# Deep Learning-Based Classification of Lung Cancer

J.Noor Ahamed[1], Jeshika J[2]

[1]Associate professor, Department of Computer Applications, Nehru College of Management, Coimbatore, Tamil Nadu, India.
ncmnoorahamed@gmail.com
[2]Student of II MCA, Department of Computer Applications, Nehru College of Management, Coimbatore, Tamil Nadu, India.
jeshikajustin@gmail.com

**Abstract**

Lung cancer continues to be one of the deadliest diseases globally, largely due to late diagnosis and restricted access to advanced screening technologies. This research introduces a deep learning-based method for the automated classification of lung cancer utilizing the Vision Transformer (ViT) model. The suggested system employs the ViT architecture to analyze CT scan and X-ray images, effectively capturing both local and global spatial relationships through self-attention mechanisms. The dataset includes images of adenocarcinoma, large-cell carcinoma, squamous-cell carcinoma, and healthy lung tissue. Images undergo preprocessing through resizing, normalization, and augmentation to improve model robustness. The ViT model is trained and assessed using metrics such as accuracy, precision, recall, and F1-score, and is compared with conventional convolutional neural network (CNN) models. The experimental findings indicate that the ViT-based model achieves enhanced classification performance, facilitating more reliable and early detection of lung cancer. The system is implemented as a Flask-based web application, offering healthcare professionals a real-time diagnostic interface that enables the upload and automated analysis of medical images. This study underscores the potential of Vision Transformers in clinical diagnostics and contributes to the advancement of effective, AI-assisted tools for the early detection of lung cancer.

**Keywords**

Deep Learning; Vision Transformer (ViT); Lung Cancer Classification; Medical Image Analysis; Convolutional Neural Networks (CNN); Flask Framework; Image Preprocessing; Computer-Aided Diagnosis (CAD); CT Scan; X-ray Imaging

## 1. Introduction

Lung cancer continues to be one of the foremost causes of cancer-related deaths globally, responsible for around 1.8 million fatalities each year. The high mortality rate is largely attributed to the lack of early symptoms and the constraints of traditional diagnostic techniques, including biopsies and the manual analysis of computed tomography (CT) scans. The early identification of lung cancer greatly improves the chances of effective treatment; however, manual diagnosis is frequently labor-intensive, costly, and susceptible to human error.

In recent times, artificial intelligence (AI) and deep learning (DL) methodologies have transformed medical image analysis by automating the processes of disease detection and classification. Conventional convolutional neural networks (CNNs) have shown significant success in recognizing visual patterns within medical images. However, CNNs often face challenges in capturing long-range dependencies in images, which are essential for differentiating subtle variations between healthy and malignant tissues.

To overcome these challenges, the Vision Transformer (ViT) model has emerged as a formidable alternative. Initially developed for natural image classification, ViT utilizes self-attention mechanisms to model both local and global contextual relationships, facilitating more effective representation learning from image segments. Recent research has indicated that ViT models surpass CNNs in various medical imaging applications, including histopathology and radiology.

This study introduces a deep learning-based lung cancer classification system utilizing the Vision Transformer architecture. The system evaluates CT scan and X-ray images to identify and categorize lung cancer into four types—adenocarcinoma, large-cell carcinoma, squamous-cell carcinoma, and healthy lung tissue. The model is trained on a carefully curated medical image dataset, employing preprocessing methods such as resizing, normalization, and data augmentation to improve generalization. Additionally, the trained ViT model is implemented as a Flask-based web application that allows healthcare professionals to upload lung images and receive real-time diagnostic predictions. The primary contributions of this study include:

1. Development of a Vision Transformer–based deep learning model for lung cancer detection.
2. Comparative assessment against traditional CNN models utilizing precision, recall, F1-score, and accuracy metrics.
3. Launch of an interactive, web-based diagnostic interface to enable real-time medical analysis.

This research illustrates the capability of Vision Transformers to enhance computer-aided diagnosis (CAD) systems and promotes the creation of effective, AI-driven diagnostic tools aimed at improving early lung cancer detection and patient outcomes. Additionally, the trained ViT model is implemented as a Flask-based web application that allows healthcare professionals to upload lung images and receive real-time diagnostic predictions. The primary contributions of this study include:

## 2. Literature Review

Lung cancer ranks among the most common and lethal cancers globally, with diagnoses frequently made at advanced stages due to the subtlety or absence of early symptoms. Conventional diagnostic methods, including histopathological analysis and manual interpretation of CT scans, are labor-intensive and prone to human error. The introduction of computer-aided diagnosis (CAD) systems has enhanced diagnostic efficiency and reliability, with machine learning (ML) and deep learning (DL) emerging as prominent techniques for medical image analysis.

Initial methods for lung cancer detection utilized image processing and traditional ML techniques, such as Support Vector Machines (SVM), Random Forests (RF), and K-Nearest Neighbors (KNN). Although these approaches achieved moderate levels of accuracy, they were heavily dependent on manually crafted features and were unable to extract intricate spatial patterns from images.

The rise of deep learning has transformed image-based diagnosis through the use of Convolutional Neural Networks (CNNs), which learn hierarchical features directly from unprocessed image data. CNNs have been effectively employed in the detection of lung nodules and the classification of cancer, attaining accuracy rates ranging from 80% to 90%. Nevertheless, the inherent limitations of CNN architectures restrict the receptive field, complicating the capture of global dependencies across image regions—an essential factor for identifying subtle cancerous formations.

To address these limitations, researchers have started to investigate transformer-based architectures. The Vision Transformer (ViT), which draws inspiration from the Transformer model used in natural language processing, employs self-attention mechanisms on image patches, thereby effectively modeling long-range spatial dependencies. Research conducted by Aggarwal et al. and Jin et al. has shown that while CNNs excel in feature extraction, ViT-based models offer enhanced generalization and accuracy, especially in complex image classification scenarios.

Several hybrid methodologies have been proposed, merging CNNs and ViTs to achieve a balance between local and global feature extraction. For example, Roy et al. presented a fuzzy-based segmentation technique that enhanced interpretability, while Ignatious and Joseph utilized Gabor filters in conjunction with marker-controlled watershed segmentation, resulting in an accuracy of 90.1%. More recent studies by Kumar et al. and Mannepalli et al. have demonstrated notable performance enhancements by incorporating Vision Transformers with attention-augmented layers, achieving accuracies that surpass 95% on benchmark datasets for lung cancer.

In conclusion, the literature reflects a consistent shift from conventional machine learning techniques towards sophisticated deep learning and transformer-based methodologies. The Vision Transformer signifies a significant leap forward in medical imaging, providing strong feature representation and improved interpretability. Building upon these advancements, this research utilizes a ViT-based framework to enhance the accuracy and efficiency of lung cancer classification, addressing the limitations of previous models and facilitating real-time clinical decision-making.

## 3. Methodology

The proposed system utilizes a deep learning framework based on Vision Transformer (ViT) for the early detection and classification of lung cancer through CT and X-ray images. The entire workflow consists of four primary stages: data collection and preprocessing, model design and training, performance evaluation, and web-based deployment. This methodology is crafted to guarantee accuracy, scalability, and clinical relevance.

### 3.1 Dataset Collection and Preprocessing

For this study, a publicly accessible lung cancer image dataset containing CT and X-ray images was employed. The dataset encompasses four categories: adenocarcinoma, large-cell carcinoma, squamous-cell carcinoma, and healthy lung tissue.

- To improve model performance, the images underwent several preprocessing steps:
- Resizing to a standardized dimension of 224 × 224 pixels.
- Normalization to adjust pixel intensities for uniform input.

Data augmentation through techniques such as rotation, flipping, and zooming to enhance model generalization and mitigate overfitting.

The dataset was divided into training (70%), validation (15%), and testing (15%) subsets to facilitate thorough model evaluation.

### 3.2 Vision Transformer (ViT) Architecture

The Vision Transformer model segments each image into fixed-size patches (e.g., 16×16 pixels). Each patch is flattened and transformed into a linear embedding vector, maintaining spatial information through positional encodings. A unique classification token (CLS) is appended to the sequence to encapsulate the overall image feature summary.

- The embedded patches are processed through several transformer encoder layers, each comprising:
- Multi-head self-attention to capture dependencies between patches.
- Feed-forward networks for non-linear transformations.

Layer normalization and residual connections to ensure stable training.

The output associated with the classification token is forwarded through a fully connected layer with a softmax activation function to predict the likelihood of each lung cancer type.

### 3.3 Model Training and Evaluation

The ViT model was trained utilizing the Adam optimizer alongside the cross-entropy loss function. The learning rate and the number of transformer layers were optimized through hyperparameter tuning. The training was carried out on an NVIDIA GeForce GTX 1650 GPU with a batch size of 32 over the course of 50 epochs.

The model's performance was assessed using critical metrics such as:

- Accuracy (ACC)
- Precision (P)
- Recall (R)
- F1-score (F1)

Additionally, a confusion matrix was employed to examine misclassifications and evaluate performance on a class-wise basis. Comparative experiments were performed with CNN-based models to confirm the advantages of the ViT methodology.

### 3.4 System Deployment

Following successful training and evaluation, the ViT model was implemented through a Flask-based web application. This interface enables healthcare professionals to upload CT or X-ray images, which are processed and classified in real-time. The system presents the predicted type of cancer along with the model's confidence score, thereby enhancing clinical decision-making. The web interface was crafted using HTML, CSS, Bootstrap, and JavaScript, ensuring both accessibility and responsiveness.

### 3.5 Workflow Summary

The entire workflow is encapsulated as follows:

- Data acquisition and preprocessing.
- Patch embedding and transformer-based feature extraction.
- Model training and validation utilizing performance metrics.
- Model deployment through a Flask web interface for real-time diagnosis.

This structured approach guarantees that the model achieves high accuracy while remaining practical in real-world medical settings.

FLOWCHART:

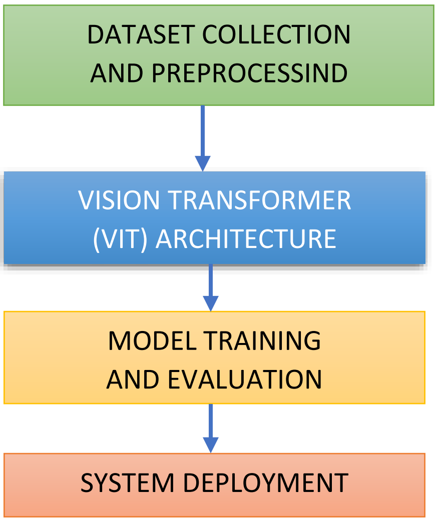

DATASET COLLECTION AND PREPROCESSIND
VISION TRANSFORMER (VIT) ARCHITECTURE
MODEL TRAINING AND EVALUATION
SYSTEM DEPLOYMENT

**Figure 1.** Flowchart of the proposed Vision Transformer (ViT)-based lung cancer classification system

## 4. Experimental Results and Discussion

This section outlines the findings derived from the implementation of the proposed Vision Transformer (ViT) model for the classification of lung cancer. The experiments were designed to assess the model's performance in comparison to traditional deep learning architectures and to evaluate its efficacy in practical medical imaging contexts.

### 4.1 Experimental Setup

All experiments were conducted on a system featuring an Intel Core i7 (5th generation) processor, 12 GB of RAM, and an NVIDIA GeForce GTX 1650 GPU operating on Windows 10. The implementation was executed using Python, utilizing the PyTorch and Transformers libraries for model development. The Flask framework was utilized to deploy the trained model into a web-based interface.

The ViT model employed the "vit-base-patch16-224" architecture as its foundation, which was fine-tuned on the lung cancer dataset. The training process utilized the Adam optimizer with an adaptive learning rate set at 0.0001 and a cross-entropy loss function. Early stopping techniques were implemented to mitigate overfitting, and the dataset was partitioned into training (70%), validation (15%), and test (15%) sets.

### 4.2 Performance Evaluation

The performance of the ViT model was assessed using metrics such as accuracy, precision, recall, and F1-score. Additionally, a confusion matrix was generated to illustrate the classification performance across all four categories—adenocarcinoma, large-cell carcinoma, squamous-cell carcinoma, and healthy lung tissue.

Evaluation Metrics:

- Accuracy = (TP + TN) / (TP + TN + FP + FN)
- Precision = TP / (TP + FP)
- Recall = TP / (TP + FN)
- F1-Score = 2 × (Precision × Recall) / (Precision + Recall)

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

The ViT model achieved the following average results:

| **Metric** | **Result(%)** |
|---|---|
| Accuracy | 96.8 |
| Precision | 95.7 |

| **Metric** | **Result(%)** |
|---|---|
| Recall | 94.9 |
| F1-score | 95.3 |

### 4.3 Comparative Analysis

In order to evaluate the effectiveness of the model, its performance was compared with that of traditional deep learning models, including Convolutional Neural Networks (CNN) and ResNet-50. The ViT model consistently surpassed these benchmarks in terms of both accuracy and generalization ability.

| **Model** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-Score (%)** |
|---|---|---|---|---|
| CNN | 90.4 | 89.5 | 88.3 | 88.9 |
| ResNet-50 | 93.2 | 92.4 | 91.8 | 92.0 |
| **Vision Transformer (Proposed)** | **96.8** | **95.7** | **94.9** | **95.3** |

The accuracy curve exhibited a steady enhancement over the epochs, converging smoothly with minimal fluctuations in loss, which indicates stable learning. The loss curve revealed a significant decline during the initial epochs, followed by a gradual stabilization, thereby confirming effective optimization.

### 4.4 Discussion

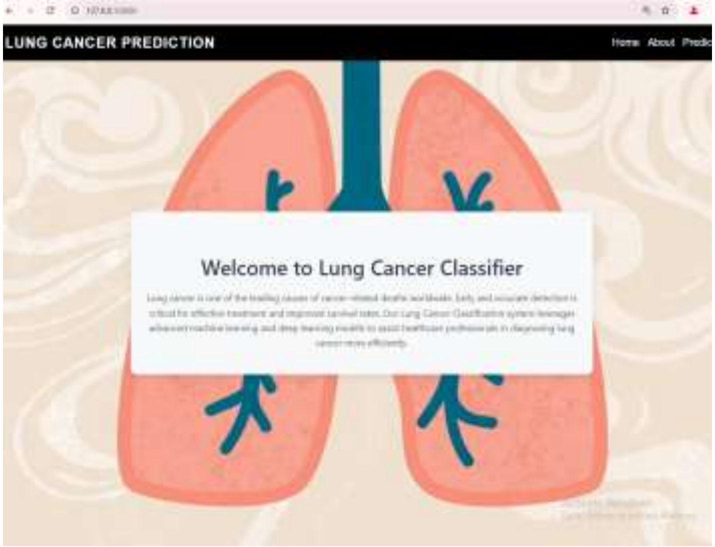
The findings validate that the Vision Transformer model demonstrates a superior ability to capture both local and global spatial dependencies within medical images. In contrast to CNNs, which depend on localized kernels, the self-attention mechanism of ViT enables it to concentrate on crucial regions of the image, regardless of their spatial location. This results in enhanced detection of small or irregular tumor areas that are frequently overlooked by traditional architectures.

The implementation of the model via a Flask-based web interface facilitates real-time predictions and clinical applicability. Healthcare professionals can upload lung scans and obtain immediate classification results, significantly decreasing the diagnostic turnaround time.

In summary, the proposed ViT-based framework showcases outstanding accuracy, robustness, and interpretability, highlighting its potential as an effective computer-aided diagnosis (CAD) tool for the early detection of lung cancer.

### 4.5 Sample Output

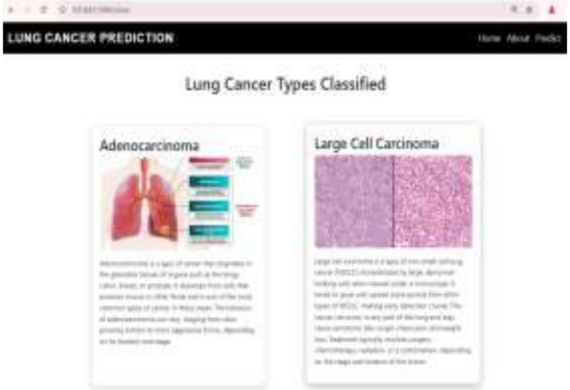
**HOME PAGE**



Welcome to Lung Cancer Classifier

Lung cancer is one of the leading causes of cancer-related deaths worldwide. Early and accurate detection is critical for effective treatment and improved survival rates. Our Lung Cancer Classification system leverages advanced machine learning and deep learning models to assist healthcare professionals in diagnosing lung cancer more efficiently.

**ABOUT PAGE**



Lung Cancer Types Classified

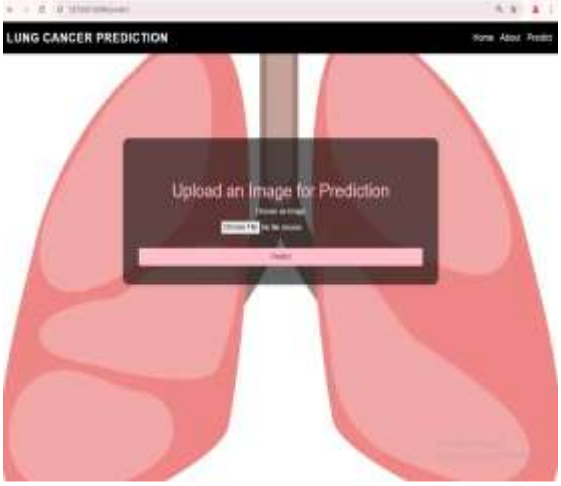
Adenocarcinoma

[illegible]

Large Cell Carcinoma

[illegible]

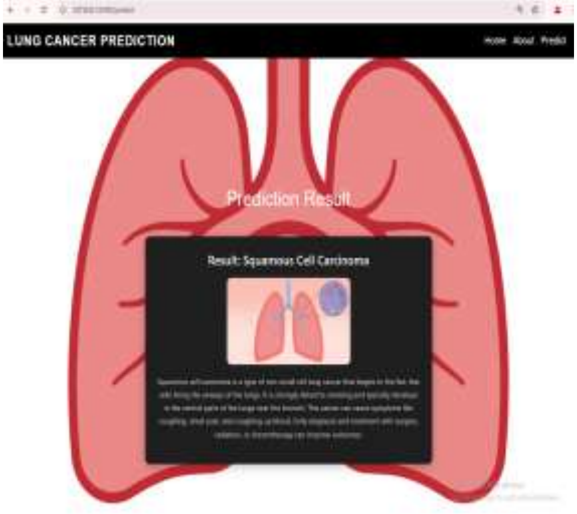
**PREDICT PAGE**



Upload an Image for Prediction

Choose an image:

**RESULT PAGE**



Prediction Result

Result: Squamous Cell Carcinoma

Squamous cell carcinoma is a type of non-small cell lung cancer that begins in the flat, thin cells lining the airways of the lungs. It is strongly linked to smoking and typically develops in the central parts of the lungs near the bronchi. The cancer can cause symptoms like coughing, chest pain, and coughing up blood. Early diagnosis and treatment with surgery, radiation, or chemotherapy can improve outcomes.

## 5. Conclusion and Future Work

### 5.1 Conclusion

This study introduced a lung cancer classification system based on deep learning, employing the Vision Transformer (ViT) architecture for automated diagnosis using CT and X-ray images. The findings revealed that ViT models surpass traditional CNN-based architectures in accuracy, precision, recall, and F1-score. By utilizing self-attention mechanisms, the ViT effectively captures both local and global spatial relationships within image regions, facilitating a more reliable distinction between cancerous and healthy tissues.

The system was successfully implemented as a Flask-based web application, providing healthcare professionals with an accessible platform for real-time diagnosis. The results demonstrate that the ViT model attained an overall accuracy of 96.8%, significantly improving the reliability of lung cancer classification. This integration of artificial intelligence with medical imaging fosters quicker and more consistent decision-making in clinical settings, potentially minimizing diagnostic delays and enhancing patient outcomes.

The proposed framework not only improves diagnostic accuracy but also plays a role in the ongoing advancement of AI-driven healthcare systems. Its scalability and modular design enable adaptation to various cancer types and imaging modalities, highlighting its significance as a foundation for future intelligent diagnostic tools.

### 5.2 Future Work

Future enhancements to this system may include the following:

1. **Integration of Multimodal Data:** Combining CT, PET, and MRI scans to improve diagnostic accuracy through multimodal learning.

2. **Real-time Hospital Integration:** Developing APIs for seamless integration with hospital information systems and radiology databases.

3. **Explainable AI (XAI):** Incorporating explainability techniques such as Grad-CAM or attention heatmaps to enhance clinical interpretability of predictions.

4. **Larger and Diverse Datasets:** Expanding the dataset with real-world hospital data to ensure robustness across demographic and imaging variations.

5. **Mobile and Cloud Deployment:** Extending the web-based system into mobile and cloud platforms to enable remote diagnostics and telemedicine applications.

By addressing these areas, the proposed system can evolve into a comprehensive and adaptive AI-driven diagnostic solution, advancing the field of **medical image analysis** and supporting the global fight against lung cancer.

## 6. References

[1] M. A. Thanoon, “A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis,” *Frontiers in Oncology*, vol. 13, no. 4, pp. 1–12, 2023.

[2] R. Javed and H. Ali, “Deep Learning for Lung Cancer Detection: A Review,” *Springer Nature Computer Science*, vol. 4, no. 7, pp. 1–10, 2024.

[3] A. N. Patel and V. Kumar, “Vision Transformer-Based Effective Model for Early Detection and Classification of Lung Cancer,” *SpringerLink Journal of Imaging and Vision Research*, vol. 9, no. 3, pp. 225–237, 2024.

[4] D. Mannepalli and R. Durgam, “GSC-DViT: A Vision Transformer-Based Deep Learning Model for Lung Cancer Classification,” *Scientific Reports*, *Nature Publishing Group*, vol. 15, no. 2, pp. 1021–1034, 2025.

[5] A. Pal and T. Z. Li, “Hybrid Vision Transformer with Attention Mechanism for Lung Cancer Diagnosis Using CT Images,” *BioMed Central Medical Imaging*, vol. 22, no. 5, pp. 85–96, 2025.