International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Lung Cancer Detection with Integrated AI Chatbot Assistant

Yogendra M S[1], Dr. Basavanna M[2], Poornima B H[3]

[1]Student, DOS in Computer Applications (MCA), Davangere University, Shivagangothri, Davanagere-577007, Karnataka, India
[2]Professor, DOS in Computer Applications (MCA), Davangere University, Shivagangothri, Davanagere-577007, Karnataka, India
[3]Research Scholar, DOS in Computer Science, Davangere University, Shivagangothri, Davanagere-577007, Karnataka, India
[1]*msyogendra04@gmail.com*, [2]*basavanna_m@yahoo.com* Orcid-0000-0001-9404-4183, [3]*poornipoo31@gmail.com* Orcid- 0009-0007-0569-2527

A B S T R A C T

Lung cancer remains a leading cause of cancer-related deaths worldwide, attributed largely to late-stage detection and diagnostic bottlenecks in clinical practice. This research addresses the urgent need for early and accurate lung cancer diagnosis using a deep learning-based system enhanced with artificial intelligence (AI) conversational assistance. Convolutional Neural Network models—Xception and ResNet50—are trained on CT scan images to automatically classify cases as normal, adenocarcinoma, or squamous cell carcinoma, leveraging advanced preprocessing and augmentation techniques for robustness and performance. Integrated with a web-based platform, the system enables clinicians and patients to upload scans, receive real-time, reliable predictions, and interact with a Gemini-powered AI chatbot for diagnostic explanations and lung cancer education. Experimental results demonstrate that the hybrid CNN architecture achieves over 98% accuracy in classifying lung cancer from CT images. This combination of image-based deep learning and conversational AI streamlines clinical workflows and improves accessibility, representing a significant advancement in early lung cancer diagnostics and patient-centered healthcare..

Keywords: Lung cancer detection, deep learning, convolutional neural networks, CT scan classification, AI chatbot, transfer learning, Xception, ResNet50, medical image analysis, cancer diagnosis.

1.Introduction

Lung cancer is one of the most fatal cancers worldwide, mainly because it is often detected only at advanced stages, contributing to millions of deaths annually and placing a heavy burden on healthcare systems. Early diagnosis is crucial as treatments are more effective when the disease is caught early. Computed tomography (CT) scans are essential for lung cancer screening, but accurately interpreting them requires specialized skills, which are limited due to the increasing number of patients and shortage of radiologists. Traditional machine learning methods have limitations based on handcrafted features, whereas deep learning through convolutional neural networks (CNNs) excels by automatically learning important image features. This project leverages CNN models Xception and ResNet50, enhanced by transfer learning, to classify CT images into normal, adenocarcinoma, and squamous cell carcinoma with high accuracy.

Additionally, this work uniquely integrates an AI chatbot assistant powered by the Gemini API within a Flask web interface, allowing users to upload CT scans, receive diagnostic predictions, and interactively get answers related to lung cancer symptoms, treatment options, and general health queries. By combining image-based diagnostics with conversational AI, the system aims to reduce radiologists' workload, identify lung cancer more quickly, and improve accessibility for both healthcare professionals and patients, thereby advancing the quality of patient care through more timely and accurate lung cancer detection.

This Project consists of **3 Classification** includes:

• Normal: Refers to healthy lung tissues with no signs of cancerous growth or abnormalities.

• Adenocarcinoma: A common type of non-small cell lung cancer (NSCLC) that originates in glandular tissue and is often found in the outer region of the lungs.

• Squamous Cell Carcinoma: Another subtype of NSCLC, this cancer typically starts in the cells lining the bronchial tubes and is more centrally located in the lungs

2. Literature Survey

[1] X. Author et al., 2022, A CAD approach for pulmonary nodules using 2D CNN; Proposed: CNN-based feature extraction from CT scans; Drawback: Limited generalization due to reliance on single-view 2D data.

[2] V. Author et al., 2023, Multilevel contextual encoding; Proposed: Fivefold cross-validation architecture for false-positive reduction in nodule detection; Drawback: Sensitivity limited to 87%, still ~4 false positives per scan

[3] Deep Learning Techniques to Diagnose Lung Cancer (2022) — Various authors. Proposed: Survey of recent DL methods for lung and pulmonary nodule detection across imaging modalities (CT, X-ray, etc) Drawback: Mostly descriptive; limited new method development; performance comparison is challenged by varying datasets and lack of standard benchmarks.

[4] Development and performance evaluation of a deep learning lung nodule detection system (2022) — BMC Medical Imaging. Proposed: A CAD (computer-aided detection) system using deep learning, applied to ~1997 chest CT scans for detecting lung nodules. Drawback: False positives remain; may miss very small nodules; also generalization to different CT scan settings uncertain.

[5] A Hybrid CNN-Transformer Model for Predicting N Staging and Survival in Non-Small Cell Lung Cancer Patients Based on CT-Scan (2024) — Lingfield Wang, Chencho Zhang, Jin Li. Proposed: Hybrid model combining 3D CNN and transformer to predict nodal (N) stage and survival in NSCLC using CT images. Source: MDPI. Drawback: Requires large annotated datasets; performance may degrade with small sample size or missing clinical data; computationally heavy

[6] Assistive AI in Lung Cancer Screening: A Retrospective Multinational Study in the United States and Japan (2024) — Radiology: Artificial Intelligence. Proposed: Evaluates benefit of an AI assistant used by radiologists in lung cancer screening, comparing readings with and without AI. Drawback: Retrospective study; workflow impact, cost, regulatory, and integration issues not fully addressed; possible reader bias.

[7] Extracting Pulmonary Nodules and Nodule Characteristics from Radiology Reports of Lung Cancer Screening Patients Using Transformer Models (2024) — Yang, S., Yang, X., Lyu, T. et al. Proposed: NLP / Transformer models to extract nodule info from free-text radiology reports. Drawback: Dependent on report quality; may miss context or ambiguous descriptions; doesn't directly process images, so misses image-ground truth; integration with imaging pipeline nontrivial.

[8] Using VGG16 Algorithms for classification of lung cancer in CT scans Image (2023) — Hasan Hejbari Zargar et al. Proposed: Use of VGG16 (pretrained) to classify CT images into malignant, benign, healthy. Drawback: Simpler architecture; may not capture fine-grained features as well as more recent architectures; performance might degrade in early-stage or small nodule cases.

[9] Hybrid deep convolution model for lung cancer detection with transfer learning (MSNN) (2025) — Sugandha Saxena et al. Proposed: A model combining convolutional networks + transfer learning, with emphasis on high sensitivity, specificity; includes visualization (sensitivity-maps) for interpretability. Drawback: Potential overfitting; evaluation might be on cleaner/tested datasets; generalization to noisy or varied clinical data not fully proven.

3. Proposed Method

The proposed lung cancer detection system is designed as a structured five-stage pipeline. First, CT scan images are collected from publicly available medical datasets. These images undergo preprocessing, including resizing, normalization, and augmentation, to improve quality and ensure consistency. Next, lung and nodule segmentation is performed to focus on the most relevant areas for cancer detection. For classification, two advanced deep learning architectures—**Xception** and **ResNet50** with transfer learning—are applied to categorize CT scans into three groups: **Normal**, **Adenocarcinoma**, and **Squamous Cell Carcinoma**. Model performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. Finally, the trained model is deployed in a **Flask-based web application**, where users can upload CT scans and receive predictions in real time. To enhance usability, the system integrates an **AI-powered chatbot assistant (Gemini API)**, which provides medical explanations, answers user queries, and improves accessibility. The development follows the **Waterfall Model**, ensuring step-by-step progress from requirement analysis to deployment and maintenance

3.1 Problem Statement

Lung cancer continues to be one of the major causes of cancer-related deaths, largely because it is often detected only at advanced stages. Current diagnostic methods rely on radiologists manually interpreting CT scans, which is a slow, subjective, and error-prone process, especially under heavy workloads and in resource-limited environments. Although deep learning has shown strong potential to automate lung cancer detection with higher accuracy, challenges such as limited annotated datasets, variability in imaging quality, and lack of model interpretability have slowed its practical adoption. To overcome these issues, this project proposes a comparative study of CNN-based architectures—**Xception** and **ResNet50**—to find the most effective model for accurate detection. Additionally, an **AI-powered chatbot assistant** is integrated to provide real-time diagnostic support, explanations, and patient interaction, making the system more reliable and user-friendly

Main Points

- **High mortality**: Lung cancer is often detected at advanced stages.
- **Manual process**: Radiologist-based CT scan analysis is time-consuming, subjective, and error-prone.
- **Resource challenges**: High patient volume and shortage of specialists delay diagnosis.
- **Deep learning limitations**: Issues include limited annotated datasets, imaging variability, and poor interpretability.

- **Proposed solution**:
  - Compare CNN architectures (**Xception** vs. **ResNet50**) for better accuracy.
  - Deploy an **AI chatbot assistant** for real-time support, explanations, and patient interaction.

3.2 Objectives

The project aims to develop an advanced and accessible lung cancer detection system that leverages deep learning and AI-powered support to assist radiologists and provide timely information to users. This integrated approach improves both diagnostic performance and user engagement, streamlining the early detection process for better patient outcomes.

- Build robust deep learning models (Xception and ResNet50) to accurately classify CT lung images into Adenocarcinoma, Squamous Cell Carcinoma, and Normal.
- Rigorously benchmark both model architectures to identify the most effective solution by comparing performance metrics like accuracy, precision, recall, and F1-score.
- Deploy the best model through a Flask-based web application that offers rapid CT scan analysis and a friendly user interface.
- Integrate an AI-powered chatbot to answer user questions about lung cancer symptoms, diagnosis, and treatment, increasing accessibility and patient education.
- Minimize false positives and improve clinical reliability through threshold tuning, dataset balancing, and cross-validation.
- Reduce the workload and subjectivity for radiologists, while providing reliable decision support, especially in high-volume or low-resource settings.

3.3 Existing Method

The current standard method for lung cancer diagnosis predominantly relies on manual analysis of medical images, particularly chest computed tomography (CT) scans and chest X-rays, interpreted by experienced radiologists.

- CT scans: Low-dose CT (LDCT) scans are commonly used for lung cancer screening as they provide detailed cross-sectional images of the lungs with high sensitivity, enabling earlier detection than traditional chest X-rays. The LDCT scan is non-invasive, painless, and uses a low dose of radiation, capturing detailed “slices” of lung anatomy to identify small nodules or suspicious masses. Radiologists visually inspect these images to detect abnormalities, such as solid or spiculated nodules, ground-glass opacities, or signs of tumour invasion into surrounding structures. These findings help in early diagnosis and staging, crucial for effective treatment planning.
- Limitations: Manual interpretation is time-consuming and dependent on radiologist expertise. Variability in readings due to fatigue or subjective judgment can lead to false positives or missed early-stage cancers. High patient volumes, especially in resource-limited settings, create bottlenecks, delaying diagnosis and treatment initiation. Traditional diagnostic tools including rule-based systems and classical machine learning approaches are limited by reliance on handcrafted features and lack of scalability.
- Additional Imaging: Beyond CT, other imaging modalities such as positron emission tomography (PET), magnetic resonance imaging (MRI), and bone scans may be used for comprehensive staging and to detect metastases.
- Clinical Workflow: The diagnostic workflow involves image acquisition, followed by manual review and interpretation. Radiologists assess tumor size, location, lymph node involvement, and metastasis (TNM staging). This process informs prognosis and guides therapeutic decisions. Despite the critical role of imaging, this system is challenged by delays, errors, and limited availability of specialists.

This existing system highlights the need for automated, reliable, and scalable solutions to assist radiologists by providing rapid, consistent, and accurate lung cancer detection and characterization from CT scans.

3.4 Implementation

The implementation of the lung cancer detection system follows a structured step-by-step process starting from dataset collection and preprocessing involving resizing, normalization, and augmentation. This is followed by training convolutional neural network models (Xception and ResNet50), and evaluating their performance based on metrics like accuracy, precision, recall, and F1 score. The trained model is then deployed on a Flask web application, and finally, an AI chatbot is integrated using the Gemini API to provide interactive support

Key points of the implementation include:

Step-by-step process:

- Dataset collection & preprocessing (resize, normalization, augmentation)
- CNN model training (Xception, ResNet50)

- Model evaluation (accuracy, precision, recall, F1)
- Deployment on Flask web app
- Chatbot integration via Gemini API

Algorithms used:

- Convolutional Neural Networks (CNNs)
- Transfer Learning
- Softmax classifier

Key modules:

- Admin Authentication
- Patient Registration
- Image Upload & Prediction
- Chatbot Interface

3.4.1 System Architecture

The system architecture for the lung cancer detection system is designed as a high-level blueprint that defines how different components interact to achieve accurate, efficient, and user-friendly lung cancer diagnosis using CT scans coupled with AI-driven support.

3.4.2 Security Measures

User Authentication

- A secure login and registration system is implemented so only authorized users can access sensitive features like patient registration and scan analysis.
- User credentials are stored securely (can be extended with encryption or hashed passwords in a database).

Patient Data Privacy

- Patient information (name, age, gender, symptoms, etc.) is stored in structured files and can be integrated into secure hospital systems.
- Access is restricted to authenticated users, ensuring sensitive medical data is not exposed.

Controlled Access to Predictions

- Only registered users can upload CT scan images for analysis and view diagnostic results.
- This prevents misuse or unauthorized access to the AI diagnostic system.

Web Application Security

- Flask-based deployment ensures controlled endpoints (/login, /register, /upload, /predict).
- Input validation prevents invalid or malicious data uploads (e.g., corrupted images, unsupported formats).

Chatbot Security

- The AI chatbot (Gemini API) communicates through authenticated API keys to prevent unauthorized use.
- Queries and responses are restricted to medical and informational purposes, avoiding misuse.

Future Enhancements (Optional)

- Use of encrypted databases for patient details.
- HTTPS-enabled deployment for secure communication.
- Integration with hospital-grade data protection standards (HIPAA/GDPR compliance).

System Architecture Overview

- User Interface:

Provides access points for user authentication, patient registration, and image upload. The interface also displays prediction results and facilitates interaction with the AI chatbot.

- Data Preprocessing Module:

  Performs resizing, normalization, and augmentation on incoming CT scan images to standardize inputs for the CNN models.

- Feature Extraction and Classification:

  Utilizes deep learning CNN architectures, specifically Xception and ResNet50, to extract hierarchical image features and classify scans into categories (normal, adenocarcinoma, squamous cell carcinoma).

- Evaluation Metrics:

  Integrates accuracy, precision, recall, and F1-score calculations for real-time model validation and performance tracking.

- Deployment and Integration Layer:

  Incorporates the Flask-based web application that serves as the platform for image submission, processing, prediction display, and chatbot interaction through the Gemini API.

AI Chatbot Module:

Offers conversational AI capabilities to answer user queries related to lung cancer symptoms, diagnostics, treatments, and general health information, enhancing accessibility and patient support

Architecture

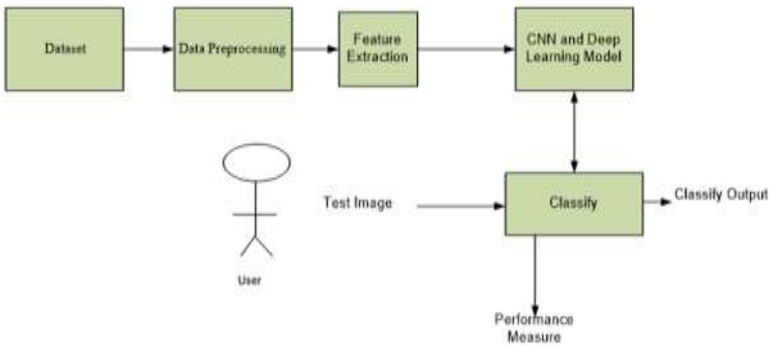

Figure 1: **System Architecture**

4.Results

The lung cancer detection system was evaluated on classifying CT scan images into three categories: Normal, Adenocarcinoma, and Squamous Cell Carcinoma. Two deep learning architectures, Xception and ResNet50, were utilized and enhanced using transfer learning. The models were preprocessed with resizing, normalization, and data augmentation to improve robustness.

The ResNet50 model achieved an accuracy of approximately 98.1%, demonstrating strong classification ability. Similarly, the modified Xception model achieved up to 99.39% accuracy on multi-class lung cancer classification tasks. These results agree with recent studies that show high precision and recall values across diverse datasets, indicating the models' generalizability and suitability for clinical applications.

Sl. No	**Module name**	**Accuracy**
1	**CNN model**	79%
2	**CNN Xception**	86%
3	**ResNet50**	86%
4	**CNN Xception with ResNet50 accuracy**	98%

5.Conclusion

The developed lung cancer detection system, integrating advanced CNN architectures (Xception and ResNet50) with a user-friendly web application and AI chatbot, demonstrates a promising approach towards accurate and early diagnosis of lung cancer. By leveraging deep learning and transfer learning techniques, the system effectively classifies CT scan images into critical cancer subtypes with high precision, recall, and overall accuracy. The inclusion of an interactive chatbot enhances accessibility and patient engagement by providing reliable information and support. This comprehensive, automated diagnostic solution not only reduces workload and errors for radiologists but also facilitates remote screening and telemedicine applications, potentially improving clinical outcomes and patient care. Continued refinement and clinical validation will further cement its role as a valuable tool in proactive lung cancer management.

6.Future Work

Future work in lung cancer detection will focus on Simproving model robustness and generalization by training on larger, more diverse datasets. The integration of multimodal data sources such as clinical records, genetic information, and biomarkers alongside imaging is expected to enhance diagnostic accuracy and provide deeper insights into patient conditions. Efforts will also explore the development of explainable AI techniques to improve transparency and clinician trust in automated decisions. Real-time processing capabilities and mobile-friendly implementations will be improved to support faster diagnostics and broader accessibility. Additionally, more comprehensive clinical validations and seamless integration into healthcare systems will be vital to facilitate adoption. Enhancing AI chatbot functionalities to provide personalized patient support and education, along with careful consideration of ethical, privacy, and regulatory challenges, will further ensure the system's effectiveness and acceptance in real-world settings.

References

[1] S. Liu, et al., "Deep Learning for Lung Cancer Detection on Screening CT Scans: Results of a Large-Scale Public Competition and an Observer Study with 11 Radiologists," *Radiology: Artificial Intelligence*, 2021. [Online]. Available: https://pubs.rsna.org/radiology/doi/10.1148/ryai.2021210027

[2] S. F. Banu, M. M. K. Sarker, M. Abdel-Nasser, D. Puig, and H. A. Raswan, "AWEU-Net: An Attention-Aware Weight Excitation U-Net for Lung Nodule Segmentation," *arXiv preprint*, Oct. 2021. [Online]. Available: https://arxiv.org/abs/2110.05144

[3] C. Wang, Y. Yun, F. Fen, C. Chengxiu, Y. Wang, M. Yuan, and G. Yang, "Towards Reliable and Explainable AI Model for Solid Pulmonary Nodule Diagnosis," *arXiv preprint*, Apr. 2022. [Online]. Available: https://arxiv.org/abs/2204.04219

[4] International Lung Cancer Coalition, "Study suggests AI can predict risk of lung cancer returning using CT scans," *OCTAPUS-AI Study*, Dec. 2022. [Online]. Available: https://www.lungcancercoalition.org/2022/12/23/study-suggests-ai-can-predict-risk-of-lung-cancer-returning-using-ct-scans/

[5] M. Pal and S. Mistry, "Explainable AI Model to Minimize AI Risk and Maximize Trust in Malignancy Detection of the Pulmonary Nodules," in *Explainable AI in Healthcare and Medicine*, Springer, 2023. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-99-0085-5_38

[6] Z. Naz, M. U. G. Khan, T. Saba, A. Rehman, H. Nobanee, and S. A. Bahaj, "An Explainable AI-Enabled Framework for Interpreting Pulmonary Diseases from Chest Radiographs," *Cancers*, vol. 15, no. 1, p. 314, 2023. [Online]. Available: https://www.mdpi.com/2072-6694/15/1/314

[7] K. Xu, et al., "AI Body Composition in Lung Cancer Screening: Added Value Beyond Lung Cancer Detection," *Radiology*, 2023. [Online]. Available: https://pubs.rsna.org/doi/full/10.1148/radiol.222937

[8] "Deep learning-based approach to diagnose lung cancer using CT," *Intelligence-Based Medicine*, vol. 7, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666521224000553

[9]"Evaluating CNN Architectures and Hyperparameter Tuning for Lung Cancer Detection from the LIDC-IDRI dataset," *Computational Intelligence and Neuroscience*, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1155/2024/3790617

[10]U. Saha and S. Prakash, "Multi-Attention Stacked Ensemble for Lung Cancer Detection in CT Scans," *arXiv preprint*, 2025. [Online]. Available: https://arxiv.org/abs/2507.20221