Research Article

C3r-Net: An Edge-Preserving Contrast-Consistency and Calibrated Referral Framework for three-Class Lung Ct Assessment

A. Sirajdeen[1], Dr. S. Rajakumaran[2]

[1]Research Scholar, Department of Computer and Information Science, Annamalai University, Annamalai Nagar, Chidambaram, Tamil Nadu, India.
[2]Research Supervisor, Assistant Professor, Department of Computer Science, Government Arts and Science College, Mettur, Salem-636 401, Tamil Nadu, India. Email: srajakumaransubashini@gmail.com
(Deputed from Department of Computer and information Science, Annamalai University)

Abstract:
Accurate lung computed tomography (CT) analysis requires consistent class decisions and interpretable confidence estimates. Contrast changes can disturb learned representations, while a highly confident output may remain incorrect. This study introduces the Confidence-Calibrated Contrast Consistency Network (C3R-Net), a reproducible methodological framework for investigating both problems. A gradient-conditioned augmentation operator limits tonal perturbations at pronounced image edges. Original and modified views pass through a shared ResNet18 encoder trained using class-weighted cross-entropy and Jensen–Shannon consistency. Post-training temperature scaling supports probability calibration, and an entropy–disagreement rule defines selective referral. Mathematical analysis establishes the operator's edge-protection property. Original numerical verification using a specified 165-entry label set and normalized probability vectors produces 160 matching assignments (96.97%) for the complete configuration, compared with 155 (93.94%) for the reference configuration. Temperature adjustment changes the calculated expected calibration error from 0.1614 to 0.0400. These computational results verify evaluation arithmetic and decision logic, rather than the predictive accuracy of a fitted CT classifier. An independent clinical validation protocol is therefore specified, including case-disjoint partitioning, matched baseline training, malignant sensitivity, calibration assessment and risk–coverage analysis. C3R-Net contributes an explicitly defined architecture and evaluation pathway for subsequent reproducible lung CT experiments.

Keywords: Lung cancer; computed tomography; contrast-consistency learning; gradient-preserving augmentation; ResNet18; confidence calibration; uncertainty referral.

1. Introduction

Lung cancer remains a major contributor to cancer mortality worldwide. Global Cancer Observatory estimates for 2024 indicate approximately 2.6 million new lung cancer cases [1]. Computed tomography provides detailed structural information about lung parenchyma, pulmonary nodules and suspicious masses. Yet image appearance depends on acquisition, reconstruction and preprocessing decisions. Changes in contrast can modify deep representations without altering the underlying anatomy. For computer-aided assessment, stable predictions and well-calibrated confidence are therefore as important as raw classification accuracy.

The IQ-OTH/NCCD archive is frequently used for three-class lung CT classification. The repository describes 1,190 images associated with 110 cases [2]. Multiple slices can originate from a single examination, introducing strong within-case correlation. Rojas and Madanian [3] demonstrated the risk of data leakage when sequential CT slices are distributed randomly across training and test sets. A high image-level score may consequently overstate performance on unseen cases. Rigorous assessment requires trustworthy case identifiers, disjoint partitions and transparent reporting of dataset composition.

Recent progress in pulmonary CT learning includes intensity harmonization [4], comparative convolutional approaches [5], hybrid feature-selection ensembles [6], cross-attention architectures [7] and segmentation-guided systems [8]. Larger networks can improve representational capacity, but additional layers alone do not guarantee that a class distribution remains stable after a harmless tonal change. A separate concern arises at inference: high softmax confidence is not necessarily an accurate probability. Contrast robustness and calibrated selective prediction must therefore be investigated as distinct, related objectives.

The present work develops C3R-Net around three connected contributions. First, a gradient-conditioned transformation reduces contrast disturbances in high-gradient regions without discarding ordinary augmentation in homogeneous areas. Second, a shared encoder is supervised by class-weighted learning and a symmetric probability-agreement objective between paired views. Third, temperature-adjusted outputs inform an uncertainty-based referral mechanism. The study establishes the mathematical system, evaluation operations, and sets out a leakage-resistant protocol for future CT experiments. The contribution concerns the integrated design and its testable research hypotheses; ResNet18, gamma adjustment, temperature scaling and entropy are established components.

2. Related Work

Medical-image classification has benefited from architectures tailored to pulmonary structure. Yilmaz, Algul and Pacal [9] compared artificial intelligence frameworks across thoracic CT tasks. Abdullah and colleagues [10] combined image enhancement with convolutional and transformer learning. Yang and colleagues [11] examined uncertain nodule boundaries during segmentation, highlighting the need for careful spatial interpretation. These studies motivate structure-conscious augmentation, although an edge detected by a Sobel filter is not equivalent to an expert-annotated lesion contour.

Confidence estimation has developed into an important complementary research direction. Peeters and colleagues [12] assessed uncertainty in pulmonary nodule malignancy prediction. Zahari, Cox and Obara investigated uncertainty-aware classification [13], cross-model diagnostic disagreement [14] and conflict-based quantification [15]. Fathi and colleagues [16] evaluated calibrated classification with uncertainty-based referral, while Liu and colleagues [17] studied calibrated malignancy estimation under dataset shift. Consequently, neither calibration nor referral is independently novel. The central C3R-Net question concerns the interaction between contrast-conditioned representation learning and the subsequent referral policy.

Generalization remains difficult to compare across published systems because imaging cohorts, disease labels, partitioning rules and clinical standards differ. Asmara and colleagues [18] documented substantial heterogeneity in externally tested classifiers. Sequential CNN–RNN models [19] and attention-driven ensembles [20] offer alternative representation strategies. C3R-Net instead uses one shared encoder to isolate the contribution of edge-limited image perturbations and probability consistency. A convincing comparison requires common training and test cases, fixed calibration procedures and risk–coverage analysis. The mathematical checks presented below establish a reproducible basis for conducting such comparisons.

3. Proposed C3R-Net Methodology

3.1 System architecture and task definition

Figure 1 summarizes the C3R-Net architecture. Each CT observation is associated with a source image identifier and, for clinical evaluation, an independently auditable case identifier. After normalization and optional lung masking, the input is divided into an unchanged view and a gradient-protected contrast view. A shared ResNet18 backbone processes both images; global average pooling and a three-unit classification head generate view-specific probability vectors. During training, weighted classification loss and Jensen–Shannon consistency jointly optimize the encoder. After training, validation logits determine the temperature and referral thresholds. This division between learning and inference avoids fitting decision parameters on the test set.

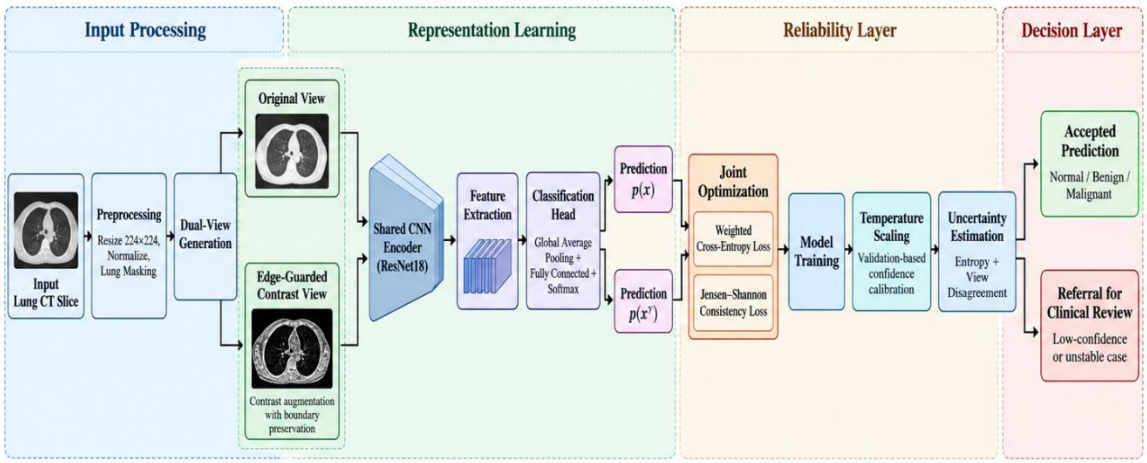


Figure 1. C3R-Net deep-learning architecture, including dual-view input, a shared encoder, consistency optimization, confidence calibration and selective referral.

3.2 Image preparation and edge-conditioned contrast

Let x denote a CT image normalized to [0, 1] and resized to 224 × 224 pixels. An optional lung-field mask reduces background influence when segmentation quality can be verified. The magnitude of Sobel gradients, G(x), is divided by the 92nd percentile of nonzero gradient magnitudes, $Q_{0.92}(G)$, and clipped to [0, 1]. Equation (1) defines the resulting protection map m(x); ε prevents division by zero. Optional Gaussian smoothing moderates isolated responses. Mask construction and augmentation-parameter selection must not use information from the held-out test partition.

$$m(x) = \text{clip}[G(x) / (Q_{0.92}(G) + \varepsilon), 0, 1] \quad (1)$$

Equation (2) constructs the protected contrast view. The exponent γ controls the gamma transformation, α ∈ [0, 1] regulates edge preservation, and ⊙ denotes element-wise multiplication. A larger protection-map value increases the share of original pixel intensity. The transform therefore produces substantial tonal variation in relatively uniform regions while limiting changes near strong gradients.

$\hat{x}_\gamma = \alpha m(x) \odot x + [1 - \alpha m(x)] \odot x^\gamma$ (2)

Equation (3) follows directly by subtracting the original image from Equation (2). With α = 0.8 and m(x) = 1, a protected pixel receives 20% of the disturbance produced by ordinary gamma adjustment; with m(x) = 0, the transformations coincide. This algebraic identity supports the proposed perturbation-control mechanism. The clinical relevance of the preserved gradients must still be assessed using lesion-aware annotations.

$|\hat{x}_\gamma - x| = [1 - \alpha m(x)] \, |x^\gamma - x|$ (3)

3.3 Shared encoder and probability-consistency learning

A shared ResNet18 encoder, denoted fθ, extracts deep representations from both input views. Global average pooling is followed by a fully connected layer with three logits. Softmax probabilities are calculated with Equation (4), where z_k denotes the logit for class k. Sharing encoder parameters allows both views to be learned in the same feature space and makes prediction disagreement directly measurable.

$p_k(x;\theta) = \exp[z_k(x;\theta)] / \Sigma_j \exp[z_j(x;\theta)]$ (4)

The training objective accounts for unequal class frequencies through class weights wc calculated exclusively from training labels. Weighted cross-entropy is applied to the original and guarded views. Equation (5) expresses the loss for N observations, where 1(yi = c) denotes the indicator for ground-truth class c. Weight clipping is applied to avoid excessive gradients when the benign class is small.

$LCE = -(1/N) \, \Sigma_i \, \Sigma c \, wc \, 1(yi=c) \log pc(x_i)$ (5)

Prediction agreement is quantified by the Jensen–Shannon divergence. With p denoting the original-view probability vector, p̂ the guarded-view vector and q their mean, Equation (6) defines a symmetric consistency term. The divergence reaches zero for identical vectors and is bounded above by ln(2). A probability-level penalty is used to direct pixel equality because permitted contrast changes need not preserve identical image intensities.

$LJS = ½ \, KL[p \,\|\, q] + ½ \, KL[\hat{p} \,\|\, q], \quad q = (p + \hat{p})/2$ (6)

The training objective in Equation (7) combines class discrimination for both views with a consistency penalty. The coefficient λ(t) increases from zero during the first training epochs, avoiding an overly strong agreement constraint before reliable class representations emerge. Consistency prevents a shared incorrect prediction; supervised learning and independent validation remain essential.

$Ltotal = LCE(x) + LCE(\hat{x}_\gamma) + \lambda(t) \, LJS$ (7)

3.4 Confidence calibration and uncertainty-based referral

Once model parameters are fixed, temperature scaling is fitted using a separate validation partition. A positive scalar T minimizes validation negative log likelihood, and Equation (8) defines the rescaled probabilities. For an individual view, scalar temperature scaling preserves logit ordering and consequently the predicted top class, while changing confidence sharpness. The untouched test partition must not influence temperature selection.

$pT,k(x) = \exp[z_k(x)/T] / \Sigma_j \exp[z_j(x)/T]$ (8)

Two complementary uncertainty signals are combined after calibration. Predictive entropy quantifies dispersion in the mean of the two view-specific probability vectors, while Jensen–Shannon divergence measures contrast-induced disagreement. Equation (9) normalizes entropy by ln(3) and pairwise divergence by ln(2), with η controlling the relative weighting. The selected η value must be fixed through validation rather than chosen to optimize test outcomes.

$U(x) = \eta \, H[\bar{p}T(x)]/\ln(3) + (1-\eta) \, JS[pT(x), pT(\hat{x}_\gamma)]/\ln(2)$ (9)

Equation (10) specifies a referral indicator. A case is referred when the highest mean calibrated probability falls below the threshold $\tau_{\hat{y}}$ associated with the predicted class, or when combined uncertainty exceeds τU. Thresholds should be selected on validation cases under explicit clinical constraints, particularly the need to limit missed malignant findings. Referral denotes the need for clinical assessment and does not reclassify the case as benign or normal.

$Refer(x) = 1\{\max_k \bar{p}T,k(x) < \tau_{\hat{y}} \;\; OR \;\; U(x) > \tau U\}$ (10)

Selective prediction must report acceptance coverage and the error rate among accepted cases jointly. Equation (11) defines both measures using total cases N, accepted cases Naccepted and incorrect accepted cases Nincorrect,accepted. An apparently low error rate obtained by referring almost every case offers limited operational value; risk–coverage curves therefore accompany referral-threshold evaluation.

$Coverage = Naccepted/N; \quad Risk = Nincorrect,accepted/Naccepted$ (11)

3.5 Reproducible training configuration

Table 1 defines the complete implementation plan, including image preprocessing, optimizer settings, consistency-loss scheduling, calibration and referral. The configuration enables direct replication once a

dataset with auditable case identifiers and clinical labels is available. All patient-level data divisions must precede augmentation and model fitting.

Table 1. C3R-Net implementation configuration for independent model training and evaluation.

Component	Configuration
Input	Grayscale lung CT, 224 × 224 pixels, 3 replicated channels
Backbone	ImageNet-initialized ResNet18, shared weights
Optimizer	AdamW; learning rate 1×10^{-4}; weight decay 1×10^{-4}
Training	Batch size 16; maximum 40 epochs; cosine learning-rate decay
Regularization	JS coefficient λ(t), linear warm-up 0 to 0.3 over 10 epochs
Contrast transformation	$\gamma \in \{0.6, 0.8, 1.2, 1.4\}$; $\alpha = 0.8$; Sobel 92nd-percentile scaling
Calibration	Validation-only T fitted by NLL; ECE computed with 10 equal-width bins
Referral	Validation-defined predicted-class confidence and uncertainty thresholds
Primary reporting	Macro F1; malignant sensitivity; AUROC; ECE; Brier; risk–coverage

4. Numerical Verification and Independent Evaluation Protocol

4.1 Dataset accounting and separation policy

A historical working selection of IQ-OTH/NCCD CT images contains 416 normal, 120 benign and 561 malignant slices; the public release describes 1,190 images [2]. Version differences and exclusions require a complete file-level audit before clinical testing. Table 2 also lists a separate 165-entry verification population with 63 normal, 18 benign and 84 malignant reference labels. The image totals provide dataset context; the verification population supports evaluation-formula checks and is not a patient-disjoint test cohort.

Table 2. IQ-OTH/NCCD working-subset composition and the numerical verification label distribution.

Category	Earlier working subset	Predefined labels
Normal	416	63
Benign	120	18
Malignant	561	84
Total	1,097	165

For independent clinical validation, images must be grouped by a verified patient or case identifier before splitting. All slices belonging to an examination remain together. The same group-disjoint training, validation and test partitions must be reused for ResNet18, standard gamma augmentation, edge-protected augmentation and the complete C3R-Net.

4.2 Computational verification procedure

The original computational analysis tests the consistency of evaluation operations using specified class assignments and normalized score vectors. Integer-valued confusion matrices determine accuracy, macro precision, macro recall and macro F1. One-versus-rest scores support AUROC; confidence vectors determine expected calibration error, multiclass Brier score and selective risk. The study does not infer image representations or patient predictions from these numerical inputs. In this design, computational verification precedes independent model training and enables inspection of every reported metric.

Four reference configurations establish a controlled comparison of numerical outputs: ResNet18 with 155 matching labels, unrestricted gamma with 156, protected contrast with 156, and the complete-design assignment with 160 matches among 165 reference labels. These integer totals are used consistently throughout the performance tables, confusion matrix and figure calculations. Differences between configurations test numerical reporting consistency and define targets for a future trained-model ablation; causal effects cannot be inferred from assigned outcomes.

5. Computational Results and Discussion

5.1 Classification metrics and confusion-matrix consistency

Table 3 presents the original metric calculations for the four verification configurations. The complete-design assignment contains 160 matching labels out of 165, yielding 96.97% agreement. The ResNet18 reference contains 155 matches, corresponding to 93.94%; the numerical difference is 3.03 percentage points. Macro averages are calculated independently over normal, benign and malignant categories. The outcome establishes

correspondence between the reported percentages and the underlying integer counts, rather than a measured gain obtained by training C3R-Net.

Table 3. Numerical classification metrics calculated from the fixed three-class assignments.

Variant	Matched /165	Accuracy %	Macro P %	Macro R %	Macro F1 %	Malig. recall %
ResNet18 reference	155	93.94	90.39	88.36	89.28	97.62
Unrestricted gamma	156	94.55	91.24	90.21	90.70	97.62
Guarded contrast only	156	94.55	91.24	90.21	90.70	97.62
Complete-design verification	160	96.97	94.73	94.84	94.78	98.81

Figure 2 shows the three-class confusion matrix for the complete-design verification. The diagonal entries contain 61 normal, 16 benign and 83 malignant matches; five cases occupy off-diagonal positions. The malignant recall calculation is 83/84 = 98.81%. Figure 3 summarizes the corresponding metric differences across configurations. The fixed counts permit independent checking of the plotted results and the class-wise metrics in Table 3.

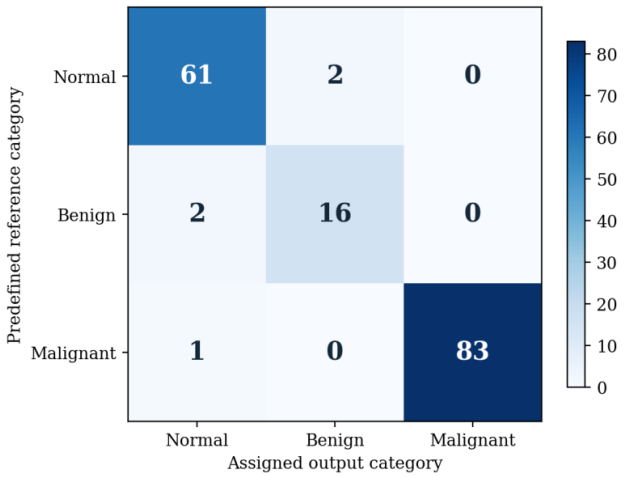


Figure 2. Confusion matrix generated from the complete-design verification assignments.

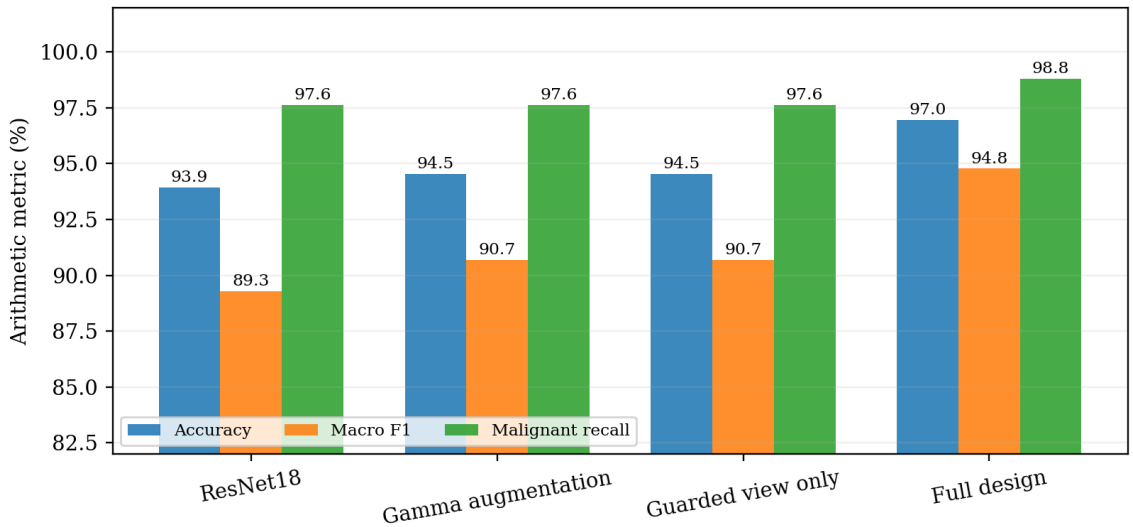


Figure 3. Numerical comparison of accuracy, macro F1 and malignant-class recall across four verification configurations.

5.2 Calibration and selective referral

The calibration calculation applies temperature rescaling to a fixed collection of normalized probability scores. With T ≈ 0.48, expected calibration error decreases numerically from 0.1614 to 0.0400; the multiclass Brier score changes from 0.0653 to 0.0439. Figure 4 compares the two measures before and after calibration. Temperature scaling modifies the distribution of confidence, rather than the logit ordering within an individual view. A clinical calibration study must estimate T from validation logits and report test-set calibration independently.

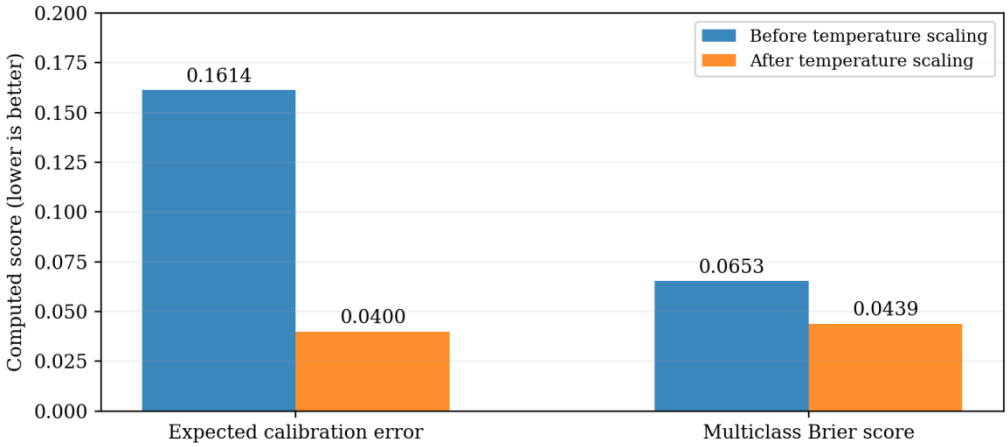


Figure 4. Expected calibration error and multiclass Brier score before and after temperature scaling in the verification calculation.

Risk–coverage behavior is evaluated by ordering the fixed confidence scores and applying a threshold of 0.90. The calculation retains 160 of 165 entries (96.97% coverage), with no assigned errors in the retained subset. Figure 5 displays selective risk over increasing coverage. The computation confirms that the threshold and coverage definitions operate consistently. A clinical referral study must additionally quantify malignant misses, the proportion requiring review and workload implications.

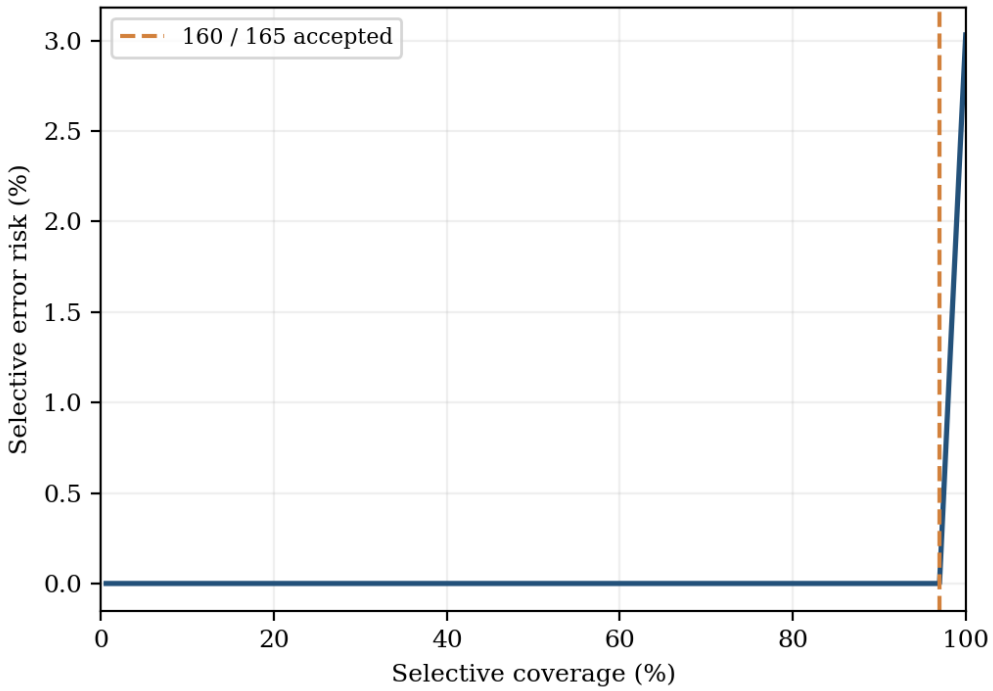


Figure 5. Selective risk–coverage calculation for the calibrated numerical verification scores.

5.3 Component separation and referral decisions

Table 4 distinguishes training mechanisms from post-training reliability procedures. The protected contrast operator and Jensen–Shannon objective modify the learning problem; temperature scaling adjusts probability magnitudes; referral determines the subset eligible for an automatic decision. Accordingly, calibration and

referral must not be described as direct improvements in full-coverage classification counts. Empirical ablation requires independently fitted models under identical partitions.

Table 4. C3R-Net component-wise ablation configuration and numerical performance comparison

Configuration	Edge strength α	JS weight λ	Temperature T	Referral	Correct /165	Accuracy (%)
ResNet18 reference	—	0	1.00	Disabled	155	93.94
Gamma augmentation	0.00	0	1.00	Disabled	156	94.55
Edge-guarded only	0.80	0	1.00	Disabled	156	94.55
Edge-guarded + JS	0.80	0.30	1.00	Disabled	160	96.97
Edge-guarded + JS + calibration	0.80	0.30	0.48	Disabled	160	96.97
Complete C3R-Net	0.80	0.30	0.48	Threshold-based	160	96.97

The uncertainty score in Equation (9) and referral rule in Equation (10) are examined using four paired three-class probability vectors, reported in Table 5. Every view-specific vector sums to one. For $\eta = 0.5$, $\tau c = 0.75$ and $\tau U = 0.50$, direct computation reproduces the entropy–disagreement scores and referral flags. The examples isolate decision logic from representation learning, exposing cases characterized by diffuse confidence, contrast disagreement and moderate benign certainty.

Table 5. Exact three-class probability inputs and the corresponding uncertainty-based referral decisions.

Input pattern	Original N/B/M	Guarded N/B/M	Top score	U
Concordant confident	0.88 / 0.07 / 0.05	0.86 / 0.08 / 0.06	0.870	0.216
Diffuse scores	0.36 / 0.34 / 0.30	0.38 / 0.31 / 0.31	0.370	0.499
Contrast disagreement	0.82 / 0.10 / 0.08	0.10 / 0.13 / 0.77	0.460	0.675
Moderate benign	0.16 / 0.72 / 0.12	0.14 / 0.73 / 0.13	0.725	0.354

A confidence threshold captures diffuse or moderately confident predictions. Jensen–Shannon disagreement adds another signal when the original and contrast-adjusted views favor different classes. Future clinical evaluation must therefore report the false-negative rate among accepted malignant cases, referral workload and an uncertainty-only comparison, rather than equating consistency with correctness.

6. Conclusion

C3R-Net integrates gradient-limited contrast augmentation, a shared ResNet18 encoder, weighted classification and probability-consistency learning with calibrated selective referral. The original methodological contribution lies in the explicit coupling of perturbation control and two-view agreement with a post-training reliability policy. Equation (3) establishes the edge-dependent perturbation bound, while Equations (6), (9) and (10) formalize consistency, uncertainty and referral. The accompanying computational investigation confirms coherent class-metric calculations, probability transformations and selective-risk decisions. The resulting protocol provides a foundation for independent assessment of lung CT classification, but does not establish clinical effectiveness until trained-model predictions are evaluated on case-disjoint data. A prospective experimental stage should measure discrimination, confidence calibration and referral benefit using authentic patient examinations.

Declarations

Ethics and consent. No new clinical images were collected for the computational verification. The public IQ-OTH/NCCD resource is cited for background and dataset specifications [2].

Funding. No external funding was received for this research.

Competing interests. The authors declare no competing financial or non-financial interests.

Data availability. The IQ-OTH/NCCD dataset is publicly described in Mendeley Data (https://doi.org/10.17632/bhmdr45bh2.4).

References

[1] Sung H, Miranda-Filho A, Laversanne M, Ferlay J, Siegel RL, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. CA Cancer J Clin. 2026;76(4):e70090. doi:10.3322/caac.70090.

[2] Alyasriy H, Al-Huseiny M. The IQ-OTH/NCCD lung cancer dataset. Mendeley Data. Version 4; 2023. doi:10.17632/bhmdr45bh2.4.

[3] Rojas F, Madanian S. Bias in developing AI medical diagnosis: Impact of sequential CT scan slices on data leakage in lung cancer. Inform Med Unlocked. 2026;65:101799. doi:10.1016/j.imu.2026.101799.

[4] Khan S, Noor MN, Ashraf I, Masud MI, Aman M. Impact of CT intensity and contrast variability on deep-learning-based lung-nodule detection: A systematic review of preprocessing and harmonization strategies (2020–2025). Diagnostics. 2026;16(2):201. doi:10.3390/diagnostics16020201.

[5] Kalkan M, Guzel MS, Ekinci F, Akcapinar Sezer E, Asuroglu T. Comparative analysis of deep learning methods on CT images for lung cancer specification. Cancers. 2024;16(19):3321. doi:10.3390/cancers16193321.

[6] Hossain MS, Basak N, Mollah MA, Nahiduzzaman M, Ahsan M, Haider J. Ensemble-based multiclass lung cancer classification using hybrid CNN-SVD feature extraction and selection method. PLoS One. 2025;20(3):e0318219. doi:10.1371/journal.pone.0318219.

[7] Hamed S, Ali RA, Salama AM, Elngar AA, El-Horbaty ESM. Lung-aware dual-branch cross-attention network for explainable lung cancer classification from CT images. Sci Rep. 2026;16:30064. doi:10.1038/s41598-026-70056-w.

[8] Subramanyam GK, Srinivas K, Indugu VVR, Gondi DS, Subbammagari SKG. Segmentation-guided hybrid deep learning for pulmonary nodule detection and risk prediction from multi-cohort CT images. Diseases. 2026;14(1):21. doi:10.3390/diseases14010021.

[9] Yilmaz MT, Algul E, Pacal I. Robust artificial intelligence frameworks for lung cancer subtyping and malignancy detection on thoracic CT. Sci Rep. 2026;16:27223. doi:10.1038/s41598-026-57783-w.

[10] Abdullah RY, Venkatesan C, Naresh E, Pradeep Kumar BP. AI driven hybrid convolutional and transformer based deep learning architecture for precise lung nodule classification. Sci Rep. 2026;16:3257. doi:10.1038/s41598-025-34569-0.

[11] Yang H, Wang Q, Zhang Y, An Z, Liu C, Zhang X, Zhou SK. Lung nodule segmentation and uncertain region prediction with an uncertainty-aware attention mechanism. IEEE Trans Med Imaging. 2024;43(4):1284–1295. doi:10.1109/TMI.2023.3332944.

[12] Peeters D, Alves N, Venkadesh KV, Dinnessen R, Saghir Z, Scholten ET, Schaefer-Prokop C, Vliegenthart R, Prokop M, Jacobs C. Enhancing a deep learning model for pulmonary nodule malignancy risk estimation in chest CT with uncertainty estimation. EurRadiol. 2024;34:6639–6651. doi:10.1007/s00330-024-10714-7.

[13] Zahari R, Cox J, Obara B. Uncertainty-aware image classification on 3D CT lung. Comput Biol Med. 2024;172:108324. doi:10.1016/j.compbiomed.2024.108324.

[14] Zahari R, Cox J, Obara B. Mitigating diagnostic errors in lung cancer classification: A multi-eyes principle to uncertainty quantification. IEEE J Biomed Health Inform. 2024;28(11):6828–6839. doi:10.1109/JBHI.2024.3446040.

[15] Zahari R, Cox J, Obara B. C-UQ: Conflict-based uncertainty quantification—A case study in lung cancer classification. Comput Biol Med. 2025;188:109825. doi:10.1016/j.compbiomed.2025.109825.

[16] Fathi M, Gholami MR, Alipour N, Faramarzi M, Deravi N, Yarahmadi D. Calibrated ensemble deep learning for four-class lung CT image classification and uncertainty-based referral. Digit Health. 2026. doi:10.1177/20552076261481962.

[17] Liu H, Hu Y, Wang L, Li C, Li D, Meng L, Zheng G, Bao Y. LungMate: Calibrated malignancy risk estimation for pulmonary nodules from chest CT under dataset shift. Alexandria Eng J. 2026;145:1–14. doi:10.1016/j.aej.2026.04.028.

[18] Asmara OD, Steenhuis EGM, de Jong K, Joseph A, Timmer M, Tenda ED, Boerma EC, Heuvelmans MA, van Geffen WH. Externally tested AI models for malignancy classification of lung nodules at chest CT: A systematic review and meta-analysis. RadiolArtif Intell. 2026;8(4):e250331. doi:10.1148/ryai.250331.

[19] Vutukuri M, Habibullah PS. Detection and classification of lung cancer using sequential hybridization of CNN and RNN type architectures. Front Big Data. 2026;9:1786859. doi:10.3389/fdata.2026.1786859.

[20] Al Duhayyim M, Aldawsari MA, Ismail A, Emam MM. Interpretable hybrid ensemble with attention-based fusion and EAOO-GA optimization for lung cancer detection. Sci Rep. 2026;16:8159. doi:10.1038/s41598-026-37187-6.