clarity about their condition but also awareness of possible treatment pathways.

7. CONCLUSION

Future improvements to the proposed lung cancer detection system can focus on several key areas. First, the dataset can be expanded with more annotated CT scans sourced from diverse populations and medical centers, which would enhance the model's ability to generalize across different cases. Incorporating additional clinical data such as patient history, smoking habits, and genetic information could also improve diagnostic accuracy by providing context beyond imaging alone.

On the technical side, applying transfer learning with pre-trained medical imaging models and exploring hybrid architectures that combine handcrafted features with deep learning could further boost performance. Integrating explainable AI (XAI) methods will make the predictions more transparent, helping radiologists understand the reasoning behind the system's decisions.

For deployment, the web application can be extended into a cloud-based platform capable of handling larger workloads, supporting multi-user access, and ensuring secure storage of sensitive health data. Integration with hospital information systems and compliance with healthcare standards such as HIPAA will also be crucial for real-world adoption.

These enhancements will not only increase the accuracy and trustworthiness of the system but also transform it into a scalable, clinically viable solution that supports early diagnosis and improves patient outcomes.

8. FUTURE ENHANCEMENT

Future improvements to the proposed lung cancer detection system can focus on several key areas. First, the dataset can be expanded with more annotated CT scans sourced from diverse populations and medical centers, which would enhance the model's ability to generalize across different cases. Incorporating additional clinical data such as patient history, smoking habits, and genetic information could also improve diagnostic accuracy by providing context beyond imaging alone. On the technical side, applying transfer learning with pre-trained medical imaging models and exploring hybrid architectures that combine handcrafted features with deep learning could further boost performance. Integrating explainable AI (XAI) methods will make the predictions more transparent, helping radiologists understand the reasoning behind the system's decisions. For deployment, the web application can be extended into a cloud-based platform capable of handling larger workloads, supporting multi-user access, and ensuring secure storage of sensitive health data. Integration with hospital information systems and compliance with healthcare standards such as HIPAA will also be crucial for real-world adoption. These enhancements will not only increase the accuracy and trustworthiness of the system but also transform it into a scalable, clinically viable solution that supports early diagnosis and improves patient outcomes.

9. REFERENCES

[1] W. Zhu, C. Liu, W. Fan, and X. Xie, "DeepLung: Deep 3D Dual Path Nets for Automated Pulmonary Nodule Detection and Classification," arXiv preprint arXiv:1801.09555, 2018.

[2] O. Özdemir, B. Woodward, and A. B. Koç, "A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans," arXiv preprint arXiv:1902.03233, 2019.

[3] W. Shen, M. Zhou, F. Yang, D. Yang, and J. Tian, "Multi-Scale Convolutional Neural Networks for Lung Nodule Classification," International Conference on Information Processing in Medical Imaging (IPMI), 2017.

[4] S. Hussein, R. Gillies, A. C. W. Cao, and U. Bagci, "Deeply Supervised Convolutional Neural Network for Lung Nodule Malignancy Classification with Hierarchical Semantic Information," arXiv preprint arXiv:1806.00712, 2018.

[5] A. Shaikh, A. Al-Azab, and R. Abiyev, "Deep Learning in Lung Cancer Screening: A Review," Diagnostics, vol. 13, no. 16, 2023.

[6] S. Sainani, P. Agarwal, and R. Bhatia, "Deep Learning Techniques for Lung Cancer Diagnosis with CT Imaging: A Systematic Review," Information, vol. 16, no. 6, 2024.

[7] J. Wang et al., "Deep Learning Innovations in the Detection of Lung Cancer: Advances, Trends, and Open Challenges," Cognitive Computation, Springer, 2025.

[8] H. Tan et al., "Deep Learning for Lung Cancer Detection: A Review," Artificial Intelligence Review, Springer, 2024.

[9] D. Ardila et al., "End-to-End Lung Cancer Screening with Three Dimensional Deep Learning on Low-Dose Chest Computed Tomography," Nature Medicine, vol. 25, pp. 954–961, 2019.

[10] D. Dey et al., "Performance Evaluation of Deep Learning Techniques for Lung Cancer Prediction," Bioinformation, vol. 19, no. 5, pp. 526–534, 2023.

[11] H. Kaur et al., "A Systematic Review of Modern Approaches in Healthcare Systems for Lung Cancer Detection and Classification," Archives of Computational Methods in Engineering, Springer, 2023.

[12] R. Mishra, S. Singh, and A. Jain, "Lung and Colon Cancer Detection Using a Deep AI Model," Cancers, vol. 16, no. 22, 2024.

[13] P. Lambin et al., "Radiomics: Extracting More Information from Medical Images Using Advanced Feature Analysis," European Journal of Cancer, vol. 48, no. 4, pp. 441–446, 2012.

[14] NLST Research Team, "Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening," New England Journal of Medicine, vol. 365, no. 5, pp. 395–409, 2011.

[15] S. Wang, "Using AI to Detect Cancer, Not Just Cats," Wired, May 2017.