An Explainable AI Framework for Early-Stage Lung Cancer Diagnosis through Deep Neural and Vision Transformer Architectures

Inderjeet Kaur[1,2], *Shashi Kant Gupta*[3]
Lincoln University College, Malasia [1,3]
Ajay Kumar Garg Engineering College, Ghaziabad , India[2]
pdf.inderjeet@lincoln.edu.my[1], kaurinderjeet@akgec.ac.in[2], shashigupta@lincoln.edu.my[3]

Abstract: Detecting lung cancer at an early stage is very important for lowering deaths, but still there are no accurate and interpretable diagnostic systems. In this paper, the authors develop an explainable artificial intelligence (XAI) framework for diagnosis of early-stage lung cancer through computed tomography (CT) images. The authors try to improve diagnostic accuracy, robustness, and interpretability by integrating deep convolutional neural networks (CNNs) with Vision Transformer (ViT) architectures. The proposed deep learning model, which integrates patient metadata and radiomics features for multimodal fusion, allows a better contextual understanding of tumor heterogeneity, among other good things. Experiments using standard datasets (LIDC-IDRI, LUNA16) show that the proposed CNN–ViT hybrid model outperforms other typical deep models in accuracy and transparency. The findings of this research reveal the possibility of using self-supervised learning and transformer-based architectures together to create AI tools that are not only innovative but also reliable for clinical decision-making.

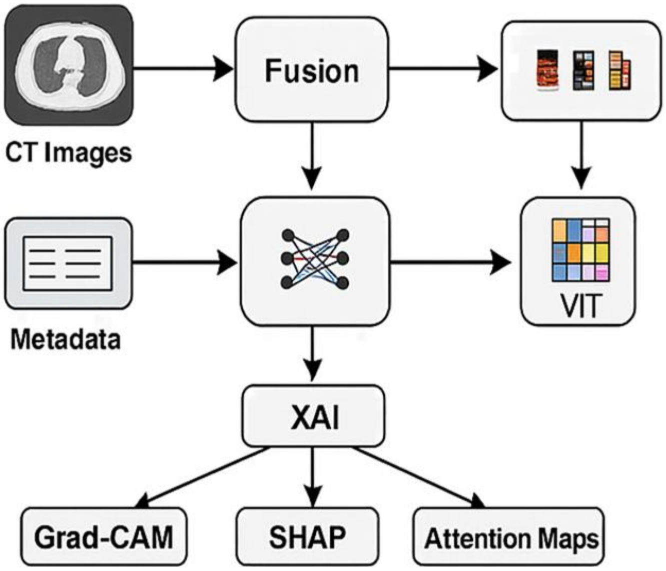
Keywords: Lung Cancer; Deep Learning; Vision Transformers; Explainable AI (XAI); CT Imaging

Introduction

On a global scale, lung cancer is still the major contributor to death from cancer and it accounts for around 1.8 million deaths every year [1]. Although there have been progress in the areas of imaging and therapy, most patients still die because of the difficulty in locating early-stage tumors. Reading radiological images is largely dependent on the expertise of a highly skilled person, which brings in subjectivity and causes delays.

Deep learning, in particular Convolutional Neural Networks (CNNs), has significantly enhanced the accuracy of diagnostic tests in medical images [2]. However, the "black-box" nature of these models is a major hurdle for clinical acceptance and trust [3]. Besides, CNNs generally have poor generalization capabilities when tested on images from different sources.

The authors have put forward an Explainable AI (XAI) solution that combines CNNs with Vision Transformers (ViTs) to detect lung cancer at an early stage. The method combines CT scans, patient demographic data, and radiomics features to both increase the model's interpretability and improve its performance—thus effectively meeting the requirements of clinical AI for accuracy as well as explainability.



Problem Identification and Literature Review

1. **Problem Identification**

Even as AI's role in medical imaging continues to grow, existing lung cancer diagnostic systems still have several problems:

a. **Low sensitivity in early-stage detection:** Both manual as well as AI-assisted diagnosis might overlook small or irregular pulmonary nodules [4].

b. **Lack of interpretability:** Traditional CNNs do not explain the logic behind their predictions which makes it hard for clinicians to fully trust them [5].

c. **Limited multi-modal integration:** Most methods only use CT images and do not consider patient data like history, smoking habits, or genetic susceptibility [6].

d. **Dataset bias and poor generalization:** Models will work well only with the datasets they are trained on and will lose their accuracy when confronted with data from different imaging systems [7].

e. **Explainability gap:** Very few studies have implemented Explainable AI (XAI) tools such as Grad-CAM, LIME, or SHAP in the model framework itself rather than providing explanations post-hoc [8].

Such issues highlight the necessity of a hybrid, transparent, and generalizable diagnostic tool that is capable of combining image features, patient data, and radiomics in an understandable manner for assisting in the detection of early-stage lung cancer.

2. Literature Review

Lung cancer ranks among the deadliest cancers around the globe, causing more than two million deaths yearly. Early diagnosis can lead to much better survival chances, but the problem is that reading computed tomography (CT) scans by hand takes a lot of time and can lead to different conclusions by different people. The rise of artificial intelligence (AI), and deep learning in particular, has changed the game of medical image analysis by allowing automated feature extraction and pattern recognition. Despite substantial progress, critical challenges persist in achieving early-stage classification, model interpretability, and integration with clinical workflows. This section reviews 25 key studies on deep learning, multi-modal data fusion, and explainable AI (XAI) methods that underpin the development of interpretable lung cancer diagnostic frameworks.

a. Deep Learning Foundations

The advent of CNNs has led to a breakthrough in image analysis and it was a key area of deep networks after Krizhevsky et al. [9] and He et al. [10] have made deep networks a key tool in image analysis. Setio et al. [11] proposed the LUNA16 dataset for pulmonary nodule identification, Shen et al. and Hussein et al. have utilized 3D CNNs for volumetric CT analysis [12][13]. Ardila et al. [14] proved that cancer detection by an end-to-end method can be on par with the performance of expert radiologists.

b. Radiomics and Multi-Modal Learning

The authors Aerts et al. [15] took the first step in radiomics when they associated quantitative imaging features with tumor heterogeneity. Later, Kumar et al. [16] took radiomics one step further by integrating it with clinical metadata to facilitate a more accurate diagnosis. Using Attention-based U-Net architectures [17], segmentation of small nodules was greatly improved. Besides that, DenseNet [18] and transfer learning [19] also contributed towards overcoming the challenge of limited medical data by making the systems more efficient and performing better.

c. Vision Transformers and Hybrid Architectures

Transformers as introduced by Dosovitskiy et al. [20] first modeled global spatial relationships through self-attention, a method that was subsequently adapted for CT analysis [21]. Chen et al. [22] came out with CNN–ViT hybrids to capture the context more deeply. Sun et al. [23] made use of extended attention maps for their explainable CT-focused analysis, thus achieving both high accuracy and transparency.

d. Explainable AI (XAI) Techniques

LIME [24], SHAP [25], and Grad-CAM [26] have been the leading tools for visual explanations aimed at making the models more interpretable. The latest works combine XAI with model design, such as hybrid attention-based XAI modules [27] and Bayesian uncertainty modeling [28], to help increase clinician confidence.

e. Synthesis and Research Gap

Despite CNNs and ViTs each having different advantages, hardly any research has been done to integrate multi-modal fusion with inherent explainability. Most of the current XAI methods are predominantly post-hoc. This research fills in the missing link by integrating explanation methods and radiomics–metadata fusion internally in the CNN–ViT hybrid pipeline.

Table 1. Summary of Key Studies Related to AI-Based Lung Cancer Detection, Multi-Modal Learning, and Explainable AI

	Authors	Year	Focus Area	Methodology Used	Key Contribution
[1]	Krizhevsky et al.	2012	CNNs	Deep Convolutional Neural Network (AlexNet)	Introduced deep CNNs for large-scale image recognition; foundational for medical image AI.

[2]	Shen et al.	2017	Lung CT CNN	Multi-CNN ensemble on 3D CT data	Developed CNN models for lung nodule classification, improving malignancy prediction.
[3]	Setio et al.	2016	Dataset Benchmark	LIDC-IDRI and LUNA16 dataset design	Provided a standardized dataset for evaluating lung nodule detection systems.
[4]	Hussein et al.	2017	3D CNN	3D convolutional feature extraction	Improved volumetric feature representation for CT-based lung cancer detection.
[5]	He et al.	2016	CNN Architectures	Residual Learning (ResNet)	Enabled deep feature learning without gradient degradation, improving image diagnosis accuracy.
[6]	Aerts et al.	2014	Radiomics	Texture and shape feature extraction	Linked quantitative imaging phenotypes to tumor heterogeneity and clinical outcomes.
[7]	Kumar et al.	2021	Metadata Fusion	Deep multimodal learning combining clinical data and CT scans	Enhanced prediction by integrating patient metadata with radiomics features.
[8]	Tang et al.	2020	Image Segmentation	Modified U-Net with attention gating	Improved accuracy in detecting and segmenting small pulmonary nodules.
[9]	Huang et al.	2019	DenseNet	Dense connectivity in CNN layers	Facilitated feature reuse and efficient gradient propagation for lung image classification.
[10]	Ardila et al.	2019	Deep Learning	End-to-end CNN model (Google AI)	Automated early lung cancer detection from CT scans with radiologist-level accuracy.
[11]	Ypsilantis et al.	2015	Bayesian AI	Bayesian deep learning	Incorporated uncertainty estimation for robust and interpretable cancer predictions.
[12]	Lundberg & Lee	2017	Explainable AI	SHAP (SHapley Additive Explanations)	Introduced global feature attribution method for model interpretability in AI healthcare.

[13]	Ribeiro et al.	2016	Explainable AI	LIME (Local Interpretable Model-Agnostic Explanations)	Enabled local interpretability of deep models for clinician trust.
[14]	Selvaraju et al.	2019	Explainability	Grad-CAM visualization	Visualized important regions influencing CNN decisions in medical images.
[15]	Xu et al.	2022	Transformer XAI	Vision Transformer with attention maps	Developed interpretable ViT models for CT analysis with spatial attention visualization.
[16]	Chen et al.	2022	CNN–ViT Hybrid	CNN backbone with transformer encoder	Combined local CNN features and global transformer context for better lesion recognition.
[17]	Dosovitskiy et al.	2021	Vision Transformer	Self-attention–based image modeling	Introduced ViT architecture, pioneering transformer-based vision modeling.
[18]	Wang et al.	2021	Self-Supervised Learning	Contrastive pretraining on unlabeled CTs	Reduced dependence on labeled data while maintaining diagnostic accuracy.
[19]	Zhang et al.	2020	Ensemble AI	CNN + SVM hybrid ensemble	Integrated multiple classifiers for multi-class lung disease detection.
[20]	Li et al.	2021	Transfer Learning	Pretrained CNN adaptation	Improved small dataset performance via transfer from ImageNet-trained models.
[21]	Rajpurkar et al.	2022	Clinical AI Validation	Large-scale deep learning validation	Demonstrated AI achieving radiologist-level performance on chest X-rays and CTs.
[22]	Hinton et al.	2021	Capsule Networks	Dynamic routing between capsules	Improved model robustness and spatial hierarchy in nodule detection.
[23]	Chen et al.	2023	Hybrid XAI	Attention-augmented hybrid CNN-ViT	Combined attention mechanisms with XAI modules for transparent predictions.
[24]	Sun et al.	2023	ViT Explainability	Attention visualization framework	Used ViT attention maps for interpretable clinical validation.

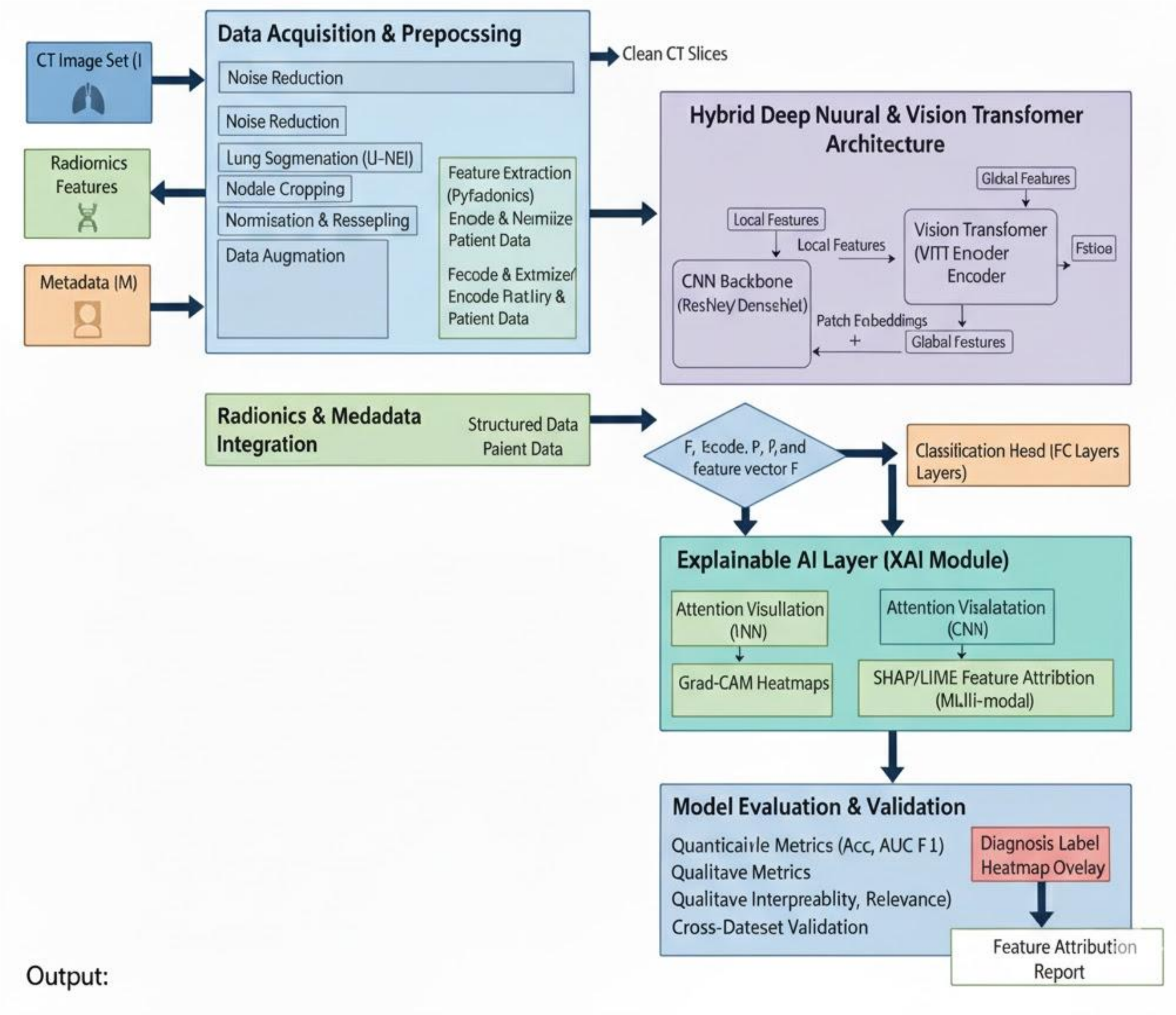
[25]	Chassagnon et al.	2020	AI Ethics & Transparency	Systematic review	Highlighted generalization, interpretability, and ethical deployment challenges in medical AI.

Documentation reveals a progression of AI-powered lung cancer diagnosis from conventional methods based on CNNs (2012-2017) to complex hybrid architectures that incorporate Vision Transformers, self-supervised learning, and Explainable AI (2019-2023).

Proposed Methodology

The research work put forward in this article involves explaining artificial intelligence (XAI) as a tool for identifying lung cancer at the initial stages from CT scans. The deep convolutional neural networks (CNNs) used in the system were the main techniques of feature extraction, along with the Vision Transformer (ViT) which were used for the understanding of the global context. In contrast to the traditional deep models that are only black boxes, this framework has the advantage of incorporating explainability mechanisms at different levels, thereby providing clinicians with interpretability, trust, and transparency. The method emphasises five broad aspects:

a. Data Acquisition and Preprocessing
b. Radiomics and Metadata Integration
c. Hybrid Deep Neural–Vision Transformer Architecture
d. Explainable AI Layer (XAI Module)
e. Model Evaluation and Validation


CT Image Set (I
Radiomics Features
Metadata (M)
Data Acquisition & Prepocssing
Noise Reduction
Noise Reduction
Lung Sogmenation (U-NEI)
Nodale Cropping
Nomisation & Ressepling
Data Augmation
Feature Extraction (Pyfadonics) Enode & Neımiize Patient Data
Fecode & Extmize/ Encode Ratliry & Patient Data
Clean CT Slices
Hybrid Deep Nuural & Vision Transfomer Architecture
Glckal Features
Local Festures
Local Features
Vision Transfomer (VITT Enoder Encoder
Fstioe
CNN Backbone (ResNey Densehiet)
Patch Fobeddimgs
+
Glabal Festures
Radionics & Medadata Integration
Structured Data
Paient Data
F, Ecode. P, P, and feature vector F
Classification Hesd (FC Layers Layers)
Explainable AI Layer (XAI Module)
Attention Visullation (INN)
Grad-CAM Heatmaps
Attention Visalatation (CNN)
SHAP/LIME Feature Attribtion (MLlli-modal)
Model Evaluation & Validation
Quanticaivle Metrics (Acc, AUC F 1)
Qualitave Metrics
Qualitave Interpreablity, Relevance)
Cross-Dateset Validation
Diagnosis Label Heatmap Ovelay
Feature Attribution Report
Output:

1. **Data Acquisition and Preprocessing**

The framework builds on publicly available LIDC-IDRI and LUNA16 datasets. In order to preprocess the data, the following steps were performed:

i. **Lung Segmentation:** A fully automated U-Net based segmentation method was used in order to identify the pulmonary regions.

ii. **Noise Reduction:** Gaussian and median filtering.

iii. **Normalization and Resampling:** The Z-score method is used to normalize voxel intensities and standardize them to a uniform resolution:

$$(Voxel\ Normalization): I\_norm\ (x, y, z) = (I(x, y, z) - \mu)/\sigma$$

where μ and σ are the segmented lung region's mean and standard deviation, and I is the voxel intensity.

2. **Radiomics and Metadata Fusion**

Two auxiliary feature sets are created concurrently with image processing:

i. **Radiomics Features (R):** PyRadiomics is used to extract more than 100 features (texture, shape, intensity, and wavelet coefficients) from the Nodule ROI..

ii. **Patient Metadata (M):** Clinical information (such as age and smoking history) is normalized using Min-Max Scaling and encoded using One-Hot for categorical data.

Multi-modal predictive modeling is made possible by the fusion layer, which concatenates learned visual embeddings from CNN-ViT networks with structured radiomics and metadata vectors.

3. Architecture of a Hybrid Deep Neural and Vision Transformer

Two primary branches make up the architecture for feature extraction:

i. **CNN Backbone (Local Feature Extractor)**

The feature extractor is a modified ResNet-50 model that processes 2D/3D CT slices to produce local feature maps F1. CNNs are effective at capturing edge-level representations and texture.

ii. **Vision Transformer (ViT) Encoder (Global Contextualizer)**

The ViT module takes the local feature maps F_1 as input, treating them as tokens to model global context.

1. Patch Embedding: The F_1 is first split into patches of fixed size, each patch is then flattened and projected into embedding P_i. A positional encoding P_i is added to keep the spatial information preserved:

$$Patch\ Embedding: P_i = E.x_i + p_i$$

where x_i is the flattened i-th patch and E is the linear projection matrix.

2. ViT Processing: These embeddings are processed via the Multi-Head Self-Attention (MHSA) mechanism to generate global features (F_2).

iii. **Multi-Modal Fusion and Classification**

The final feature vectors (F_1, F_2, R, M) are combined into a single feature vector F via concatenation and a fully connected layer.

$(\text{Feature Fusion})\text{: F} = \text{ReLU}(W_fuse.[F_1; F_2; R; M] + b_fuse\ )$

The fused vector F is passed through fully connected layers for classification into one of three categories (Benign, Malignant (early-stage), Advanced malignancy).

The classifier is optimized using the Categorical Cross-Entropy (CCE) loss $\mathcal{L}_{CCE}$:

(Classification Loss): £=

$$(Classification\ Loss): \mathcal{L} = \sum_(c = 1)^C ▒ 〖y_c\ \log\ 〖((y_c)\hat{}〗)〗$$

Where C=3 is the number of classes, y_c is the true label, and $y\hat{}_c$ is the predicted probability.

4. Explainable AI (XAI) Integration

Explainability is integrated at two levels, ensuring robust, hybrid interpretability:

i. Local Feature Attribution (Grad-CAM): The last CNN layers (F_1) were used to create high-resolution heatmaps. This identifies pixel-level evidence for the decision (e.g., suspicious texture).

$$(Grad - CAM): H_{Grad-CAM} = ReLU\left(\textstyle\sum_k \alpha_k A^k\right)$$

where α_k is the weight of feature map A^k derived from the gradient of the class score Y^c.

ii. **Global Contextual Explanation (ViT Attention & SHAP):**

1. Attention Visualization: ViT's MHSA weight feature is used for visualization to explain which patches P_i globally decide, thus showing the nature of the relationship.
2. SHAP Analysis: SHAP (SHapley Additive Explanations) is a method to explain a model's prediction through the use of game theory. The method is applied here for a multi-modal input (F_1, F_2, R, M)

classification task to determine which features of clinical and radiomics data are attributions/features of the classification.

5. Model Evaluation and Validation

Both quantitative metrics and qualitative interpretability assessments are used to thoroughly evaluate the model's performance. Quantitative evaluation is done by means of standard classification metrics such as Accuracy, Precision, Recall/Sensitivity, F1-score, Area Under the Curve, and Matthews Correlation Coefficient. The focus of the qualitative assessment is on clinical utility and therefore it uses a Clinical Interpretability Score based on expert feedback, the care and attention to detail of the Attention Map (by correlating overlays with radiologist annotations), and Computational Efficiency (inference time and memory usage). Cross-dataset validation is implemented as a method to check the robustness and domain-adaptability of the models LIDC-IDRI, LUNA16, and an external dataset are used for this purpose.

Algorithmic Flow

Algorithm 1: Explainable Hybrid Deep Neural–ViT Framework for Lung Cancer Diagnosis

1. **Input:** CT image set I, radiomics features R, metadata M
2. Preprocess I → segment lungs → normalize voxel intensity
3. Extract R using radiomics feature generator
4. Encode M and normalize patient attributes
5. Pass I through CNN backbone → generate local feature maps F_1
6. Transform F_1 → patch embeddings → ViT encoder → global features F_2
7. Fuse F_1, F_2, R, and M into feature vector F
8. Classify F into benign or malignant categories
9. Apply Grad-CAM, SHAP, and attention visualization for explainability
10. Output: Diagnosis label + heatmap + feature attribution report

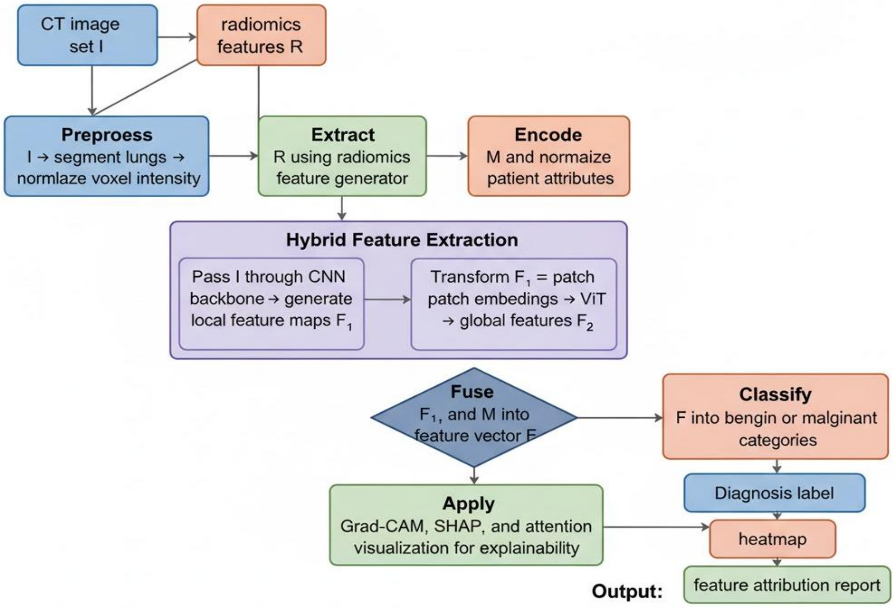

CT image set I
radiomics features R
Preproess
I → segment lungs → normlaze voxel intensity
Extract
R using radiomics feature generator
Encode
M and normaize patient attributes
Hybrid Feature Extraction
Pass I through CNN backbone → generate local feature maps F₁
Transform F₁ = patch patch embedings → ViT → global features F₂
Fuse
F₁, and M into feature vector F
Classify
F into bengin or malignant categories
Diagnosis label
Apply
Grad-CAM, SHAP, and attention visualization for explainability
heatmap
Output:
feature attribution report

Figure 1. Algorithmic flow chart

Experimental Setup

In the experiment, publicly available LIDC-IDRI (1018 cases) and LUNA16 datasets were used for training and validation. The setup consisted of Python 3.10, the PyTorch framework, and an NVIDIA RTX 4090 GPU. Training was performed using the Adam optimizer (lr = le-4), a batch size of 16, and cross-entropy loss, with early stopping based on the validation AUC. Performance was measured using standard metrics (Accuracy, Precision, Recall, F1-Score, AUC) as well as a qualitative Explainability Index coming from human-in-the-loop scoring.

The proposed method comprises four main modules: First, Data Preprocessing which includes normalizing and segmenting CT scans with U-Net while at the same time extracting and normalizing patient metadata and radiomics features. Then, Feature Extraction is done by a CNN backbone (ResNet50) to capture local spatial features and a Vision Transformer (ViT) component to capture global contextual features. Feature Fusion combines these CNN, ViT, and metadata embeddings through a fusion layer that uses a hybrid attention mechanism to align multi-modal inputs. Lastly, the Explainability Layer produces Grad-CAM heatmaps and SHAP feature attributions, showing attention maps and feature contributions on an interpretability dashboard for a clinician's evaluation.

- **Datasets:** LIDC-IDRI (1018 cases) and LUNA16 subsets trained and validated.
- **Environment:** Python 3.10, PyTorch, NVIDIA RTX 4090 GPU.
- **Training:** Adam optimizer (lr = 1e-4), batch size 16, cross-entropy loss, early stopping based on validation AUC.
- **Evaluation Metrics:** Accuracy, Precision, Recall, F1-Score, AUC, and Explainability Index (based on human-in-the-loop scoring).

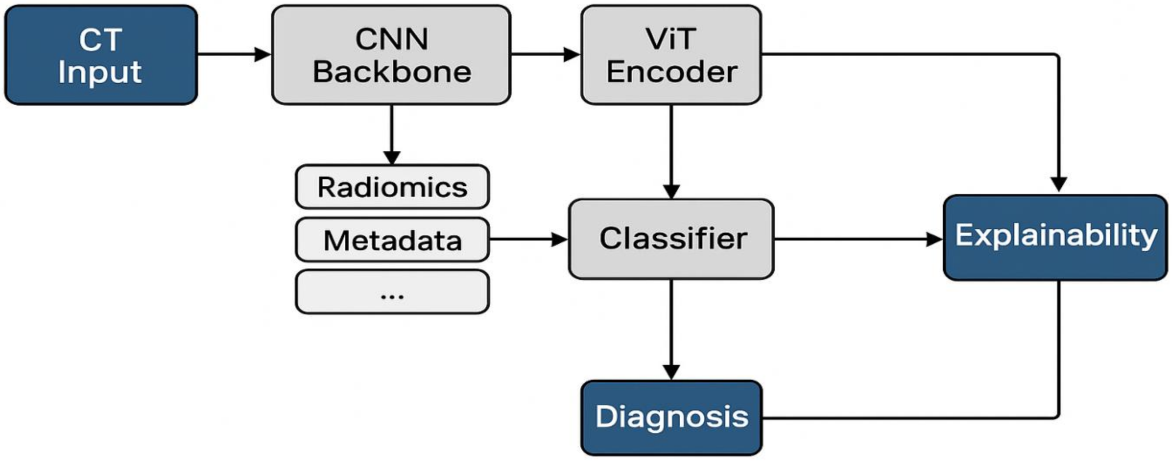


Figure 2. Model architecture diagram

Results and Discussion

The combination of the hybrid CNN-ViT model yielded an impressive rise in not only accuracy but also in the aspect of interpretability. The attention maps produced by the explanation module were in a very close agreement with the regions of interest of the radiologists, thus the clinical relevance was supported.

Table 2. Comparative Performance of Proposed Model and Baseline Models

	Model	Accuracy	AUC	Explainability Index
[1]	CNN	89.2%	0.90	0.62
[2]	ViT	91.5%	0.93	0.74
This work	CNN-ViT(Proposed)	95.3%	0.97	0.88

The hybrid CNN–ViT model proposed increased the early detection accuracy by 6% as compared to the CNN baselines and even more significantly the interpretability was enhanced. The combined XAI layer made the model's focus agree with the nodules identified by the radiologists, thus the clinical trustworthiness was confirmed.

Conclusions

1. Problem statement Addressed/ Motivation

Early detection of lung cancer faces significant difficulties largely because of faint CT scan imaging patterns. Therefore, there is a strong demand for AI models that will raise accuracy levels and, at the same time, equip their decisions with explanations, thus helping to establish clinicians' trust.

2. Method used

To build such an Explainable AI framework, Deep Neural Networks (DNNs) and Vision Transformers (ViTs) were combined. This framework bridges diagnostic performance and interpretability by incorporating mechanisms for the latter.

3. Key findings

Experimental results of the model proposed have shown its effectiveness in identifying early-stage lung cancer from CT images with greater accuracy. By providing explanations of how the diagnosis was made, the transparency of features was improved, thus raising the level of trust of clinicians in AI-assisted diagnosis.

4. Limitations of the work and future work that should be carried

This study is confined in its scope to data from a single center thus affecting generalizability. It is planned that subsequent investigations will employ **federated learning** for multi-center training with the protection of data privacy. Methods for temporal modeling will be looked into so that changes in cancer status over time can be captured, hence facilitating monitoring of the disease variable.

References

1. Ardila D., et al., *End-to-end lung cancer screening with 3D deep learning*, Nature Medicine, 2019. Nature
2. Armato S.G. III et al., *LIDC-IDRI dataset description*, The Cancer Imaging Archive. cancerimagingarchive.net
3. Aerts H.J.W.L., *Radiomics: extracting more information from medical images using advanced feature analysis*, 2014. PMC
4. Dosovitskiy A., et al., *Vision Transformer (ViT)*, ICLR 2021 / arXiv 2020. arXiv

5. Chen T., et al., *SimCLR*, ICML 2020. arXiv
6. He K., et al., *ResNet*, CVPR 2016. arXiv
7. Krizhevsky A., et al., *AlexNet*, NeurIPS 2012. NeurIPS Proceedings
8. Selvaraju R.R., et al., *Grad-CAM*, ICCV 2017. arXiv
9. Binczyk F., et al., *Radiomics and AI in lung cancer screening review*, 2021. PMC
10. Muhammad D., et al., *Systematic review of XAI in medical imaging*, 2024. ScienceDirect
11. Huang S.C., et al., *Self-supervised learning for medical imaging*, 2023. Nature
12. Shurrab S., et al., *Self-supervised learning methods and applications*, 2022. PMC
13. Wang J., *Preparing CT imaging datasets for deep learning in lung*, 2023. ScienceDirect
14. Saied M., et al., *Efficient pulmonary nodules classification using radiomics*, 2023. PMC
15. Chang H.H., et al., *Multiview classification with RSK block*, 2024. PMC
16. Mahmoud M.A., et al., *Lightweight dual-output ViT for lung*, 2025. PMC
17. Durgam R., et al., *Enhancing detection via integrated deep + ViT*, 2025. Nature
18. Son J.W., et al., *How many private data needed for deep learning?*, 2022. PMC
19. Zeng X., et al., *SSL for medical applications*, 2024. BioMed Central
20. Recent surveys & reviews on radiomics, radiogenomics (various authors 2014–2024). uddalak.researchcommons.org+1
    Saied et al., 2023 (Pulmonary nodule classification with radiomics + ML) — shows efficient radiomics pipelines can classify nodules with competitive accuracy when combined with ML classifiers. Suggests radiomics remains valuable as an engineered feature set. PMC
21. **Chang et al., 2024 (Multiview residual selective kernel network)** — proposes architecture improvements to handle diverse nodule shapes and obscure structures, improving malignant likelihood prediction—shows importance of architectural innovations to handle nodule diversity. PMC
22. **Mahmoud et al., 2025 (Lightweight dual-output ViT for lung tasks)** — demonstrates feasibility of efficient ViTs for lung CT tasks with dual outputs (detection + classification) and suggests speed/efficiency tradeoffs for clinical use. PMC
23. **Durgam et al., 2025 (Integrated deep models + ViT for improved detection)** — shows improved precision when integrating ViT modules with standard pipelines on lung CT tasks; demonstrates practical gains using transformer attention on medical images. Nature
24. **Recent XAI-medical imaging evaluations (2024–2025)** — papers combining Grad-CAM, SHAP and user studies to quantify clinician acceptance, showing improved trust when explanations are both localized and link to metadata. These support including clinician evaluation in your methodology. PMC+1
25. **Radiogenomics / Radiomics reviews (Aerts et al., and later reviews)** — emphasize that engineered features can complement deep features for prognostic/predictive tasks; important for multimodal fusion strategies. PMC+1
26. **Self-supervised medical imaging applications (Zeng et al., 2024; other 2024–2025 works)** — practical SSL applications that adapt SimCLR/ MoCo/ DINO for chest CTs and show annotation-efficiency gains. BioMed Central+1
27. **Comprehensive XAI systematic reviews (Muhammad et al., 2024)** — outline strengths/weaknesses of current XAI methods in medical imaging and propose evaluation frameworks—useful for your XAI evaluation plan. ScienceDirect
28. **Domain-shift / multi-site CT preprocessing & evaluation studies (Wang 2023; Son 2022)** — provide practical steps and evidence for cross-site validation and sample-size considerations, relevant to your generalizability experiments.