Multimodal AI for Pulmonary Nodule Classification: A Review

Vinod Kumar[1], Aleem Ali[2]

[1] Postdoctoral Researcher, Lincoln University College, Petaling Jaya, Selangor 47301, Malaysia
[2] MM Engineering College, Maharishi Markandeshwar (Deemed) University, Mullana, India
[3] University Institute of Engineering, Chandigarh University, Mohali, Punjab, India
[1,2] pdf.vinodkumar@lincoln.edu.my, [3] aleem.e12948@cumail.in
ORCID: [1,2] 0000-0002-2333-8931; [3] 0000-0001-5298-4775

Abstract: The classification of pulmonary nodules by computed tomography (CT) is a crucial step in computer-aided lung cancer diagnosis, but it is still time-consuming and highly inter-observer dependent, and pertinent literature of convolutional neural networks (CNN) for this problem is extensive, diverse, and hard to synthesize. This paper presents a systematic review and categorizes the CNN related pulmonary nodule classification studies into an eight-stage pipeline with an additional stage of explainable AI, in order to make it easier to refer to different studies: Dataset curation, CT preprocessing, ROI extraction, CNN backbone selection and ensembling, attention based feature fusion, radiomic and clinical feature integration, multimodal classification, uncertainty quantification, explainable AI, and clinical validation. The backbone architectures and attention-based fusion are fairly well developed, whereas uncertainty quantification, validation of explainability, and eventual clinical utility are less mature, with pooled external-validation studies reporting greater than 85% sensitivity and less than 80% specificity. There are nine identified research gaps that range from diversity of datasets to harmonization of preprocessing, benchmarking in fusion, and outcome-related validation. The results provide a useful guide for researchers and clinical-AI developers looking for pulmonary nodule classification systems that can extrapolate from the benchmark set and have clinical confidence.

Keywords: Nodule Classification; Convolutional Neural Networks; Deep Feature Fusion; Radiomics; Explainable AI; Clinical Validation.

Introduction

Lung cancer is still the biggest killer of all cancers globally and early identification of pulmonary nodules via low-dose CT is one of the most promising tools to improve survival rates – five-year survival drops from >60% at an early stage to <15% at an advanced stage [9, 1]. Interpretation of the malignancy of a nodule by a radiologist is time-consuming and has been shown to be interobserver variable, especially for small, sub-solid and indeterminate nodules [1]. They have motivated the research on computer-aided diagnosis based on CNN over a decade, ranging from a single backbone CNN to CNN ensemble, attention-based CNN fusion, radiomics-enhanced CNN and multimodal CNN, which incorporate imaging information, radiomics, and clinical information.

This literature is spread out over the areas of radiology, biomedical engineering, and computer vision, making it difficult to identify which aspects of the methodology are ripe for development and which are still open problems. Instead this review will focus on a conceptual pipeline of nine stages that reflects how a contemporary classification is actually created, and will identify research gaps at each stage of the process, thus avoiding the blanket approach that considers the field as a whole. It serves three purposes: a progressive review of the CNN literature, a streamlined list of nine research gaps, and a list of future trends to fill the gap between best performance and clinically viable systems.

Related Work

Recently, a number of reviews have been published on the various stages of the lung-nodule analysis pipeline using CNN, but these reviews focus on a smaller portion of the pipeline. Although Gao et al. [4] give a systematic review of the deep learning algorithms for nodule detection and segmentation, they do not move on to the classification and fusion stage. Forte et al. [9] conducted a meta-analysis of the diagnostic accuracy of CNN but it does not cover the uncertainty or explainability of the CNN. While there are a number of studies that have addressed external validation of AI models, such as those done by Asmara et al. [1] and Arun et al. [2], they have all concentrated on the external validation of AI models without introducing findings into a complete pipeline. These reviews are compared with the present work in Table 1 on four organizing dimensions.

Table 1. Comparison of this review with related systematic reviews

Review	Pipeline coverage	calibration	Explainability	External validation
Gao et al. [4]	Detection & segmentation only	No	No	Partial
Forte et al. [9]	Diagnostic accuracy (meta-analysis)	No	No	No
Asmara et al. [1]	External validation only	No	No	Yes
Arun et al. [2]	External validation only	No	No	Yes
This work	All nine pipeline stages	Yes	Yes	Yes

Key Contribution:

- A nine-stage conceptual pipeline for dataset curation to clinical validation for CNN-based classification of pulmonary nodules.
- A staged assessment of evidence maturity, with well studied components (backbones, ensembling, attention fusion) and under developed components (uncertainty, explainability validation, outcome linked utility).
- Nine concrete research gaps identified as being linked to the stages, and five future directions identified to bridge the gap between the accuracy of the benchmark and the clinically deployable and trusted system.

Proposed framework:

The nine-stage pipeline illustrated in figure 1 was inductively derived as a common structure of end-to-end pulmonary nodule classification systems reported in the literature and was used to organize this review.

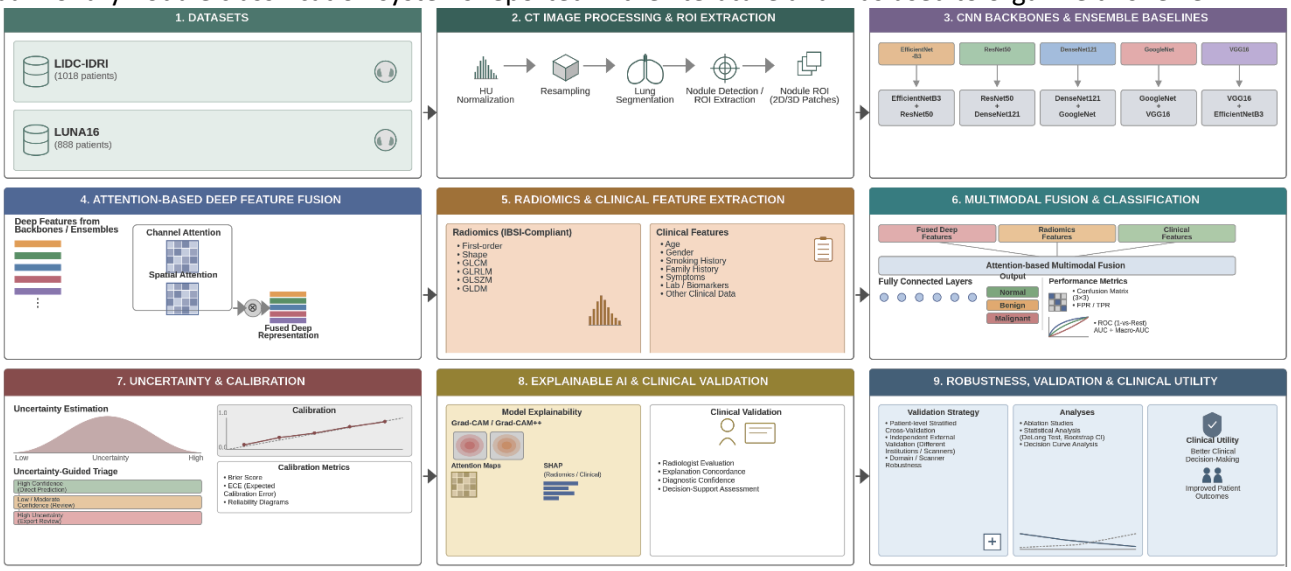


Figure 1. Pulmonary nodule classification system architecture based on CNN.

Datasets and Preprocessing:

The two public datasets are LIDC-IDRI (approximately 1018 patients) and LUNA16 (888 curated scans) [4]. They are used almost universally, so they can be used to make comparisons across studies, but the specificity of the datasets decreases to 70-80% when sensitivity is high (as shown by the external testing of models), and the accuracy of independent studies ranges from 68-99.6%, which indicates sensitivity to the specific study chosen as well as to the model design. Upstream, robustness is affected by intensity variability across scanners which can lead to a loss of performance of 10-20% in detection under domain shift, and there is a lack of complete and detailed reporting of the various preprocessing steps used.

Backbones, ensembles and attention fusion. Backbones of EfficientNet, ResNet, DenseNet, GoogLeNet, and VGG variants are recur, with a focus on either accuracy or efficiency, gained with residual connections, dense connectivity and/or compound scaling. A simple baseline is to ensemble multiple different architectures, and ensembling with channel/spatial attention has further reduced false positives when compared to single CNN detectors [3]. Cross-modal fusion is extended in more elaborate schemes, such as I-VISTA, which combines a Swin-Transformer branch, a convolutional-autoencoders branch, and a 3-D radiomic branch using crisscross attention that achieves 93.9% accuracy and AUC 0.93 on a sub solid-nodule invasiveness task [8], and multi-feature attention fusion that has been evaluated across multiple hospital sites, showing superior cross-scanner

generalization compared to naive feature concatenation [10]. Most of the ensemble comparisons have not been "parameter count controlled", however, and most attention weights have never been validated against regions annotated by radiologists.

Mathematical Model of the Proposed Framework:

Overall formulation

Let a patient CT examination be represented by

$$\mathcal{D} = \{(X_i, y_i)\}_{i=1}^{N}$$

where

- X_i = CT volume of the i^{th} patient,
- y_i = corresponding diagnostic class,
- N = number of patient cases.

For pulmonary nodule classification, $y_i \in \{0,1\}$

Where $y_i = \begin{cases} 0, & \text{Benign nodule} \\ 1, & \text{Malignant nodule} \end{cases}$

If the framework is extended to normal/benign/malignant classification, $y_i \in \{0,1,2\}$. The proposed model can be expressed as

$$\boxed{\hat{y} = F_\Theta(X, R(X), \Phi_{CNN}(X), \Phi_{Rad}(R), \Phi_{Clin}(C))}$$

where

- $R(X)$ = extracted region of interest/nodule,
- Φ_{CNN} = deep CNN feature extractor,
- Φ_{Rad} = radiomic feature extractor,
- Φ_{Clin} = clinical feature representation,
- F_Θ = multimodal fusion and classification function,
- Θ = trainable parameters.

Stage I — CT preprocessing

Let the original CT volume be $X \in \mathbb{R}^{H\times W\times D}$.

Each voxel intensity is represented in Hounsfield Units (HU):

$$HU = mP + b$$

where P is the stored pixel value and m, b are the DICOM rescale parameters.

The normalized CT volume can be defined as $X_N = \frac{X-\mu_X}{\sigma_X}$

Where $\mu_X = \frac{1}{M}\sum_{j=1}^{M} X_j$

and $\sigma_X = \sqrt{\frac{1}{M}\sum_{j=1}^{M}(X_j - \mu_X)^2}$.

Alternatively, after lung-window clipping,

$$X_C = \min(\max(X, L), U)$$

where L and U are lower and upper HU limits.

The normalized image is then $X_P = \frac{X_C - L}{U - L}$.

The manuscript specifically identifies HU normalization and reproducible preprocessing as important requirements for robust models.

Stage II — ROI / pulmonary nodule extraction

Let the segmentation function be $S = f_S(X_P; \theta_S)$

Where $S \in [0,1]^{H\times W\times D}$ is the predicted nodule probability map.

The binary nodule mask is $M(x) = \begin{cases} 1, & S(x) \geq \tau_s \\ 0, & S(x) < \tau_s \end{cases}$

where τ_s is the segmentation threshold.

The extracted ROI is $R = X_P \odot M$ where $\odot$ represents element-wise multiplication. Thus, $\boxed{R = f_R(X_P)}$ provides the localized pulmonary nodule used by subsequent feature-extraction stages.

Stage III — CNN deep feature extraction

Let R be passed through K CNN backbones:

$$z_k = \Phi_k(R; \theta_k), \qquad k = 1, \dots, K.$$

For example,

$$\Phi_1 = \text{ResNet},$$
$$\Phi_2 = \text{DenseNet},$$
$$\Phi_3 = \text{EfficientNet}.$$

The combined deep representation is $Z_D = [z_1, z_2, \dots, z_K]$.

Stage IV — Attention-based deep feature fusion

Instead of simple concatenation, assign an attention weight to each CNN representation.

For each feature vector, $e_k = v^T \tanh(W_k z_k + b_k)$

and normalize the attention coefficients using softmax:

$$\alpha_k = \frac{\exp(e_k)}{\sum_{j=1}^{K}\exp(e_j)}.$$

Therefore, $\sum_{k=1}^{K}\alpha_k = 1$. The fused deep representation becomes $\boxed{Z_A = \sum_{k=1}^{K}\alpha_k z_k}$ where Z_A represents the attention-weighted deep feature representation. This is preferable to assuming that every CNN contributes equally. The manuscript reports that attention-based fusion and ensembling are relatively mature components but also identifies the lack of rigorous validation of attention maps as a research gap.

Stage V — Radiomic feature representation

From the segmented ROI R, extract a radiomic feature vector:

$$Z_R = [r_1, r_2, \dots, r_m]^T.$$

The radiomic vector can include

$$Z_R = [r_{\text{shape}}, r_{\text{first-order}}, r_{\text{texture}}, r_{\text{GLCM}}, r_{\text{GLRLM}}, r_{\text{GLSZM}}]^T.$$

Standardization is applied: $r_j' = \frac{r_j - \mu_j}{\sigma_j}$.

Hence, $Z_R' = \left[\frac{r_1-\mu_1}{\sigma_1}, \dots, \frac{r_m-\mu_m}{\sigma_m}\right]^T$.

Stage VI — Clinical feature representation

Let the clinical vector be $C = [c_1, c_2, \dots, c_q]^T$ where variables may include demographic and clinical characteristics such as age, sex, smoking

history and biomarkers, consistent with the manuscript's multimodal formulation.

Normalize continuous clinical variables: $c_j' = \frac{c_j - \mu_{c_j}}{\sigma_{c_j}}$.

The clinical representation is $Z_C = \Phi_C(C; \theta_C)$.

Stage VII — Multimodal feature fusion

The three information streams are

Z_A = deep attention features,
Z_R = radiomic features,
Z_C = clinical features.

A basic multimodal representation is $Z_F = [Z_A \| Z_R \| Z_C]$

where ‖ denotes concatenation. However, a stronger formulation is weighted late fusion:

$$\boxed{Z_F = \beta_D Z_A + \beta_R Z_R + \beta_C Z_C}$$

subject to $\beta_D + \beta_R + \beta_C = 1,\ \beta_D, \beta_R, \beta_C \geq 0$.

The fusion coefficients therefore quantify the contribution of each modality. This follows the manuscript's observation that late fusion is generally preferred for heterogeneous modalities, while comparisons of fusion strategies remain limited.

Stage VIII — Classification model

The fused representation is passed to a classification layer:

$$z_o = W_o Z_F + b_o.$$

For binary classification, the malignancy probability is

$$\boxed{p(y = 1|X) = \sigma(z_o)}$$

Where $\sigma(z) = \frac{1}{1+e^{-z}}$. Thus, $p_M = \frac{1}{1+e^{-z_o}}$ represents the predicted probability of malignancy.

The final prediction is $\hat{y} = \begin{cases} 1, & p_M \geq \tau_c \\ 0, & p_M < \tau_c. \end{cases}$

For three-class classification, use softmax:

$$p_k = \frac{e^{z_k}}{\sum_{j=1}^{3} e^{z_j}}$$

With $k \in$ {Normal,Benign,Malignant}.

Stage IX — Uncertainty quantification

This is an important addition because the manuscript identifies uncertainty quantification as comparatively underdeveloped.

For T stochastic model predictions,

$p_t = P(y = 1|X, \Theta_t), \quad t = 1, \ldots, T.$

The mean prediction is $\bar{p} = \frac{1}{T}\sum_{t=1}^{T} p_t$.

Prediction variance is $U = \frac{1}{T}\sum_{t=1}^{T}(p_t - \bar{p})^2$.

Thus, $\boxed{U(X) = \mathrm{Var}[P(y|X)]}$ can be used as an uncertainty indicator. A confidence score can be defined as $C(X) = 1 - U(X)$.

Cases satisfying $C(X) < \tau_U$ can be flagged for expert/radiologist review rather than being treated as high-confidence automated decisions. This directly operationalizes the manuscript's proposed uncertainty-guided triage concept.

Stage X — Explainable AI

For the final prediction $f(X, Z_R, Z_C)$, an explanation function can be defined as $E = \Psi(X, Z_R, Z_C; f)$.

For image-based explanation, a saliency/activation map is

$$H(x, y, z) = \left|\frac{\partial f}{\partial X(x, y, z)}\right|.$$

After normalization, $H_N = \frac{H - H_{\min}}{H_{\max} - H_{\min}}$. The explanation should ideally concentrate on clinically meaningful nodule regions. For radiomic/clinical features, SHAP-type contribution can be represented as $f(x) = \phi_0 + \sum_{j=1}^{M} \phi_j$

where

- ϕ_0 = baseline prediction,
- ϕ_j = contribution of feature j.

Unified mathematical objective

The entire proposed framework can therefore be expressed compactly as

$$\boxed{\hat{y} = F_\Theta\left[\sum_{k=1}^{K} \alpha_k\, \Phi_k(R(X_P)),\ \Phi_R(R(X_P)),\ \Phi_C(C)\right]}$$

Where

$$X_P = P(X),$$
$$R(X_P) = X_P \odot f_S(X_P),$$
$$\alpha_k = \frac{\exp(e_k)}{\sum_j \exp(e_j)},$$

And $Z_F = \beta_D Z_A + \beta_R Z_R + \beta_C Z_C$.

The final probability is $\boxed{p_M = \sigma(W_F Z_F + b_F)}$

and the complete prediction is $\boxed{\hat{y} = \mathbb{I}[p_M \geq \tau_c].}$

Training objective

A practical total loss for the proposed framework can be formulated as

$$\boxed{\mathcal{L} = \lambda_1 \mathcal{L}_{cls} + \lambda_2 \mathcal{L}_{seg} + \lambda_3 \mathcal{L}_{att} + \lambda_4 \mathcal{L}_{cal} + \lambda_5 \| \Theta \|_2^2}$$

where: Classification loss

$$\mathcal{L}_{cls} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log p_i + (1 - y_i)\log(1 - p_i)].$$

Segmentation loss

A Dice loss can be used: $\mathcal{L}_{seg} = 1 - \frac{2\sum_x M_x \widehat{M}_x + \epsilon}{\sum_x M_x + \sum_x \widehat{M}_x + \epsilon}$.

Calibration loss

A calibration penalty can be introduced as $\mathcal{L}_{cal} = (\mathrm{ECE})^2$. The manuscript explicitly discusses calibration and gives an example target of ECE < 0.08, although this should not be presented as a universally established threshold for the proposed model.

Clinical decision layer

The final framework should not simply output *malignant/benign*. It should output:

$$O(X) = [p_M,\ 1 - p_M,\ U,\ E]$$

where

- p_M = malignancy probability,
- $1 - p_M$ = benign probability,
- U = predictive uncertainty,
- E = explanation.

A decision policy can then be defined as

$$D(X) = \begin{cases} \text{Expert review}, & U > \tau_U \\ \text{Malignant/high-risk}, & U \le \tau_U,\ p_M \ge \tau_M \\ \text{Benign/low-risk}, & U \le \tau_U,\ p_M < \tau_M. \end{cases}$$

This is more defensible than forcing every CT case into an automated binary decision.

Proposed framework in one mathematical flow

The complete mathematical architecture can be summarized as:

$$X \xrightarrow{P} X_P \xrightarrow{S} R$$

$$R \to \begin{cases} \Phi_1(R) \\ \Phi_2(R) \\ \vdots \\ \Phi_K(R) \end{cases} \to Z_A$$

$$R \to Z_R, \qquad C \to Z_C$$

$$(Z_A, Z_R, Z_C) \xrightarrow{\text{Multimodal Fusion}} Z_F \xrightarrow{\text{Classifier}} p_M \xrightarrow{\text{Uncertainty}} (U, C) \xrightarrow{\text{XAI}} E$$

Finally, CT → Preprocessing → ROI → CNN Ensemble → Attention Fusion → Radiomics → Clinical Fusion → Classification → Uncertainty → XAI → Clinical Validation

Table 2. Representative studies plotted on stages of the pipeline.

Study	Method	Dataset(s)	Key result
UrRehman et al. [3]	Dual channel-spatial attention CNN	LIDC-IDRI, LUNA16	Fewer false positives vs. plain CNN
Lin et al. [5]	3D CNN + radiomics + clinical (stacked ensemble)	In-house chest CT cohort	Outperformed single-modality models
Deng et al. [10]	Multi-feature attention fusion	Multi-center GGN cohort (3 sites)	Validated on two external test sets
Khademi et al. [8]	Swin-Transformer + CAE + 3D radiomics (I-VISTA)	In-house subsolid-nodule cohort	AUC 0.93, Acc. 93.9%
Asmara et al. [1]	Meta-analysis, 21 externally tested models	Multiple external CT cohorts	Pooled sensitivity 88%, specificity 75%

Incorporating radiomics with clinical data and multimodal classification. A parallel body of work complements CNN with IBSI-compliant radiomic descriptors and with clinical variables like age, sex, smoking history and laboratory biomarkers. Combining deep, radiomic and clinical features in tri-modal stacked ensembles has shown higher accuracy for benign/malignant classification and pathological subtyping compared to using single-modality baselines [5]. For heterogeneous modalities, late fusion (modality independent processing followed by combination) is usually preferred over early fusion, but there are limited comparisons of fusion strategies within cohorts, and not always IBSI verification of radiomic extraction. A summary of representative studies of these stages of pipelines is shown in Table 2.

Uncertainty, Explainability, and Clinical Validation:

Calibration and uncertainty quantification are no longer just a must: it has clinical implications and is a malignancy call. Calibration (expected calibration error < 0.08) for ensembled decomposition into aleatoric and epistemic uncertainty is successful in neighboring tasks in thoracic imaging, and uncertainty guided triage is proposed to send out low-confidence cases for expert review, but explicit reporting of calibration is still not common for nodule classifiers in particular. For the activations of malignant lesions, the explainability methods employed are the dominant, and in their most common use, SHAP and its variants are applied to the radiomic and clinical feature streams, with malignant lesions generally covering larger, more peripheral areas than benign ones are. But such heat maps are hardly ever evaluated for quantitative faithfulness or stability, or structured concordance studies with radiologists. During validation phase, a meta-analysis of 21 externally tested classifiers revealed pooled sensitivity of 88% but pooled specificity of only 75% [1] and an associated scoping review indicated that none of the studies reviewed assessed downstream clinical outcomes such as change in management or cost-effectiveness.

Discussions:

This stage-wise synthesis reveals three gaps that run across the synthesis. A first generalization problem: models trained on LIDC-IDRI/LUNA16 and single-institution datasets tend to have poor performance on external data, however, external validation is uncommon. Second, a validation-depth gap: ever more complex fusion and attention mechanisms are suggested, but their assessment is done with less rigor, as attention maps are not usually compared with radiologist annotations and fusion strategies are not usually benchmarked head-to-head and uncertainty metrics are not reported. Third, a translation gap: few studies relate performance to a decision analytic or outcome measure, for example with respect to net benefit or change in management, and thus provide limited evidence basis for clinical adoption, compared to other AI-assisted oncology workflows that are already seeing advances in precision diagnosis and prognosis [6].

These gaps drive five priorities to motivate. Patient privacy should be maintained to expose models to scanner and protocol diversity outside LIDC-IDRI/LUNA16 along with demographic diversity in federated, multi-institutional consortia. Preprocessing pipelines need to be standardized, IBSI-compliant, and provided with a level of detail that would allow the pipelines to be exactly reproduced, including explicit HU normalization and sampling pipelines. Benchmarking the fusion strategies and ensembles should only be done within the same cohort and compared to models with parameters matching the one(s) of the model used in the training of the ensemble. Calibration and uncertainty indicators should be routinely reported along with discrimination indicators; uncertainty-driven triage for clinical integration. Lastly, explainability should be validated with prospective validation based on outcome measures (decision curve analysis), and tested for its faithfulness and stability. From related machine-learning studies of lung-cancer risk stratification based on non-imaging behavioral and hematological parameters, where multiple classifiers are evaluated using hybrid and ensemble designs [7] comes some complementary evidence that rigorous, multi-classifier benchmarking practices already prevalent outside the CNN imaging literature provide a template that pulmonary nodule classification research has yet to fully embrace.

Conclusions:

CNN-based pulmonary nodule classification is key to the ability to identify lung cancer using lungs CT images. Using a nine stage conceptual pipeline, inductively developed from reviewed literature, was employed to map the curation of datasets from clinical validation, and to grade the evidentiary maturity of each stage. There are relatively few and well-developed studies on backbone architectures, ensembling, and attention-based fusion.

References

[1] O. D. Asmara, E. G. M. Steenhuis, K. de Jong, et al., "Externally tested AI models for malignancy classification of lung nodules at CT: a systematic review and meta-analysis," Radiol. Artif. Intell., vol. 8, no. 4, art. no. e250331, 2026. https://doi.org/10.1148/ryai.250331

[2] S. Arun, M. Grosheva, M. Kosenko, J. L. Robertus, O. Blyuss, R. Gabe, D. Munblit, and J. Offman, "Systematic scoping review of external validation studies of AI pathology models for lung cancer diagnosis," NPJ Precis. Oncol., 2025. https://doi.org/10.1038/s41698-025-00940-7

[3] UrRehman Z, Qiang Y, Wang L, Shi Y, Yang Q, Khattak SU, Aftab R, Zhao J. Effective lung nodule detection using deep CNN with dual attention mechanisms. Scientific Reports. 2024;14:3934. https://doi.org/10.1038/s41598-024-51833-x

[4] C. Gao et al., "Deep learning in pulmonary nodule detection and segmentation: a systematic review," Eur. Radiol., vol. 35, no. 1, pp. 255-266, 2025. https://doi.org/10.1007/s00330-024-10907-0

[5] C.-Y. Lin, S.-M. Guo, J.-J. J. Lien, W.-T. Lin, Y.-S. Liu, C.-H. Lai, I.-L. Hsu, C.-C. Chang, and Y.-L. Tseng, "Combined model integrating deep learning, radiomics, and clinical data to classify lung nodules at chest CT," Radiol. Med., vol. 129, no. 1, pp. 56-69, Jan. 2024. https://doi.org/10.1007/s11547-023-01730-6

[6] Kumar V, Prabha C, Sharma P, Mittal N, Askar SS, Abouhawwash M. Unified deep learning models for enhanced lung cancer prediction with ResNet-50–101 and EfficientNet-B3 using DICOM images. BMC medical imaging. 2024 Mar 18;24(1):63. https://doi.org/10.1186/s12880-024-01241-4

[7] V. Kumar, C. Prabha, D. Gupta, S. Juneja, S. Kumari, and A. Nauman, "Multi-model machine learning framework for lung cancer risk prediction: a comparative analysis of nine classifiers with hybrid and ensemble approaches using behavioral and hematological parameters," SLAS Technol., vol. 33, art. no. 100314, Aug. 2025. https://doi.org/10.1016/j.slast.2025.100314

[8] Kumar V, Bakariya B. Classification of malignant lung cancer using deep learning. Journal of Medical Engineering & Technology. 2021 Feb 17;45(2):85-93. https://doi.org/10.1080/03091902.2020.1853837

[9] G. C. Forte, S. Altmayer, R. F. Silva, M. T. Stefani, L. L. Libermann, C. C. Cavion, A. Youssef, R. Forghani, J. King, T. L. Mohamed, R. G. F. Andrade, and B. Hochhegger, "Deep learning algorithms for diagnosis of lung cancer: a systematic review and meta-analysis," Cancers, vol. 14, no. 16, art. no. 3856, Aug. 2022. https://doi.org/10.3390/cancers14163856

[10] H. Deng, W. Huang, X. Zhou, T. Zhou, L. Fan, and S. Liu, "Prediction of benign and malignant ground glass pulmonary nodules based on multi-feature fusion of attention mechanism," Front. Oncol., vol. 14, art. no. 1447132, Oct. 2024. https://doi.org/10.3389/fonc.2024.1447132