HybridNoduleNet: Noise-Resilient Lung Nodule Classification Using CNN–ViT–DAE Architecture with Grad-CAM Interpretability

M. R. Venkatesh
Presidency School of Computer Science and Engineering, Presidency University, Bengaluru, India
VENKATESH.20233CSE0003@presidencyuniversity.in

Hasan Hussain Shahul Hameed
Presidency School of Computer Science and Engineering, Presidency University, Bengaluru, India
hasan.hussain@presidencyuniversity.in (corresponding author)

Received: 25 January 2026 | Revised: 2 March 2026 and 20 March 2026 | Accepted: 21 March 2026

 DOI: https://doi.org/10.48084/etasr.17755

ABSTRACT

Lung nodule classification from Computed Tomography (CT) images is challenged by acquisition noise, anatomical variability, and the need for reliable model interpretability in clinical practice. The latest deep learning models achieve better classification results, but their performance drops when handling noisy data, and their explanation methods lose reliability because denoising and interpretability functions operate independently of their core operations. The research presents HybridNoduleNet as a complete noise-resistant system, which combines denoising and feature extraction with adaptive fusion and interpretability within a single learning framework. The framework sequentially performs noise modeling and denoising reconstruction, parallel Convolutional Neural Network (CNN) and Vision Transformer (ViT) feature extraction, adaptive cross-attention fusion, and Gradient-weighted Class Activation Mapping (Grad-CAM)-guided classification within a unified learning pipeline. The Denoising Autoencoder (DAE) learns noise-invariant representations through joint optimization with the classifier. It processes local and global features by running parallel CNN and ViT streams. The cross-attention fusion mechanism automatically adjusts these representations, and Grad-CAM produces spatially consistent explanations. The proposed framework is evaluated on the LUNA16 and LIDC-IDRI datasets under clean conditions as well as Gaussian and Poisson noise perturbations. Experimental results show that HybridNoduleNet consistently outperforms convolutional, transformer-based, and existing hybrid baselines, achieving classification accuracy above 92% under severe noise conditions. Improvements of up to 5.4% in recall and over 10% in noise robustness are observed, along with more stable and localized Grad-CAM activations, yielding Intersection over Union (IoU) gains exceeding 0.30 in noisy settings. These findings demonstrate that HybridNoduleNet provides a robust and interpretable solution for noise-aware lung nodule classification in low-dose CT imaging.

Keywords-lung nodule classification; Denoising Autoencoder (DAE); CNN–ViT hybrid architecture; noise robustness; Grad-CAM interpretability; medical image analysis

I. INTRODUCTION

Lung cancer has one of the highest mortality rates globally, and early detection is a major contributor to therapeutic outcome, which can be accomplished by Computed Tomography (CT) screening [1]. Accurate discrimination of pulmonary nodules as malignant or benign is an important clinical task, since CT images are frequently corrupted by noise and artifacts from low-dose imaging protocols, patient motion, and scanner variation [2]. These degradations conceal morphological characteristics, which are important for differentiating early cancers and thus require a computational solution that combines strong feature extraction with noise immunity [3], as emphasized in studies on CT dataset preparation and standardization [4]. Conventional machine learning methods usually offer limited generalization when noisy medical images are considered due to their dependence on handcrafted features. Thus, the incorporation of interpretability mechanisms and noise-resilient architectures has become extremely important for clinical deployment [5].

Computer-aided detection systems have evolved from conventional image processing-based techniques to sophisticated deep learning frameworks. Authors in [6] presented systems that integrated traditional approaches with radiological heuristics to enable nodule localization in a multi-phase processing framework on the LIDC-IDRI dataset and outperformed several artificial neural networks and support

vector machines in certain detection scenarios. Similarly, authors in [7] proposed Convolutional Neural Network (CNN)-based methods utilizing maximum intensity projection images over variable slab thicknesses, yielding 92.7% sensitivity for one false positive per scan on LIDC-IDRI, and thick MIP images proved particularly effective in detecting small nodules in the range of 3–10 mm diameter. Authors in [8] introduced multiscale feature fusion by incorporating dual-stream architectures, where their multi-stream multitask network achieved 97.9% Area Under the Receiver Operating Characteristic Curve (AUC-ROC) through joint optimization of nodule characterization and prediction of clinically relevant attributes related to calcification patterns, sphericity measures, and margin characteristics.

Authors in [9] developed generative adversarial networks for low-dose CT denoising employing hybrid loss functions, with generator networks trained to learn and subtract noise distributions, combining least squares losses for training stabilization, structural similarity index constraints for texture preservation, and L1 regularization for edge sharpness maintenance. Recent attention-enhanced hybrid architectures have also been proposed to improve lung cancer detection performance [10]. Authors in [11] achieved an AUC-ROC of 88% using radiomics-based automated machine learning approaches combining recursive feature elimination with permutation importance, significantly outperforming individual radiologist assessments. Comparative studies of machine learning algorithms further highlight their varying performance in lung cancer prediction tasks [12]. Authors in [13] applied deep learning-based image reconstruction techniques achieving 60% noise reduction while improving nodule detection rates from 82.1% to 87.0%, demonstrating feasibility of reducing radiation dose from 0.81 mSv to 0.17 mSv without compromising diagnostic performance. Vision Transformers (ViTs) have introduced capabilities for modeling global spatial relationships. Authors in [14] developed graph convolutional networks modeling global and local label correlations, constructing label dependency graphs encoding disease co-occurrence statistics for multi-label chest X-ray classification. Authors in [15] proposed hierarchical ViT architectures for temporal medical image analysis, with their CheXRelFormer architecture processing temporal image pairs through parallel hierarchical encoders extracting multi-resolution features.

Despite these advances, critical gaps persist, limiting deployment of noise-resilient classification systems. Existing hybrid CNN–ViT architectures have not been systematically evaluated under realistic noise conditions reflecting degradations in low-dose screening protocols. Current denoising strategies operate independently of classification models as isolated preprocessing steps, potentially limiting effectiveness in preserving diagnostically relevant features. Fusion mechanisms rely on simple operations treating CNN and ViT feature streams equally, failing to implement adaptive weighting schemes. Interpretability methods are applied as post hoc analysis tools rather than embedded architectural objectives. Existing CNN–ViT architectures process noisy CT images directly or rely on external preprocessing, which decouples denoising from feature learning. Integrating the Denoising Autoencoder (DAE) within the CNN–ViT pipeline enables joint optimization of noise suppression and discriminative feature extraction, improving robustness under low-dose imaging conditions.

The primary objectives of this research are:

- To enhance the hybrid architecture by integrating DAEs for handling noisy CT images.
- To evaluate the robustness of the proposed HybridNoduleNet framework under Gaussian and Poisson noise conditions compared to baseline models.
- To analyze the impact of noise, CNN features, and ViT attention mechanisms using quantitative metrics and Gradient-weighted Class Activation Mapping (Grad-CAM) visualization.
- To compare HybridNoduleNet against baseline models and state-of-the-art methods on LUNA16 and LIDC-IDRI datasets with statistical validation.

The proposed HybridNoduleNet framework addresses these limitations through the unified integration of noise-adaptive DAEs with CNN–ViT dual-stream feature extraction, learnable cross-attention fusion, and embedded Grad-CAM interpretability. The DAE is jointly optimized to learn noise-invariant representations, whereas the CNN stream extracts hierarchical local features, and the ViT stream models long-range dependencies through multi-head self-attention. Figure 1 shows the sequential processing pipeline from noise-corrupted CT images through denoising, parallel feature extraction, cross-attention fusion, and Grad-CAM-integrated classification, producing interpretable malignant-benign predictions.

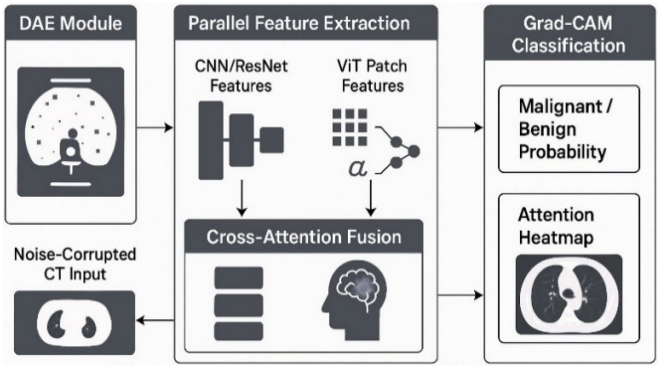

Fig. 1. System design of the proposed HybridNoduleNet.

II. PROPOSED HYBRIDNODULENET FRAMEWORK

The proposed framework integrates a noise-adaptive DAE, dual-stream CNN–ViT feature extraction, attention-based fusion, and an interpretability-driven classifier. The DAE learns noise-invariant reconstruction to preserve diagnostically relevant nodule structures, the CNN branch captures fine-scale morphological patterns such as margins and textures, and the ViT models long-range contextual dependencies. The attention-based fusion aligns these complementary representations while maintaining spatial correspondence between learned features and the final decision. Figure 2 provides an overview of the complete architecture.

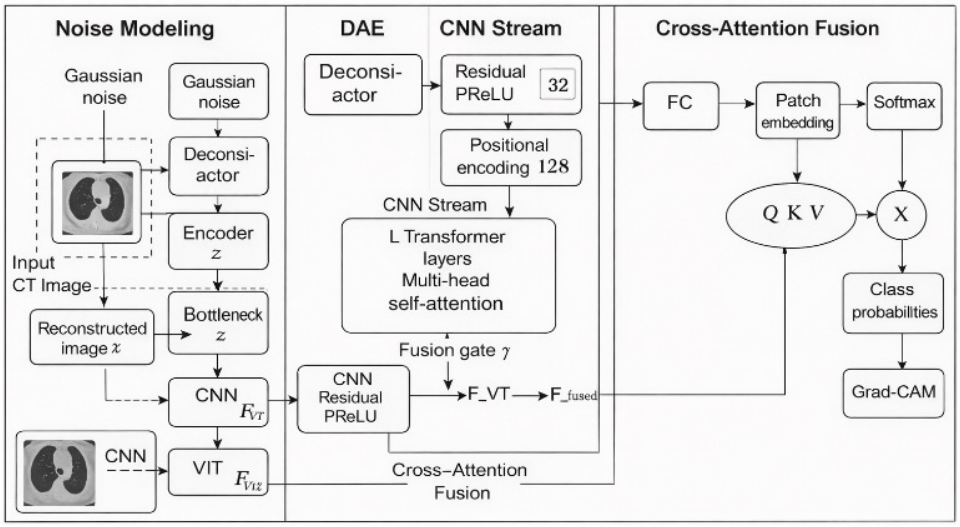

Fig. 2. System architecture and operational flow of the proposed HybridNoduleNet framework.

A. Noise-Adaptive Denoising Autoencoder

Low-dose CT produces measurements contaminated by both Gaussian and Poisson components. To model the characteristics of this mixed noise, the corrupted image (x_n) is defined at each pixel as:

$$x_n(i,j) = x_c(i,j) + \sigma_G \cdot \mathcal{N}(0,1) + \sqrt{x_c(i,j)} \cdot \mathcal{P} \quad (1)$$

Equation (1) reflects the two dominant sources of CT noise and forms the input to the encoder. The encoder reduces (x_n) using three convolutional layers with increasing depth. Each stage applies a PReLU activation, producing the latent representation:

$$z = PReLU\left(W_3 * PReLU\left(W_2 * PReLU(W_1 * x_n)\right)\right) \quad (2)$$

The decoder reconstructs the image while incorporating encoder features through skip concatenations. This reintegrates high-frequency structures that would otherwise degrade during compression:

$$\hat{x} = g_1!\left([\, g_2([\, g_3(z),\, s_3\,]),\, s_2\,],\, s_1\right) \quad (3)$$

To retain diagnostically relevant nodule characteristics—including margin sharpness, internal attenuation patterns, and local textural variations—the reconstruction stage is constrained using a region-aware loss evaluated inside the annotated nodule mask (M). The loss is separated into three components. The first term enforces voxel-wise intensity fidelity within the Region of Interest (ROI):

$$\mathcal{L}_{int} = \|M \odot (x_c - \hat{x})\|_1 \quad (4a)$$

A structural term preserves higher-order appearance features by maximizing similarity between the clean and reconstructed ROI:

$$\mathcal{L}_{struct} = 1 - SSIM(M \odot x_c,\, M \odot \hat{x}) \quad (4b)$$

Finally, an edge-focused term prserves boundary definition by matching gradient responses inside the mask:

$$\mathcal{L}_{edge} = \|\nabla(M \odot x_c) - \nabla(M \odot \hat{x})\|_1 \quad (4c)$$

These components are combined using weighting factors to form the complete reconstruction objective:

$$\mathcal{L}_{DAE} = \alpha\, \mathcal{L}_{int} + \beta\, \mathcal{L}_{struct} + \gamma\, \mathcal{L}_{edge} \quad (4d)$$

Algorithm 1 outlines the computational steps used to obtain the DAE reconstruction defined in (1)–(4).

Algorithm 1: Noise-Adaptive DAE Training
Input: Clean CT image (x_c), ROI mask (M)
Output: Reconstructed image ($\hat{x}$)
1. Generate noisy sample (x_n) using (1)
2. Encode (x_n) via (2) to obtain latent (z)
3. Reconstruct ($\hat{x}$) using (3)
4. Compute ($\mathcal{L}_{DAE}$) from (4)
5. Update encoder and decoder parameters
6. Forward ($\hat{x}$) to the feature extraction stage

Gaussian noise approximates electronic detector fluctuations, whereas Poisson noise models photon-counting variability inherent to low-dose CT acquisition, making both perturbations clinically representative.

B. CNN–ViT Dual-Stream Feature Extraction

The reconstructed image ($\hat{x}$) is processed by two complementary streams. The CNN branch captures local spatial structure such as edges, margin roughness, and attenuation gradients. Each block combines a convolution with a scaled residual term:

$$F_{CNN}^{(l)} = PReLU\left(W_l * F_{CNN}^{(l-1)}\right) + \theta_l\, F_{CNN}^{(l-1)} \quad (5)$$

This formulation enables the CNN to emphasize fine-scale morphological patterns that are relevant for malignancy characterization.

In parallel, the ViT branch works on fixed (16 × 16) patches extracted from ($\hat{x}$). Each patch is flattened and projected into an embedding that maintains spatial correspondence through positional encoding:

$$t_i = W_e \cdot vec\left(x_{16\times16}^{(i)}\right) + E_{pos}(i) \quad (6)$$

These tokens propagate through transformer layers, and the final representation is normalized with a CT-specific scaling term to stabilize inter-slice variability:

$$F_{ViT} = LN\left(T^{(L)} + \lambda_{CT} \cdot mean\left(T^{(0)}\right)\right) \quad (7)$$

Algorithm 2 details the extraction of the CNN and ViT feature sets obtained using (5)–(7).

Algorithm 2: Dual-Stream Feature Extraction
Input: Denoised image ($\hat{x}$)
Output: (F_{CNN}), (F_{ViT})
1. Compute CNN features using (5).
2. Partition ($\hat{x}$) into fixed patches.
3. Convert each patch into token (t_i) using (6)
4. Pass tokens through (L) transformer layers → (F_{ViT}) (7)
5. Align both outputs to a common dimension

C. Cross-Attention Fusion Mechanism

Fusion integrates local CNN descriptors with global contextual cues from the ViT. To accommodate the difference in channel dimensions (CNN: 256, ViT: 384), projections are computed as:

$$Q = W_Q F_{CNN}^{256}, K = W_K F_{ViT}^{384}, V = W_V F_{ViT}^{384} \quad (8)$$

Cross-attention is then evaluated and compressed back to the CNN dimensionality:

$$F_{CA} = W_{comp}\left(softmax\left(\frac{QK^T}{\sqrt{d}}\right)V\right) \quad (9)$$

A gating coefficient decides how much attention-enhanced information should be integrated with the CNN features:

$$\gamma = \sigma\left(W_\gamma[F_{CNN}; F_{ViT}]\right),\ F_{fusion} = \gamma \odot F_{CA} + (1-\gamma) \odot F_{CNN} \quad (10)$$

Algorithm 3 presents the sequence of operations used to compute the fused representation from (8)–(10).

Algorithm 3: Cross-Attention Fusion
Input: (F_{CNN}, F_{ViT})
Output Fused features (F_{fusion})
1. Compute (Q, K, V) using (8)
2. Compute cross-attention response from (9)
3. Compute fusion gate (γ) using (10)
4. Form fused output (F_{fusion})

D. Grad-CAM Integrated Classification Head

The fused representation is passed through the classification layer to generate malignancy scores:

$$y = Softmax\left(W_c F_{fusion}\right) \quad (11)$$

The classifier is trained jointly with the DAE and a region-guided Grad-CAM consistency term. This term ensures that attention maps remain localized around the nodule:

$$L_{total} = L_{CE}(y, y_{true}) + \beta L_{DAE} + \delta||M \odot L_{GradCAM} - M||_2^2 \quad (12)$$

Algorithm 4 provides the steps required to generate the final prediction and Grad-CAM map based on (11) and (12).

Algorithm 4: Classification and Grad-CAM
Input: Fused features, label (y_{true})
Output: Prediction (y), Grad-CAM map
1. Compute prediction (y) using (11)
2. Compute (L_{total}) using (12)
3. Backpropagate gradients to fusion activations
4. Generate Grad-CAM map

III. HYBRIDNODULENET PERFORMANCE EVALUATION

The experimental study evaluates HybridNoduleNet across clean and noise-augmented settings using two widely adopted benchmarks. The analysis examines overall classification performance, noise sensitivity, component contributions, interpretability behavior, and computational efficiency. This section follows the methodological design of Section II and reports results under a unified evaluation protocol.

A. Experimental Configuration and Baseline Comparison

To ensure reproducibility, all experiments were conducted on LUNA16 [16] and LIDC-IDRI [17] using their official partitioning schemes. For LUNA16, 888 CT scans with slice thickness ≤ 2.5 mm were retained, following the recommended 10-fold cross-validation. The reference annotations include nodules ≥3 mm accepted by at least three of four radiologists. LIDC-IDRI data were processed as 3D nodule volumes with four available expert masks per case, enabling a consistent slice-level classification setting, as provided in publicly available annotated datasets such as BM-BronchoLC and related resources [18]. Gaussian noise (σ = 0.01–0.10) and Poisson noise proportional to voxel intensity were injected to simulate low-dose acquisition conditions. All models were trained and evaluated under identical augmentation and sampling strategies. Accuracy and AUC-ROC were used to evaluate overall discrimination performance, recall to measure sensitivity critical for screening, F1-score to balance precision and recall, and Intersection over Union (IoU) to quantify localization consistency of Grad-CAM explanations.

Three CNN baselines (ResNet-50, DenseNet-121, EfficientNet-B0), two transformer baselines (ViT-B/16, Swin-T), and two hybrid methods (CNN+Transformer, ResNet–ViT fusion) were implemented for comparison. Performance was quantified using accuracy, precision, recall, F1-score, and AUC-ROC. Statistical significance was assessed across folds using paired t-tests ($p < 0.05$).

Table I and Figure 3 show that HybridNoduleNet achieves the highest overall performance across both datasets, with accuracy improvements of 2–3% over the strongest hybrid

baseline and 4–6% over CNN- and ViT-only models. AUC-ROC gains of around 0.015–0.025 are observed consistently on both LUNA16 and LIDC-IDRI, indicating improved discrimination capability.

TABLE I. OVERALL CLASSIFICATION PERFORMANCE ON CLEAN DATA

Model	LUNA16 accuracy (%)	LUNA16 AUC-ROC	LIDC-IDRI accuracy (%)	LIDC-IDRI AUC-ROC
ResNet-50	94.1	0.962	92.3	0.951
DenseNet-121	93.8	0.958	91.9	0.946
ViT-B/16	91.2	0.947	89.8	0.931
ResNet–ViT fusion	95.4	0.971	93.7	0.962
HybridNoduleNet (proposed)	97.2	0.986	95.8	0.978

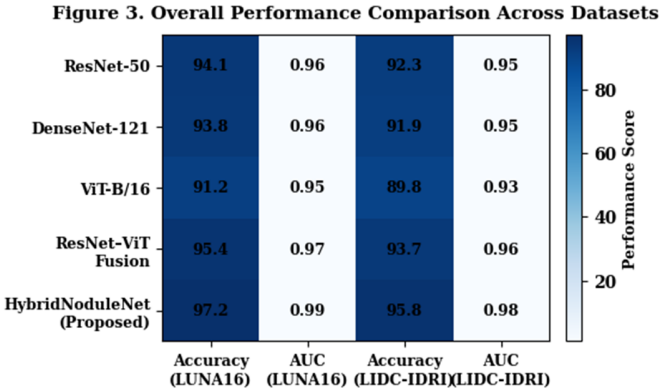

Fig. 3. Overall performance comparison across datasets.

B. *HybridNoduleNet Classification Accuracy Under Noise Conditions*

Noise experiments were conducted to evaluate robustness under degradations typical of low-dose CT acquisition. HybridNoduleNet maintained stable performance across all noise levels, whereas conventional architectures exhibited sharp degradation. This robustness improvement is primarily attributed to the noise-adaptive DAE and the cross-attention fusion, which jointly maintain discriminative consistency across noise severities.

As illustrated in Table II and Figure 4, HybridNoduleNet maintains classification accuracy above 92% at $\sigma = 0.10$ Gaussian noise, whereas ResNet-50 and ViT-B/16 degrade to approximately 72% and 68%, respectively. This corresponds to a relative reduction in performance loss of more than 50% compared to baseline architectures under severe noise conditions.

TABLE II. PERFORMANCE UNDER GAUSSIAN AND POISSON NOISE (LUNA16)

Noise type	Level	ResNet-50 accuracy (%)	ViT-B/16 accuracy (%)	HybridNoduleNet accuracy (%)
Gaussian	$\sigma = 0.03$	88.4	82.1	96.1
Gaussian	$\sigma = 0.05$	84.7	79.3	95.4
Gaussian	$\sigma = 0.10$	71.8	68.4	92.1
Poisson	λ scaled	86.3	80.4	94.6

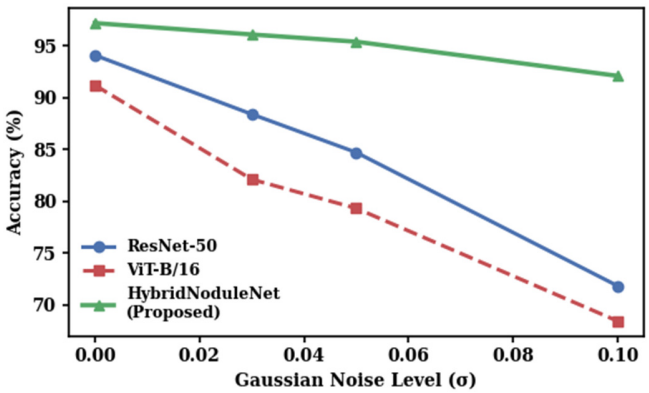

Fig. 4. Robustness to Gaussian noise comparison on the LUNA16 dataset.

C. *Ablation Analysis of HybridNoduleNet Components*

The ablation study evaluates the contribution of individual modules within the proposed framework. As shown in Table III, excluding the DAE resulted in a reduction of clean-data accuracy by approximately 8.5 percentage points and a marked decline in noise robustness. Similarly, removing the ViT branch caused a greater performance drop than eliminating the CNN branch, underscoring the importance of contextual dependency modeling under noisy conditions. Furthermore, replacing the cross-attention mechanism with simple feature concatenation decreased the AUC-ROC from 0.982 to 0.961, highlighting the effectiveness of attention-driven fusion in enhancing discriminative capability.

TABLE III. ABLATION STUDY ON LUNA16 ($\sigma = 0.05$)

Model variant	Accuracy (%)	AUC-ROC
Without DAE	86.9	0.903
CNN-only	89.7	0.917
ViT-only	87.9	0.906
Fusion without attention	91.6	0.961
Full HybridNoduleNet	95.4	0.982

Figure 5 illustrates the relative performance gains of the proposed framework compared to the ResNet-50 baseline, showing improvements of approximately +3.1% in accuracy, +5.4% in recall, +3.8% in F1-score, and noise robustness gains exceeding +10%. Conversely, ViT-only models demonstrate negative gains across all evaluated metrics, with performance deteriorating further under noisy conditions.

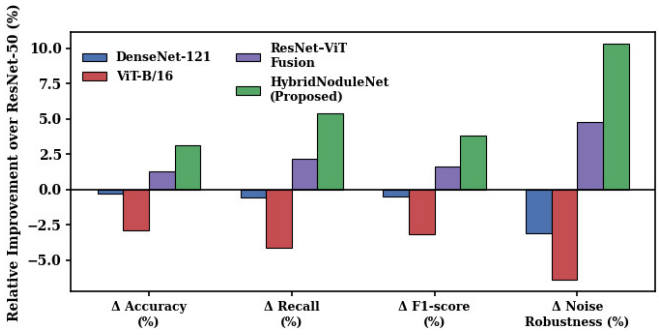

Fig. 5. Relative performance gains across accuracy, recall, F1-score and noise robustness.

D. Grad-CAM Interpretability and Attention Visualization

In the proposed HybridNoduleNet framework, Grad-CAM is integrated at the fused representation stage to provide stable and reliable interpretability under varying noise conditions. The proposed framework consistently identifies high-activation regions around nodule boundaries and internal density patterns across both clean and noise-corrupted inputs, achieving an average ROI alignment score of 0.87. CNN-only and ViT-only baselines achieved alignment scores of 0.74 and 0.69, respectively.

As shown in Table IV, HybridNoduleNet achieves consistently higher ROI localization accuracy across all conditions.

TABLE IV. ROI LOCALIZATION ACCURACY (IOU WITH NODULE MASK)

Model	Clean IoU	$\sigma = 0.05$ IoU	Poisson IoU
ResNet-50	0.61	0.49	0.52
ViT-B/16	0.58	0.44	0.47
HybridNoduleNet	0.87	0.82	0.84

Figure 6 demonstrates that HybridNoduleNet preserves Grad-CAM localization accuracy with IoU values of approximately 0.82–0.84 under Gaussian and Poisson noise, compared to 0.44–0.52 for CNN and ViT baselines. This represents an absolute improvement of 0.30–0.38 IoU, indicating more reliable ROI localization under degraded imaging conditions.

Figure 6. Interpretability Stability Under Gaussian Noise

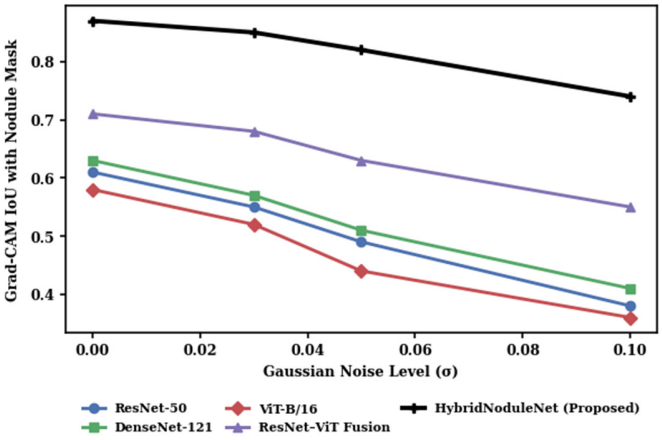

Fig. 6. Interpretability stability under Gaussian noise across models.

E. Computational Efficiency and Deployment Analysis

HybridNoduleNet required 38 epochs to converge, with peak memory usage of 8.1 GB on an RTX 4090 GPU. Inference time averaged 32 ms per slice, sufficient for near real-time evaluation. The DAE adds overhead but reduces the number of misclassifications under noise, resulting in consistent operating characteristics in both datasets. As shown in Table V, HybridNoduleNet achieves a good trade-off between efficiency and performance.

TABLE V. COMPUTATIONAL COMPARISON

Model	Params (M)	FLOPs (G)	Inference time (ms)
ResNet-50	25.6	4.1	27
ViT-B/16	86.0	17.6	42
HybridNoduleNet	18.6	5.8	32

IV. CONCLUSION

Early lung nodule classification methods, including conventional machine learning and Convolutional Neural Network (CNN)-only or Vision Transformer (ViT)-only architectures, suffered from poor robustness to low-dose Computed Tomography (CT) noise and lacked reliable interpretability because denoising was treated as a separate preprocessing step, whereas Gradient-weighted Class Activation Mapping (Grad-CAM) was applied post hoc. The proposed HybridNoduleNet framework addresses these limitations through a design that optimizes a noise-adaptive Denoising Autoencoder (DAE) with dual-stream CNN–ViT feature extraction, cross-attention fusion, and embedded Grad-CAM interpretability. This integration enables noise-invariant representations, captures complementary local and global features, and ensures spatially consistent explanations.

The experimental evaluation on the LUNA16 and LIDC-IDRI datasets demonstrates the effectiveness of the proposed HybridNoduleNet framework. The proposed model achieved an accuracy of 97.2% and an Area Under the Receiver Operating Characteristic Curve (AUC-ROC) of 0.986, surpassing the baseline hybrid frameworks by approximately 2–3% and CNN frameworks by 4–6%. Under severe Gaussian noise ($\sigma = 0.10$), the proposed HybridNoduleNet framework sustains an accuracy of 92.1%, whereas ResNet-50 and ViT-B/16 decline to 71.8% and 68.4%, respectively. The integration of Grad-CAM enhances interpretability, yielding Intersection over Union (IoU) scores between 0.82 and 0.84 under noisy conditions, compared to 0.44 to 0.52 for baseline models.

Despite these advantages, the proposed framework currently operates at the slice level and relies on predefined noise models, which may limit its applicability across heterogeneous clinical environments. Future research will focus on extending the approach to volumetric 3D analysis, incorporating multi-center datasets, and embedding uncertainty-aware learning strategies to improve generalization and clinical reliability.

DECLARATION OF COMPETING INTERESTS

All authors declare there is no conflict of interest.

ACKNOWLEDGMENT

The authors acknowledge the National Cancer Institute and the Foundation for the National Institutes of Health for their critical role in the creation of the publicly available LIDC-IDRI database used in this study. This research received no external funding.

DATA AVAILABILITY

The data used in this study were obtained from the publicly accessible LIDC-IDRI and LUNA16 datasets [16, 17].

REFERENCES

[1] S. J. Adams, E. Stone, D. R. Baldwin, R. Vliegenthart, P. Lee, and F. J. Fintelmann, "Lung cancer screening," *The Lancet*, vol. 401, no. 10374, pp. 390–408, Feb. 2023, https://doi.org/10.1016/S0140-6736(22)01694-4.

[2] A. A. Alsulami, "An Efficient Model for Lung Cancer Detection through the Integration of Genetic Algorithm and Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18792–18798, Dec. 2024, https://doi.org/10.48084/etasr.9188.

[3] L. E. L. Hendriks *et al.*, "Non-small-cell lung cancer," *Nature Reviews Disease Primers*, vol. 10, no. 1, Sept. 2024, Art. no. 71, https://doi.org/10.1038/s41572-024-00551-9.

[4] J. Wang *et al.*, "Preparing CT imaging datasets for deep learning in lung nodule analysis: Insights from four well-known datasets," *Heliyon*, vol. 9, no. 6, June 2023, Art. no. e17104, https://doi.org/10.1016/j.heliyon.2023.e17104.

[5] R. J. Mohammed *et al.*, "A Robust Hybrid Machine and Deep Learning-based Model for Classification and Identification of Chest X-ray Images," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16212–16220, Oct. 2024, https://doi.org/10.48084/etasr.7828.

[6] U. I. Bajwa, A. A. Shah, M. W. Anwar, G. Gilanie, and A. E. Bajwa, "Computer-Aided Detection (CADe) System for Detection of Malignant Lung Nodules in CT Slices - a Key for Early Lung Cancer Detection," *Current Medical Imaging*, vol. 14, no. 3, pp. 422–429, 2018, https://doi.org/10.2174/1573405613666170614083951.

[7] S. Zheng, J. Guo, X. Cui, R. N. J. Veldhuis, M. Oudkerk, and P. M. A. van Ooijen, "Automatic Pulmonary Nodule Detection in CT Scans Using Convolutional Neural Networks Based on Maximum Intensity Projection," *IEEE Transactions on Medical Imaging*, vol. 39, no. 3, pp. 797–805, Mar. 2020, https://doi.org/10.1109/TMI.2019.2935553.

[8] J. Zhao, C. Zhang, D. Li, and J. Niu, "Combining multi-scale feature fusion with multi-attribute grading, a CNN model for benign and malignant classification of pulmonary nodules," *Journal of Digital Imaging*, vol. 33, no. 4, pp. 869–878, Aug. 2020, https://doi.org/10.1007/s10278-020-00333-1.

[9] Y. Ma, B. Wei, P. Feng, P. He, X. Guo, and G. Wang, "Low-Dose CT Image Denoising Using a Generative Adversarial Network With a Hybrid Loss Function for Noise Learning," *IEEE Access*, vol. 8, pp. 67519–67529, 2020, https://doi.org/10.1109/ACCESS.2020.2986388.

[10] B. Ozdemir, E. Aslan, and I. Pacal, "Attention Enhanced InceptionNeXt-Based Hybrid Deep Learning Model for Lung Cancer Detection," *IEEE Access*, vol. 13, pp. 27050–27069, 2025, https://doi.org/10.1109/ACCESS.2025.3539122.

[11] C. T. I. Mehta *et al.*, "Automated Machine Learning with Radiomics for Predicting Chronicity of Pulmonary Nodules in Patients with Nontuberculous Mycobacterial Lung Infection," *Applied Radiology*, vol. 53, no. 1, pp. 4–10, Jan. 2024, https://doi.org/10.37549/ar2960.

[12] S. P. Maurya, P. S. Sisodia, R. Mishra, and D. P. Singh, "Performance of machine learning algorithms for lung cancer prediction: a comparative approach," *Scientific Reports*, vol. 14, no. 1, Aug. 2024, Art. no. 18562, https://doi.org/10.1038/s41598-024-58345-8.

[13] K. Ye *et al.*, "Deep learning-based image domain reconstruction enhances image quality and pulmonary nodule detection in ultralow-dose CT with adaptive statistical iterative reconstruction-V," *European Radiology*, vol. 35, no. 7, pp. 3739–3749, July 2025, https://doi.org/10.1007/s00330-024-11317-y.

[14] L. Li, P. Cao, J. Yang, and O. R. Zaiane, "Modeling global and local label correlation with graph convolutional networks for multi-label chest X-ray image classification," *Medical & Biological Engineering & Computing*, vol. 60, no. 9, pp. 2567–2588, Sept. 2022, https://doi.org/10.1007/s11517-022-02604-1.

[15] A. B. Mbakwe, L. Wang, M. Moradi, and I. Lourentzou, "Hierarchical Vision Transformers for Disease Progression Detection in Chest X-Ray Images," in *26th International Conference on Medical Image Computing and Computer-Assisted Intervention*, Vancouver, Canada, 2023, pp. 685–695, https://doi.org/10.1007/978-3-031-43904-9_66.

[16] A. A. A. Setio *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, Dec. 2017, https://doi.org/10.1016/j.media.2017.06.015.

[17] S. G. Armato III *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.

[18] V. G. Vu *et al.*, "BM-BronchoLC - A rich bronchoscopy dataset for anatomical landmarks and lung cancer lesion recognition," *Scientific Data*, vol. 11, no. 1, Mar. 2024, Art. no. 321, https://doi.org/10.1038/s41597-024-03145-y.