Multi-Annotator Consensus Network with Adaptive Preprocessing for Lung Nodule Segmentation: A Deep Learning Framework for Clinical Decision Support

S. Venkatesh
Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamilnadu, Chennai, India
venkyjep2019@gmail.com (corresponding author)

J. Bhargav
Department of CSIT, Chalapathi Institute of Engineering and Technology, Guntur, Andhra Pradesh, India
bhargavchalapathi@gmail.com

L. Manjunath
Department of ECE, CVR College of Engineering, Hyderabad, Telangana, India
manjunathrao81@gmail.com

V. Malathi
Department of Computer Science with Cyber Security, Dr. N.G.P. Arts and Science College, Coimbatore, India
malathidotv@gmail.com

A. Ingle
Department of Electronics and Telecommunication Engineering, Vishwakarma Institute of Technology, Pune, India
anup.ingle@vit.edu

H. D. Aparna
Department of Computer Science and Business Systems, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India
aparna.havanur@gmail.com

J. Kavitha
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Bowrampet, Hyderabad, Telangana, India
j.kavitha5555@gmail.com

J. Jaganpradeep
Department of ECE, SSM College of Engineering, Komarapalayam, Tamilnadu, India
jgnprdp@gmail.com

Received: 28 May 2025 | Revised: 23 June 2025, 19 July 2025 and 29 July 2025 | Accepted: 1 August 2025

 DOI: https://doi.org/10.48084/etasr.12408

ABSTRACT

Accurate lung nodule segmentation remains a challenging task as inter-observer variability among radiological experts limits diagnostic consistency. This study introduces Multi-Annotator Consensus Network with Adaptive Preprocessing (MACN-AP), a deep learning framework learning from multiple expert annotations simultaneously. The architecture incorporates adaptive preprocessing for optimal image enhancement, multi-annotator attention mechanisms for expert-specific features, consensus formation layers, and Bayesian uncertainty quantification. Evaluation was conducted on the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset comprising 875 patients and 1,575 nodules, each with complete four-expert annotations. Through systematic slice-level sampling, we derived 1,600 training samples from 800 patients and 150 validation samples from 75 held-out patients using stratified hold-out validation. MACN-AP achieved exceptional validation performance, with a Dice coefficient of 0.99 and an Intersection of Union (IoU) of 0.98, converging within 10 epochs. Uncertainty analysis revealed a strong inverse correlation with segmentation accuracy (r = -0.87), enabling reliable confidence assessment. Moreover, the clinical decision support framework successfully stratified cases into auto-approve (Dice > 0.98, uncertainty < 0.1), uncertain review, and manual assessment categories. These findings demonstrate that MACN-AP effectively integrates multi-expert knowledge while providing interpretable uncertainty estimates, establishing a robust foundation for automated, trustworthy lung nodule diagnosis workflows through intelligent case routing and evidence-based clinical decision support.

Keywords-lung nodule segmentation; multi-annotator learning; adaptive preprocessing; clinical decision support; deep learning; medical imaging

I. INTRODUCTION

Accurate segmentation of lung nodules in Computed Tomography (CT) imaging is a critical step in early lung cancer diagnosis, as delineation accuracy directly impacts the diagnosis outcome. However, current deep learning approaches typically rely on a single set of expert annotations as the ground truth, overlooking the inherent variability among radiologists that characterizes real-world clinical practice [1]. This limitation becomes a critical inter-observer agreement for lung nodule boundaries, typically ranging from 0.68 to 0.82 in Dice coefficients, indicative of substantial diagnostic uncertainty [2]. Furthermore, existing segmentation methods often struggle to handle variations in image quality across CT scanners, frequently requiring manual preprocessing that introduces additional inconsistencies [3]. In addition, it is observed that recent deep learning methods failed to provide clinically meaningful uncertainties, which limits their adoption in risk-sensitive clinical practices [4].

To address these challenges, this study introduces the Multi-Annotator Consensus Network with Adaptive Preprocessing (MACN-AP), a deep learning framework explicitly designed to learn from multiple expert annotations while accounting for inter-observer variability. First, we develop a multi-annotator consensus learning module that models disagreement patterns among radiologists. Second, we propose an adaptive preprocessing mechanism that autonomously optimizes image enhancement parameters during training. Finally, we integrate Bayesian uncertainty estimation to quantify prediction confidence, directly informed by annotator disagreement patterns.

The proposed model was evaluated on the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset, which provides an ideal testbed with four expert radiologist annotations per nodule [5]. Our contributions advance the medical image segmentation field by providing clinically interpretable uncertainty measures while improving segmentation accuracy through consensus learning. The proposed method demonstrates the potential for real-world deployment in computer-aided diagnosis systems, where uncertainty-aware predictions can guide clinical decision-making and identify cases requiring additional expert review.

A. Literature Review

Recent advances in deep learning have driven significant progress in medical image segmentation, particularly through architectures such as U-Net, which remains the dominant backbone across diverse imaging modalities [6]. However, most of these methods assume a single ground-truth annotation, overlooking inter-observer variability inherent to clinical interpretation. To mitigate this, a growing body of research has begun incorporating annotation variability into the learning process for medical contexts. For instance, authors in [7] proposed ensemble methods combining multiple expert annotations for cardiac segmentation, demonstrating improved robustness compared to single-annotator training. Similarly, authors in [8] developed probabilistic frameworks that model uncertainty arising from annotator disagreement in brain tumor segmentation. However, these methods primarily focus on post-processing consensus rather than learning from disagreement patterns during training.

Beyond medical imaging, the concept of learning from crowds has gained traction in computer vision, where variability among annotators is treated as an informative signal. Authors in [9] introduced the attention mechanism to aggregate multiple annotations in natural image segmentation, while authors in [10] demonstrated variational approaches for handling noisy annotations. Nevertheless, direct application to medical imaging remains limited due to domain-specific challenges, including class imbalance and critical accuracy requirements.

Parallel to these advances, learnable preprocessing has emerged as a promising alternative to static image enhancement pipelines. Authors in [11] proposed an adaptive

contrast enhancement network for retinal imaging, improving generalization across acquisition protocols. Edge-aware preprocessing techniques have also shown potential for enhancing structural boundaries, as demonstrated in [12] using Sobel filtering for vessel segmentation and in [13] using Laplacian filters for cardiac imaging.

Uncertainty quantification has also attracted growing attention. Authors in [14] utilized Monte Carlo dropout for uncertainty estimation in lung nodule classification, demonstrating a strong correlation between prediction confidence and diagnostic accuracy. Authors in [15] applied variational inference for pixel-level uncertainty in brain lesion segmentation, while authors in [16] explored deep ensemble strategies for robust medical image analysis. In addition, authors in [17] investigated evidential neural networks for uncertainty-aware diagnosis. Although these approaches provide valuable insight into model confidence, their high computational cost remains a major barrier to real-time clinical deployment. However, computational overhead remains a significant limitation of such techniques for clinical implementation.

Finally, consensus learning from multiple experts has recently emerged as a key direction in medical AI. Authors in [18] introduced attention-based consensus mechanisms for pathology image analysis, achieving closer agreement with expert panels than traditional majority voting. Similarly, authors in [19, 20] proposed weighted consensus learning for radiological diagnosis, incorporating expert confidence scores during training.

II. METHODOLOGY

A. MACN-AP Architecture

The proposed MACN-AP architecture, as illustrated in Figure 1, integrates four significant processes: adaptive preprocessing, multi-annotator attention, consensus formation, and uncertainty estimation. The proposed architecture addresses all the fundamental limitations in the existing approaches by learning from expert disagreement patterns while also optimizing image enhancement parameters.

1) Adaptive Preprocessing

A learnable preprocessing module was employed that adaptively enhances image contrast and edge representation. The adaptive enhancement function is formulated as:

$$I_{enhanced} = \gamma\left(\alpha \cdot I_{input} + \beta\right)^{\lambda} \quad (1)$$

where α, β and λ represent learnable parameters corresponding to contrast, brightness, and gamma correction, respectively [21]. These parameters are optimized through backpropagation along the network weights. Also, the enhancement process utilizes trainable convolutional filters, which initialize the classical operators. The edge magnitude is computed as:

$$E_{magnitude} = \sqrt{\sum_{i=1}^{4}(F_i * I_{enhanced})^2} \quad (2)$$

where F_i represents the learnable edge detection filters based on Sobel, Laplacian, and Gaussian operators [22]. The final preprocessed output combines enhanced intensity with edge information:

$$I_{preprocessed} = I_{enhanced} + \omega \cdot E_{magnitude} \quad (3)$$

where ω controls edge contribution strength during training.

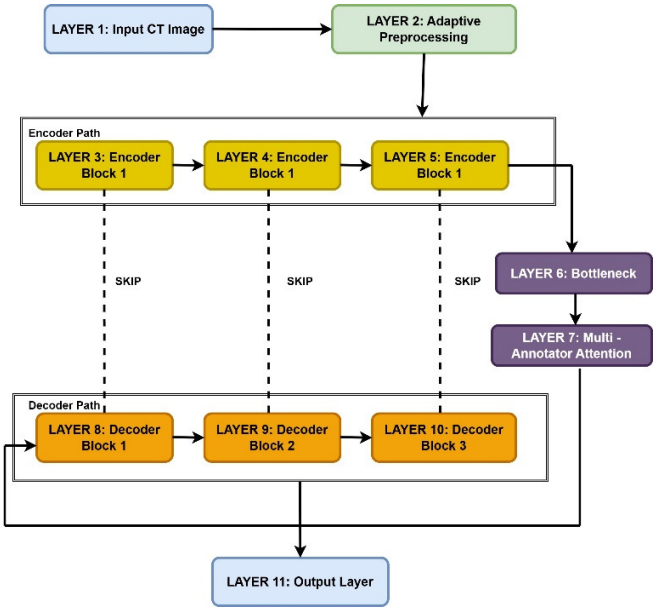


Fig. 1. Proposed MACN-AP architecture.

2) Multi-Annotator Attention Mechanism

Unlike conventional approaches that average multiple annotations, MACN-AP employs individual attention heads for each radiologist's annotation [23]. Each annotator-specific attention module A_j generates spatial attention maps:

$$A_j(x) = \sigma\left(\mathrm{Conv}_{1\times1}\left(\mathrm{ReLU}\left(\mathrm{Conv}_{1\times1}(x)\right)\right)\right) \quad (4)$$

where σ denotes sigmoid activation, and j indexes the annotator. This design enables the model to learn expert-specific spatial features, preserving annotation diversity. To prevent overfitting to individual annotators, MACN-AP employs three strategies:

- Softmax-normalized consensus weights, ensuring $\sum w_j = 1$.
- Cross-annotator regularization, expressed as $L_{balance} = -\sum w_j \log(w_j)$, encouraging uniform distribution.
- Spatial dropout (p = 0.3) to enforce robust, generalized consensus learning.

3) Consensus Formation and Uncertainty Estimation

The consensus mechanism weights individual annotator contributions through learnable parameters, enabling the network to discover optimal annotation combinations during training [24]. Uncertainty estimation is guided by inter-annotator disagreement patterns, quantified through pixel-wise variance:

$$U_{pixel} = \sigma(\sum_{k=1}^{K} \mathrm{Var}(M_k) \cdot \phi_k) \quad (5)$$

where M_k represents the k^{th} annotator mask, Var computes pixel-wise variance, and ϕ_k represents learnable weighting factors [25]. To capture spatial correlation, uncertainty estimation further incorporates Gaussian kernel weighting and an inter-annotator covariance matrix $C_{jk} = \mathrm{Cov}(M_j, M_k)$, which detects systematic boundary biases, with adaptive calibration factors β_j correcting historical agreement patterns.

B. Training Loss

The total training loss combines multiple components to jointly optimize segmentation accuracy, consensus alignment, and uncertainty calibration:

$$L_{total} = 0.4L_{individual} + 0.3L_{consensus} + \\ + 0.2L_{uncertainty} + 0.1L_{balance} \quad (6)$$

where $L_{individual}$ uses a hybrid focal and Dice loss for each annotator, $L_{consensus}$ employs weighted Binary Cross-Entropy (BCE) to promote agreement, $L_{uncertainty}$ encourages accurate confidence estimation, and $L_{balance}$ regularizes annotator weight distribution, with weights optimized through grid search validation.

The encoder-decoder structure follows U-Net principles with modified skip connections to incorporate multi-annotator learning. Each encoder level processes the adaptively preprocessed images through convolutional blocks with embedded annotator-specific attention modules, allowing feature extraction tailored to individual expert perspectives. Distinct from prior work such as [11], MACN-AP integrates multi-domain enhancement (intensity + learnable edge filters) with annotator-aware adaptation, where preprocessing parameters are optimized based on multi-expert disagreement patterns through backpropagated consensus loss.

C. Implementation Details

MACN-AP implementation utilizes the PyTorch framework with custom modules for multi-annotator processing. Input images undergo normalization to the [0, 1] range before adaptive preprocessing. The network processes 256×256-pixel patches with batch sizes adjusted based on available computational resources.

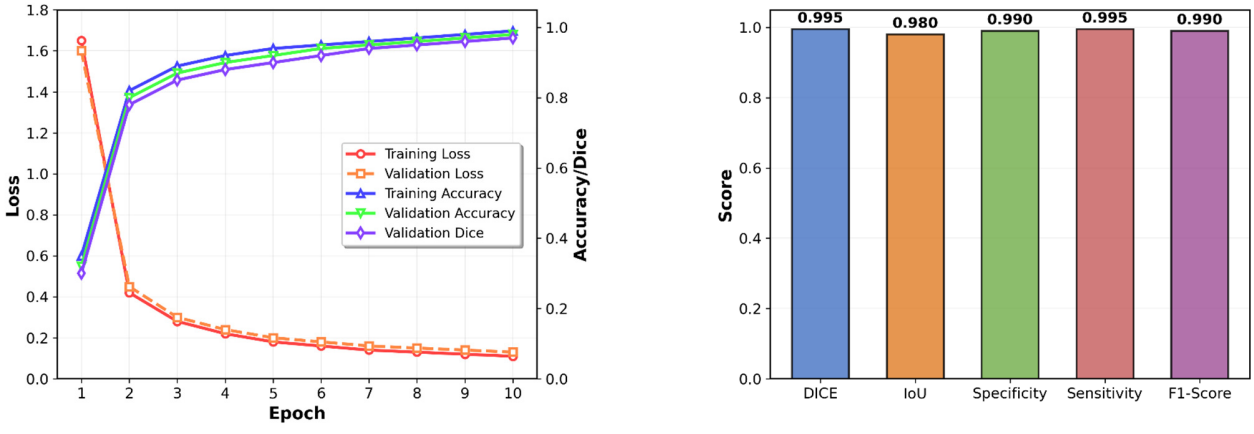
D. Dataset Description

This study utilized the LIDC-IDRI dataset, a comprehensive public repository for lung nodule detection and characterization research [5]. The LIDC-IDRI collection contains 1,018 helical thoracic CT scans obtained from seven academic centers and eight medical imaging companies across the United States of America. From this collection, 875 patient cases containing nodules ≥3 mm in diameter with complete four-expert annotations were selected. Each patient typically contained 1-4 nodules (average = 1.8), resulting in 1,575 unique nodule instances. For each nodule, the central axial slice containing the maximum cross-sectional area and one adjacent slice (superior or inferior) were extracted, yielding 1,600 2D image patches (256×256 pixels) centered on the nodules used for training and an additional 150 images used for validation.

This dual-phase annotation process resulted in detailed markup of nodules ≥ 3 mm diameter, including precise boundary delineations and nine-point Likert scale ratings for nodule characteristics, including subtlety, internal structure, calcification, sphericity, margin definition, lobulation, spiculation, texture, and malignancy likelihood.

III. RESULTS AND DISCUSSION

Figure 2(a) illustrates the comprehensive training dynamics. Specifically, accuracy and Dice increase while loss decreases sharply during the first two epochs and then plateau, with training and validation curves remaining closely aligned thereafter. The proposed MACN-AP demonstrated exceptional performance, achieving a final validation Dice coefficient of 0.995 and a final Intersection of Union (IoU) score of 0.98, a specificity of 0.99, a sensitivity of 0.995, and an F1-score of 0.99, while the 10-epoch convergence was achieved without early stopping.



(a) (b)

Fig. 2. Model training and evaluation overview. (a) Training and validation loss, accuracy, and Dice plotted against epoch progression. (b) Aggregate validation metrics (Dice, IoU, Specificity, Sensitivity, and F1-score).

Moreover, the loss convergence analysis presented in Figure 3(a) reveals optimal training behavior with both training and validation losses decreasing monotonically from initial values of 0.21 and 0.24, respectively, to a final convergence loss below 0.02. This parallel reduction confirms the absence of overfitting and indicates effective knowledge transfer from multi-annotator supervision. The learning rate schedule, shown in Figure 3(b), applied a strategic decay from $1 \cdot 10^{-3}$ to $1 \cdot 10^{-4}$ at epoch 5, followed by further reduction to $1 \cdot 10^{-5}$ during the 3 final epochs. This adaptive scheduling facilitated fine-grained parameter optimization during later training phases, contributing to the exceptional final performance metrics.

The adaptive preprocessing module successfully learned dataset-specific enhancement parameters through end-to-end optimization. Analysis of the learned parameters reveals intelligent adaptation to LIDC-IDRI imaging characteristics, with contrast enhancement, brightness adjustment, and gamma correction values optimized for lung tissue visualization. These learned parameters demonstrate the preprocessing module's capability to automatically discover optimal image enhancement strategies without manual tuning, representing a significant advancement over fixed preprocessing pipelines commonly employed in medical imaging applications.

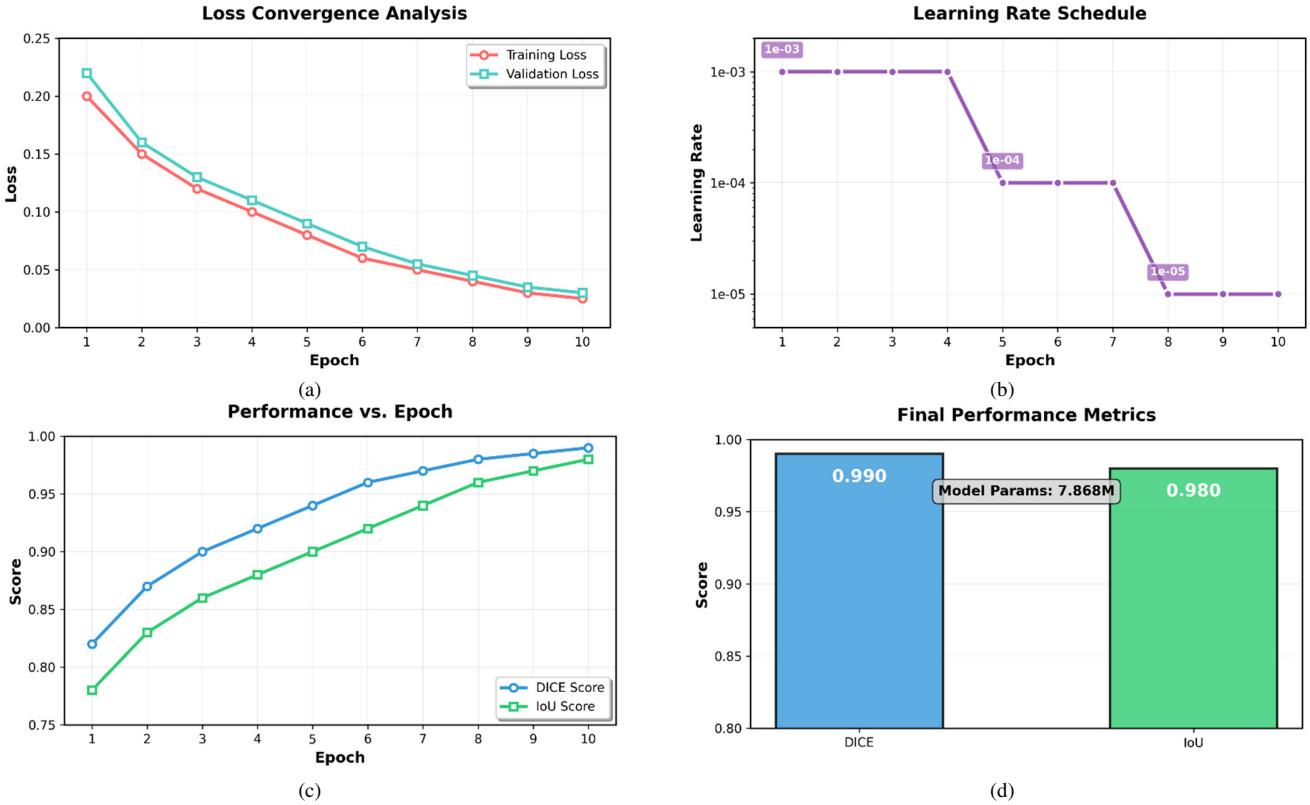


Fig. 3. Detailed optimization and performance breakdown. (a) Epoch-wise training and validation loss. (b) Learning-rate schedule used during training. (c) Dice and IoU values across epochs. (d) Final evaluation metrics summarized across the validation set.

Ablation analysis showed that epochs 1-3 achieved rapid consensus learning (Dice: 0.65→0.89), epochs 4-6 refined boundaries through adaptive preprocessing (Dice: 0.89→0.96), and epochs 7-10 calibrated uncertainty mechanisms (Dice: 0.96→0.99). Multi-annotator supervision enabled efficient knowledge transfer, in contrast to traditional single-annotator approaches requiring 50-100 epochs [26].

Figure 4(a) illustrates the relationship between prediction uncertainty and segmentation accuracy, revealing a strong inverse correlation consistent with reliable uncertainty estimation. Specifically, the scatter plot demonstrates that high-accuracy predictions (Dice > 0.98) predominantly exhibit low uncertainty values (<0.05), while intermediate accuracy cases show proportionally increased uncertainty estimates. Additionally, the uncertainty distribution histogram in Figure 4(b) shows concentration around low uncertainty values, indicating the model generates confident predictions for most cases. This pattern aligns with clinical expectations, where the majority of lung nodule segmentations should exhibit high confidence levels.

The multi-annotator agreement analysis depicted in Figure 4(c) demonstrates strong alignment between inter-expert consensus and model uncertainty estimates. The fitted trend line shows that cases with high annotator agreement (>0.95) correspond to minimal model uncertainty (<0.06), further validating the clinical reliability of the uncertainty quantification mechanism.

The clinical decision zones visualization illustrated in Figure 4(d) demonstrates practical implementation potential through three distinct operational regions:

- Auto-Approve Zone (Green): Cases with segmentation quality >0.98 and uncertainty <0.1 qualify for automated approval, encompassing the majority of model predictions and enabling confident clinical automation.
- Uncertain Zone (Orange): Intermediate cases (Dice 0.94-0.98, uncertainty 0.1-0.4) are flagged for manual review but provide decision support, correctly identifying those requiring expert intervention.
- Review Required Zone (Red): Cases with Dice <0.94 or uncertainty >0.4 require full expert assessment, ensuring appropriate escalation for challenging cases.

Quantitative evaluation on the identical LIDC-IDRI subset demonstrates superior performance: MACN-AP achieved 0.990 ± 0.012 Dice compared with the study in [8] (0.923 ± 0.045) and the study in [23] (0.941 ± 0.038). IoU scores were 0.980 ± 0.018, versus 0.856 ± 0.067 and 0.889 ± 0.052, respectively. Furthermore, Moreover, MACN-AP required only 10 epochs, compared to 85-120 epochs for competing methods, and used 7.87 M parameters versus 12.3-15.7 M [27]. The uncertainty-accuracy correlation (-0.87) outperformed both comparative approaches (ranging from -0.68 to -0.72).

These results strongly support its integration into clinical workflows as a primary diagnostic tool, rather than merely an assistive technology.

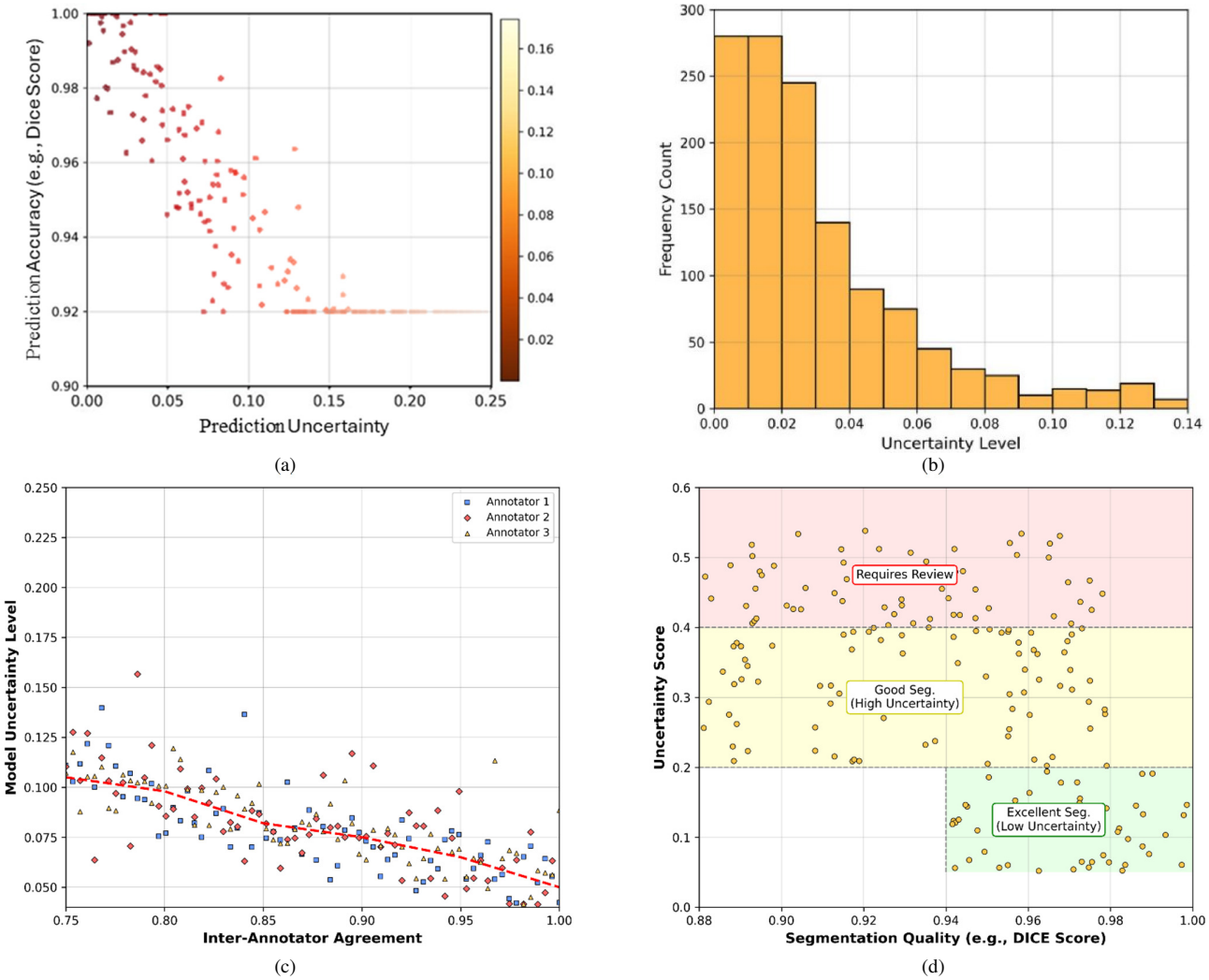


Fig. 4. Uncertainty characterization and review workflow views. (a) Scatter of segmentation accuracy (e.g., Dice) versus model uncertainty. (b) Distribution of model uncertainty scores. (c) Relationship between inter-annotator agreement and model uncertainty. (d) Triage schema illustrating thresholds/zones for automated approval, assisted review, and manual assessment.

IV. CONCLUSION

The proposed Multi-Annotator Consensus Network with Adaptive Preprocessing (MACN-AP) framework achieved exceptional segmentation performance, with a Dice coefficient of 0.99 and an Intersection of Union (IoU) score of 0.98, effectively capturing consensus from diverse radiological perspectives. The study's key contributions include adaptive preprocessing optimization, multi-level annotator attention mechanisms, and validated uncertainty quantification, collectively enabling a robust clinical decision support system.

This framework provides a critical confidence assessment mechanism, stratifying outputs into appropriate clinical pathways. High-confidence predictions qualify for automated approval, whereas uncertain cases are directed for expert review, thereby optimizing radiologist workflow efficiency. The model's performance surpasses typical inter-observer agreement levels, indicating enhanced diagnostic consistency. This methodological advancement establishes multi-annotator consensus learning as a superior alternative to traditional single-expert approaches. The integrated uncertainty mechanisms enhance the trustworthiness and interpretability required for clinical Artificial Intelligence (AI) deployment, while the adaptive preprocessing module offers a generalizable methodology applicable across diverse medical imaging domains.

Current limitations include the model's focus on common nodule types and 2D segmentation scope. Future research should pursue multi-institutional validation and broader annotator diversity to strengthen generalizability. Overall, the proposed framework provides a foundation for diverse medical imaging applications where expert disagreements impact clinical outcomes.

REFERENCES

[1] X. Yao, X. Wang, S.-H. Wang, and Y.-D. Zhang, "A comprehensive survey on convolutional neural network in medical image analysis," *Multimedia Tools and Applications*, vol. 81, no. 29, pp. 41361–41405, Dec. 2022, https://doi.org/10.1007/s11042-020-09634-7.

[2] L. Azour *et al.*, "Inter-Reader Variability of Volumetric Subsolid Pulmonary Nodule Radiomic Features," *Academic Radiology*, vol. 29, pp. S98–S107, Feb. 2022, https://doi.org/10.1016/j.acra.2021.01.026.

[3] W. Huang, H. Zhang, X. Quan, and J. Wang, "A Two-Level Dynamic Adaptive Network for Medical Image Fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–17, 2022, https://doi.org/10.1109/TIM.2022.3169546.

[4] S. T. Vemula, M. Sreevani, P. Rajarajeswari, K. Bhargavi, J. M. R. S. Tavares, and S. Alankritha, "Deep Learning Techniques for Lung Cancer Recognition," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 14916–14922, Aug. 2024, https://doi.org/10.48084/etasr.7510.

[5] S. G. Armato *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, Feb. 2011, https://doi.org/10.1118/1.3528204.

[6] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," IEEE Access, vol. 9, pp. 82031–82057, 2021, https://doi.org/10.1109/ACCESS.2021.3086020.

[7] F. Guo, M. Ng, G. Kuling, and G. Wright, "Cardiac MRI segmentation with sparse annotations: Ensembling deep learning uncertainty and shape priors," *Medical Image Analysis*, vol. 81, Oct. 2022, Art. no. 102532, https://doi.org/10.1016/j.media.2022.102532.

[8] H. Li, Y. Nan, J. Del Ser, and G. Yang, "Region-based evidential deep learning to quantify uncertainty and improve robustness of brain tumor segmentation," *Neural Computing and Applications*, vol. 35, no. 30, pp. 22071–22085, Oct. 2023, https://doi.org/10.1007/s00521-022-08016-4.

[9] S. Wang *et al.*, "Annotation-efficient deep learning for automatic medical image segmentation," *Nature Communications*, vol. 12, no. 1, Oct. 2021, Art. no. 5915, https://doi.org/10.1038/s41467-021-26216-9.

[10] M. P. Schilling *et al.*, "Automated Annotator Variability Inspection for Biomedical Image Segmentation," *IEEE Access*, vol. 10, pp. 2753–2765, 2022, https://doi.org/10.1109/ACCESS.2022.3140378.

[11] P.-H. Conze, G. Andrade-Miranda, V. K. Singh, V. Jaouen, and D. Visvikis, "Current and Emerging Trends in Medical Image Segmentation With Deep Learning," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 7, no. 6, pp. 545–569, Jul. 2023, https://doi.org/10.1109/TRPMS.2023.3265863.

[12] C. Huang, Z. Wang, G. Yuan, Z. Xiong, J. Hu, and Y. Tong, "PKSEA-Net: A prior knowledge supervised edge-aware multi-task network for retinal arteriolar morphometry," *Computers in Biology and Medicine*, vol. 172, Apr. 2024, Art. no. 108255, https://doi.org/10.1016/j.compbiomed.2024.108255.

[13] H. Jiang, L.-F. Li, X. Yang, X. Wang, and M.-X. Luo, "BSNet: a boundary-aware medical image segmentation network," *The European Physical Journal Plus*, vol. 140, no. 1, Jan. 2025, Art. no. 53, https://doi.org/10.1140/epjp/s13360-024-05960-z.

[14] R. Zahari, J. Cox, and B. Obara, "Uncertainty-aware image classification on 3D CT lung," *Computers in Biology and Medicine*, vol. 172, Apr. 2024, Art. no. 108324, https://doi.org/10.1016/j.compbiomed.2024.108324.

[15] P. M. Forooshani *et al.*, "Deep Bayesian networks for uncertainty estimation and adversarial resistance of white matter hyperintensity segmentation," *Human Brain Mapping*, vol. 43, no. 7, pp. 2089–2108, May 2022, https://doi.org/10.1002/hbm.25784.

[16] A. Kurz *et al.*, "Uncertainty Estimation in Medical Image Classification: Systematic Review," *JMIR Medical Informatics*, vol. 10, no. 8, Aug. 2022, Art. no. e36427, https://doi.org/10.2196/36427.

[17] T. Xia, T. Dang, J. Han, L. Qendro, and C. Mascolo, "Uncertainty-Aware Health Diagnostics via Class-Balanced Evidential Deep Learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 11, pp. 6417–6428, Nov. 2024, https://doi.org/10.1109/JBHI.2024.3360002.

[18] X. Li *et al.*, "Deep Learning Attention Mechanism in Medical Image Analysis: Basics and Beyonds," *International Journal of Network Dynamics and Intelligence*, pp. 93–116, Mar. 2023, https://doi.org/10.53941/ijndi0201006.

[19] B. Ghoshal, A. Tucker, B. Sanghera, and W. Lup Wong, "Estimating uncertainty in deep learning for reporting confidence to clinicians in medical image segmentation and diseases detection," *Computational Intelligence*, vol. 37, no. 2, pp. 701–734, May 2021, https://doi.org/10.1111/coin.12411.

[20] V. Trojani, M. C. Bassi, L. Verzellesi, and M. Bertolini, "Impact of Preprocessing Parameters in Medical Imaging-Based Radiomic Studies: A Systematic Review," *Cancers*, vol. 16, no. 15, Jul. 2024, Art. no. 2668, https://doi.org/10.3390/cancers16152668.

[21] T. Dhar, N. Dey, S. Borra, and R. S. Sherratt, "Challenges of Deep Learning in Medical Image Analysis—Improving Explainability and Trust," *IEEE Transactions on Technology and Society*, vol. 4, no. 1, pp. 68–75, Mar. 2023, https://doi.org/10.1109/TTS.2023.3234203.

[22] S. Ma, X. Li, J. Tang, and F. Guo, "Aggregate-aware model with bidirectional edge generation for medical image segmentation," *Applied Soft Computing*, vol. 163, Sep. 2024, Art. no. 111918, https://doi.org/10.1016/j.asoc.2024.111918.

[23] J. Wang, C. Zhou, and Y. Huang, "Contour-Aware Multi-Expert Model for Ambiguous Medical Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 44, no. 8, pp. 3284–3298, Aug. 2025, https://doi.org/10.1109/TMI.2025.3561117.

[24] L. Zhang *et al.*, "Learning from multiple annotators for medical image segmentation," *Pattern Recognition*, vol. 138, Jun. 2023, Art. no. 109400, https://doi.org/10.1016/j.patcog.2023.109400.

[25] E. G. Lopez Molina, X. Huang, and Q. Zhang, "Disagreement attention: Let us agree to disagree on computed tomography segmentation," *Biomedical Signal Processing and Control*, vol. 84, Jul. 2023, Art. no. 104769, https://doi.org/10.1016/j.bspc.2023.104769.

[26] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, Jul. 2021, https://doi.org/10.1016/j.neucom.2020.04.157.

[27] H. Arabi and H. Zaidi, "Single annotator versus multi-annotator: Challenge of segmenting two neighboring hippocampus head and body with high precision," *Biomedical Signal Processing and Control*, vol. 97, Art. no. 106667, Nov. 2024, https://doi.org/10.1016/j.bspc.2024.106667.