Towards Clinically Deployable Lung Nodule Classification Using a Hybrid Multi-Scale 3D CNN–Radiomics Framework with Explainable Artificial Intelligence

Dr. Inderjeet Kaur[1] | Dr. Shashi Kant Gupta[2]

[1]Postdoctoral Researcher, Lincoln University College, Malaysia

[1]Professor, Ajay Kumar GArg Engineering College, Ghaziabad

[2]Adjunct Professor, Department of Computer Science and Engineering, Lincoln University College, Malaysia

Abstract

Lung cancer is one of the leading causes of cancer-related deaths worldwide, making early diagnosis essential for improving patient outcomes. Although Low-Dose Computed Tomography (LDCT) supports early detection, manual interpretation of volumetric CT scans is time-consuming and subject to considerable inter-observer variability. Existing deep learning approaches also face challenges related to model interpretability, limited use of radiomic information, varying nodule sizes, and computational complexity. This paper presents a hybrid framework that combines multi-scale 3D Convolutional Neural Networks (3D CNNs), handcrafted radiomic features, a Convolutional Block Attention Module (CBAM), and 3D Grad-CAM for explainable lung nodule classification. The model was evaluated on 2,661 annotated nodules from 1,018 CT scans in the publicly available LIDC-IDRI dataset using a strict patient-level data split. It achieved 93.8% accuracy, 91.4% sensitivity, 95.2% specificity, and an AUC of 0.9489, while requiring only ~450K parameters and an average inference time of 58 ms. Quantitative evaluation of the generated 3D Grad-CAM heatmaps produced an Intersection-over-Union (IoU) of 0.77 with expert annotations, indicating good agreement between model explanations and clinically relevant regions. These results demonstrate that combining multi-scale feature learning with radiomic descriptors can provide accurate and computationally efficient lung nodule classification.

Keywords-Lung Cancer, Lung Nodule Classification, Explainable Artificial Intelligence, Deep Learning, Multi-scale 3D CNN, Radiomics, CBAM Attention, Grad-CAM, Medical Image Analysis, Computer-Aided Diagnosis.

Introduction

Lung cancer continues to be a major public health concern, accounting for more than **2.8 million new cases** and nearly **1.9 million deaths** each year. Detecting the disease at an early stage has a substantial impact on patient outcomes. When identified during **Stage I**, the five-year survival rate exceeds **90%**, whereas advanced metastatic disease is associated with survival rates below **10%**. Screening using **Low-Dose Computed Tomography (LDCT)** has therefore become an important tool for early diagnosis. However, each CT examination contains hundreds of slices, making manual interpretation both time-consuming and demanding. In addition, subtle lesions such as ground-glass opacities, together with overlapping anatomical structures, contribute to inter-observer variability that has been reported between **10% and 40%**.

Deep learning has considerably improved the performance of **Computer-Aided Diagnosis (CAD)** systems for lung nodule analysis. In particular, **3D Convolutional Neural Networks (3D CNNs)** are well suited to volumetric CT data because they learn spatial information directly from three-dimensional image volumes. Despite these advances, several factors continue to limit their routine clinical use.

One limitation is that many existing models rely on a **single input patch size**. While effective for certain nodules, fixed receptive fields often fail to represent both fine internal textures and the surrounding anatomical context when lesion size varies.

Another limitation is the limited integration of **radiomic features**. Quantitative descriptors describing texture, shape, and intensity contain clinically relevant information but are frequently excluded from deep learning pipelines or processed independently of learned image features.

Model interpretability also remains an important concern. Although visualization methods such as **Grad-CAM** are commonly used, many studies rely on qualitative inspection of heatmaps without comparing them quantitatively with expert annotations. This makes it difficult to assess whether highlighted regions correspond to clinically meaningful structures.

A further issue relates to **dataset partitioning**. Several published studies divide data at the nodule level instead of the patient level, allowing scans from the same patient to appear in both training and testing sets. Such data leakage can lead to overly optimistic performance estimates that may not reflect real clinical performance.

Finally, model complexity continues to be a practical consideration. Modern Vision Transformer (ViT) architectures often contain more than **100 million parameters**, making deployment challenging on standard clinical workstations without dedicated GPU infrastructure.

To address these limitations, this work presents a hybrid framework that combines **multi-scale 3D CNN feature extraction**, **radiomic feature fusion**, **CBAM attention**, and **quantitatively evaluated 3D Grad-CAM** within a compact architecture containing approximately **450,000 parameters**. The framework is designed to balance classification performance, computational efficiency, and model interpretability while following a strict **patient-level evaluation protocol**.

Related work

Recent research on AI-based lung nodule classification has explored several directions, including **3D Convolutional Neural Networks (3D CNNs)**, **Vision Transformers (ViTs)**, **radiomics-assisted models**, **Explainable Artificial Intelligence (XAI)**, and lightweight deep learning architectures. Each approach addresses different aspects of lung nodule analysis, ranging from feature representation and classification performance to model interpretability and computational efficiency. Table I summarizes representative studies published between **2022 and 2026**, together with their main contributions and reported limitations.

Table I. Representative Studies on AI-Based Lung Nodule Classification (2022–2026)

Year	Methodology / Reference	Key Strengths	Identified Limitations
2022	Self-Supervised 3D Vision Transformer	Learns rich global contextual features	High computational complexity; detection-oriented

2023	Comprehensive ViT Thoracic Review	Summarizes multi-modal ViT advances	Highlights substantial translation barriers
2024	TiCNet (CNN–Transformer Hybrid)	Effective multi-scale feature fusion	Focuses on detection rather than malignancy scoring
2025	GSC-DVIT (Dual-Attention ViT)	Reduced parameter footprint	Lacks quantitative XAI validation
2025	CNN + SHAP + CatBoost Pipeline	Incorporates feature-level XAI	No quantitative spatial localization metrics
2025	CLAHE + CNN–Radiomics Pipeline	Contrast & radiomic feature enhancement	Lacks robust multi-scale contextual learning
2025	Hybrid Vision Transformer Pipeline	Combines local and global representations	High computational latency (>200 ms)
2026	Multi-View Radiomics Transformer	Multi-view spatial feature integration	Minimal evaluation in clinical workflows
2026	Multi-Modal Explainable AI	Enhanced visual interpretability	General focus on overall pulmonary disease
Prop.	**Multi-Scale 3D CNN + Radiomics + CBAM + 3D Grad-CAM**	**High accuracy (93.8%), 450K params, 58 ms runtime, quantitative IoU = 0.77**	**Requires prospective multi-center trial validation**

The comparison shows that transformer-based models are effective at learning global contextual information but usually require large computational resources. Hybrid CNN–Transformer architectures improve feature representation, although many have been designed primarily for nodule detection rather than malignancy classification. Other studies have incorporated radiomic features or explainability techniques; however, quantitative validation of model explanations remains limited. Lightweight architectures have reduced computational requirements, but balancing efficiency, interpretability, and diagnostic performance is still an open research problem.

Key Contribution

1. A review of recent literature identifies several challenges that continue to limit the practical adoption of AI-based lung nodule classification systems. These include limited multi-scale feature representation, inadequate integration of radiomic features with deep learning, insufficient quantitative evaluation of explainability methods, inconsistent validation protocols, and the high computational cost of modern transformer-based architectures.
2. To address these issues, this work introduces a hybrid framework that combines multi-scale volumetric feature learning, handcrafted radiomic descriptors, attention-based feature

refinement, and quantitative explainability within a computationally efficient architecture. The principal contributions of this study are summarized below.

3. A **multi-scale 3D CNN architecture** is developed using three parallel input volumes (**32^3, 48^3, and 64^3 voxels**) to capture fine texture, lesion morphology, and surrounding anatomical context across nodules of different sizes.
4. **Handcrafted radiomic descriptors** extracted using PyRadiomics are integrated with deep features through a unified feature fusion strategy, allowing quantitative imaging biomarkers to complement learned representations.
5. A **Convolutional Block Attention Module (CBAM)** is incorporated after feature fusion to enhance diagnostically relevant feature channels while suppressing less informative responses.
6. Model interpretability is improved using **3D Grad-CAM**, and the generated attention maps are quantitatively evaluated against expert radiologist annotations using the **Intersection-over-Union (IoU)** metric.
7. The framework follows a **strict patient-level data partitioning strategy**, preventing patient overlap between training and testing datasets and reducing the possibility of data leakage.
8. The proposed architecture achieves **93.8% classification accuracy**, **91.4% sensitivity**, **95.2% specificity**, and an **AUC of 0.9489**, while requiring only **approximately 450K trainable parameters** and an average inference time of **58 ms**.

Table II summarizes the identified research challenges together with the corresponding design choices incorporated into the proposed framework.

Table II. Research Gap Analysis & Proposed Architectural Solutions

Research Challenge	Existing Approaches	Identified Limitation	Proposed Solution
Multi-Scale Feature Extraction	Single-scale CNN / fixed receptive field	Fails to capture heterogeneous nodule sizes	Three parallel 3D CNN branches (32^3, 48^3, 64^3 voxels) capture texture, morphology & context
Feature Fusion Strategy	Deep features or radiomics used independently	Sub-optimal feature synergy	Hybrid concatenation of 384D deep features with 28 PyRadiomics descriptors
XAI Rigor	Qualitative visual heatmaps	No objective localization verification	3D Grad-CAM heatmaps quantitatively evaluated via IoU against radiologist annotations
Validation Integrity	Nodule-level dataset partitioning	Data leakage; overly optimistic metrics	Strict patient-level data partitioning ensuring

			zero inter-set patient overlap
Deployment Viability	Heavy Transformer architectures	High latency (>200 ms) and huge parameters (>100M)	Ultra-lightweight ~450K parameter model achieving 58 ms inference per scan

Methodology

The proposed framework follows six sequential processing stages: **(1) Preprocessing and Resampling, (2) Multi-Scale 3D CNN Feature Learning, (3) Radiomic Feature Extraction, (4) Hybrid Feature Fusion with CBAM Attention, (5) Softmax Classification, and (6) Quantitative Evaluation using 3D Grad-CAM.** The complete processing pipeline is illustrated in **Figure 2**.

A. CT Image Preprocessing

The input DICOM CT scans are first converted to **Hounsfield Units (HU)** and windowed between **−1000 HU and +400 HU** to highlight the lung parenchyma. To maintain a consistent spatial resolution across all scans, each volume is resampled to an isotropic voxel spacing of $\mathbf{1 \times 1 \times 1\ mm^3}$. Using the annotated lesion centroids provided in the LIDC-IDRI dataset, three concentric cubic Volumes of Interest (VOIs) are extracted with sizes of $\mathbf{32^3}$, $\mathbf{48^3}$, and $\mathbf{64^3}$ **voxels**. These patches capture complementary information, including fine texture and margins, lesion morphology, and the surrounding anatomical context.

B. Multi-Scale 3D CNN Branches and Radiomics

Each of the three parallel CNN branches uses the same architecture consisting of **Conv3D(16) → Conv3D(32) → MaxPool → Conv3D(64) → Global Average Pooling (GAP) → FC(128)**. The outputs from the three branches are concatenated to form a **384-dimensional deep feature vector (FCNN)**.

In parallel, **PyRadiomics** extracts **28 quantitative radiomic descriptors (FR)** from each nodule. These descriptors include **shape features, first-order statistics, Gray-Level Co-occurrence Matrix (GLCM), Gray-Level Run Length Matrix (GLRLM), and Gray-Level Size Zone Matrix (GLSZM)** features, providing additional quantitative information about lesion characteristics.

C. Hybrid Feature Fusion and CBAM Module

The deep feature vector and handcrafted radiomic descriptors are combined to obtain a **412-dimensional hybrid feature vector (FH = FCNN ⊕ FR)**. This representation is then processed using the **Convolutional Block Attention Module (CBAM)**, which applies channel attention followed by spatial attention. These attention mechanisms emphasize diagnostically important features while reducing the influence of less informative background responses before the final classification stage.

Mathematical Formulation

Let the input CT volume be represented by $\mathbf{X} \in \mathbb{R}^{H \times W \times D}$, from which three volumetric patches $\mathbf{X_{32}}$, $\mathbf{X_{48}}$, and $\mathbf{X_{64}}$ are extracted. The mathematical formulation of the proposed framework is given below.

(1)

(2)

(3)

(4)

(5)

(6)

(7)

The proposed framework was evaluated using the **LIDC-IDRI** dataset, which contains **1,018 thoracic CT scans** and **2,661 annotated lung nodules (≥3 mm)**. Nodules with radiologist consensus scores of **≤2** were considered benign, whereas nodules with scores of **≥4** were labeled malignant. Cases with an intermediate score of **3** were excluded from the experiments.

To prevent data leakage, a **strict patient-level partition** was adopted, allocating **60%** of the patients for training, **15%** for validation, and **25%** for testing.

The network was trained using the **Adam optimizer** with early stopping. Model performance was evaluated using **Accuracy, Sensitivity, Specificity, AUC-ROC, McNemar's statistical test ($p < 0.001$), 10,000 bootstrap confidence interval iterations,** and **IoU-based localization analysis** for the generated Grad-CAM heatmaps.

The proposed framework was compared with representative volumetric deep learning models commonly used for lung nodule classification. The quantitative performance comparison is presented in **Table III**.

Table III. Comparative Diagnostic Performance Analysis on Held-Out Test Set

Model Architecture	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC	Parameters	Inference (ms)
ResNet50 (3D)	89.2	87.3	90.8	0.9380	11 M	95
DenseNet121 (3D)	91.5	89.1	93.4	0.9560	7 M	87
Vision Transformer (ViT)	90.8	88.5	92.7	0.9520	>150 M	>200
Proposed Framework	**93.8**	**91.4**	**95.2**	**0.9489**	**450 K**	**58**

As shown in **Table III**, the proposed framework achieved the highest overall classification accuracy (**93.8%**) while maintaining a balanced sensitivity (**91.4%**) and specificity (**95.2%**). Compared with DenseNet121 and Vision Transformer architectures, the proposed model requires substantially fewer parameters (**450K**) and achieves a lower average inference time (**58 ms**). Statistical analysis using McNemar's test confirmed that the observed performance improvement was significant ($p < 0.001$), and the **95% bootstrap confidence interval** for the AUC ranged from **0.934 to 0.962**.

Ablation Study

To assess the contribution of each component, an ablation study was performed by progressively incorporating multi-scale feature learning, radiomic feature fusion, and CBAM attention into the baseline model.

Table IV. Ablation Study of Component Contributions

Model Configuration	AUC-ROC	Absolute Gain
Baseline Single-Scale 3D CNN (32^3 input)	0.8842	—
+ Multi-Scale Feature Learning (32^3, 48^3, 64^3)	0.9142	+3.00%
+ Radiomics Feature Fusion (28 PyRadiomics descriptors)	0.9342	+2.00%
+ CBAM Attention Module (Proposed Full Pipeline)	**0.9489**	**+1.47%**

The results indicate that **multi-scale feature learning** provides the largest improvement in classification performance, increasing the AUC by **3.00%** over the baseline model. Incorporating **radiomic feature fusion** contributes an additional **2.00%**, while the inclusion of **CBAM attention** further improves the AUC by **1.47%**. Overall, the complete framework achieves a cumulative improvement of **6.47%** compared with the baseline architecture.

Discussions

The experimental results indicate that combining multi-scale feature learning with radiomic descriptors improves lung nodule classification compared with conventional single-scale architectures. Processing three input resolutions enables the network to capture both fine-grained internal texture and broader anatomical context, which is important because pulmonary nodules vary considerably in size and appearance.

The inclusion of handcrafted radiomic features provides additional quantitative information that complements features learned by the deep neural network. This combination contributes to improved classification performance while preserving clinically meaningful imaging characteristics.

The quantitative evaluation of 3D Grad-CAM further strengthens the interpretability of the proposed model. Rather than relying only on visual inspection, the IoU comparison with expert radiologist annotations demonstrates that the highlighted regions correspond closely to clinically relevant areas of the lesion.

Another important observation is the balance achieved between accuracy and computational efficiency. Although transformer-based models often report strong classification performance, they generally require significantly larger parameter counts and longer inference times. In contrast, the proposed framework achieves competitive diagnostic performance using only **approximately 450K parameters** and an average inference time of **58 ms**, making it more suitable for deployment on standard clinical workstations.

While the experimental results are encouraging, further evaluation using larger multi-center datasets and prospective clinical studies would provide additional evidence of the framework's robustness across different imaging protocols and patient populations.

Conclusions

• **Problem addressed / Motivation**

Manual interpretation of volumetric lung CT scans is time-consuming and subject to inter-observer variability. Existing deep learning methods also face challenges related to model interpretability, computational complexity, and reliable evaluation.

• **Method used**

A hybrid framework combining multi-scale 3D CNNs, handcrafted radiomic features, CBAM attention, and quantitatively validated 3D Grad-CAM was developed and evaluated using a strict patient-level partition of the LIDC-IDRI dataset.

• **Key findings**

- The proposed framework achieved **93.8% classification accuracy**, **91.4% sensitivity**, **95.2% specificity**, and an **AUC of 0.9489**.
- The network required only **approximately 450K parameters** and achieved an average inference time of **58 ms**.
- Quantitative evaluation of 3D Grad-CAM produced an **IoU score of 0.77**, indicating good agreement between model explanations and expert annotations.
- Multi-scale feature learning, radiomic feature fusion, and CBAM attention each contributed to improved classification performance.

• **Limitations and Future Work**

- The evaluation was performed using a single public dataset.
- Prospective clinical validation across multiple institutions was not included.
- Additional studies involving larger and more diverse patient populations are needed.
- Future work will investigate multi-center validation, integration of multimodal clinical and genomic information, and deployment within hospital PACS environments for routine clinical evaluation.

References

[1] S. G. Armato III, G. McLennan, L. Bidaut *et al*., "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.
https://doi.org/10.1118/1.3528204

[2] A. A. A. Setio, A. Traverso, T. de Bel *et al*., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Medical Image Analysis*, vol. 42, pp. 1–13, 2017.
https://doi.org/10.1016/j.media.2017.06.015

[3] G. Kang, K. Liu, B. Hou and N. Zhang, "3D multi-view convolutional neural networks for lung nodule classification," *PLoS ONE*, vol. 12, no. 11, Art. no. e0188290, 2017.
https://doi.org/10.1371/journal.pone.0188290

[4] R. Dey, Z. Lu and Y. Hong, "Diagnostic classification of lung nodules using 3D neural networks," *Proc. IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 774–778, 2018.
https://doi.org/10.1109/ISBI.2018.8363693

[5] H. Jung, B. Kim, I. Lee *et al*., "Classification of lung nodules in CT scans using three-dimensional deep convolutional neural networks with a checkpoint ensemble method," *BMC Medical Imaging*, vol. 18, Art. no. 48, 2018.
https://doi.org/10.1186/s12880-018-0286-0
[6] S. Shen, S. X. Han, D. R. Aberle, A. A. T. Bui and W. Hsu, "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification," *Expert Systems with Applications*, vol. 128, pp. 84–95, 2019.
https://doi.org/10.1016/j.eswa.2019.03.031
[7] P. Monkam, S. Qi, H. Ma, W. Gao, Y. Yao and W. Qian, "Detection and classification of pulmonary nodules using convolutional neural networks: A survey," *IEEE Access*, vol. 7, pp. 78075–78091, 2019.
https://doi.org/10.1109/ACCESS.2019.2921977
[8] K. Mehta, A. Jain, J. Mangalagiri, S. Menon, P. Nguyen and D. R. Chapman, "Lung nodule classification using biomarkers, volumetric radiomics and 3D CNNs," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 10, pp. 1813–1824, 2021.
https://doi.org/10.1007/s11548-021-02413-9
[9] Y. Gu, J. Chi, J. Liu *et al*., "A survey of computer-aided diagnosis of lung nodules from CT scans using deep learning," *Computers in Biology and Medicine*, vol. 137, Art. no. 104806, 2021.
https://doi.org/10.1016/j.compbiomed.2021.104806
[10] S. Tomassini, N. Falcionelli, P. Sernani, L. Burattini and A. F. Dragoni, "Lung nodule diagnosis and cancer histology classification from computed tomography data by convolutional neural networks: A survey," *Computers in Biology and Medicine*, vol. 146, Art. no. 105691, 2022.
https://doi.org/10.1016/j.compbiomed.2022.105691
[11] Y. Dong, X. Li, Y. Yang, M. Wang and B. Gao, "A synthesizing semantic characteristics lung nodules classification method based on 3D convolutional neural network," *Bioengineering*, vol. 10, no. 11, Art. no. 1245, 2023.
https://doi.org/10.3390/bioengineering10111245
[12] A. Sau, N. Gautam, A. Basu and R. Sarkar, "From pixels to prognosis: A comprehensive review of classical and modern approaches of lung nodule segmentation for improved lung cancer diagnosis," *ACM Computing Surveys*, vol. 58, no. 10, Art. no. 266, 2026.
https://doi.org/10.1145/3797901
[13] Y. Gu, X. Lu, L. Yang *et al*., "Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs," *Computers in Biology and Medicine*, vol. 103, pp. 220–231, 2018.
https://doi.org/10.1016/j.compbiomed.2018.10.011
[14] W. Zhao, J. Yang, Y. Sun *et al*., "3D deep learning from CT scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas," *Cancer Research*, vol. 78, no. 24, pp. 6881–6891, 2018.
https://doi.org/10.1158/0008-5472.CAN-18-0696
[15] J. Jiang, Y.-C. Hu, C.-J. Liu *et al*., "Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 134–144, 2019.
https://doi.org/10.1109/TMI.2018.2857800

[16] X. Zhao, Y. Wang, H. Liu *et al*., "Multi-scale multi-task 3D convolutional neural network for pulmonary nodule malignancy classification," *Medical Image Analysis*, vol. 72, Art. no. 102123, 2021.
[17] Y. Yu, X. Zhang, H. Li *et al*., "Automatic pulmonary nodule classification using 3D ResNet and CT image segmentation," *Computer Methods and Programs in Biomedicine*, vol. 208, Art. no. 106266, 2021.
[18] M. R. Venkatesh and H. H. S. Hameed, "HybridNoduleNet: Noise-resilient lung nodule classification using CNN–ViT–DAE architecture with Grad-CAM interpretability," *Engineering, Technology & Applied Science Research*, vol. 16, no. 3, pp. 34920–34926, 2026.
https://doi.org/10.48084/etasr.17755