Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.088036

ARTICLE

Actor-Critic Reinforcement Learning-Guided Adaptive Lesion Refinement Network for Brain Stroke Lesion Segmentation

Zhanar Azhibekova[1], Akbayan Aliyeva[2,*], Daniyar Rakhymberdiyev[3,4] and Meruert Dauletova[3,4,*]

[1]Department of Information and Communication Technologies, Kazakh National Medical University named after S.D. Asfendiyarov, Almaty, Kazakhstan

[2]Department of Computer Engineering, Khoja Akhmet Yassawi Kazakh-Turkish University, Turkistan, Kazakhstan

[3]School of Digital Technologies, Narxoz University, Almaty, Kazakhstan

[4]School of Physical Education and Sports, International University of Tourism and Hospitality, Turkistan, Kazakhstan

*Corresponding Authors: Akbayan Aliyeva. Email: akbayan.aliyeva@ayu.edu.kz; Meruert Dauletova. Email: dauletova.meruert@iuth.edu.kz

Received: 27 June 2026; Accepted: 11 September 2026; Published: 21 October 2026

ABSTRACT: Accurate stroke-lesion segmentation is difficult because lesion morphology, boundary contrast, and acquisition characteristics differ across magnetic resonance imaging (MRI), computed tomography perfusion (CTP), and non-contrast computed tomography (NCCT). This study proposes an Actor-Critic Reinforcement Learning-Guided Adaptive Lesion Refinement Network (ACRL-Net) that combines a multi-scale encoder-decoder with an uncertainty-aware actor-critic agent and applies up to four region- and boundary-level mask refinements after the initial prediction. Unlike supervised single-pass segmentation, the policy selects sequential corrective actions, while the critic is used during training to estimate long-term action value and is omitted at inference. Evaluation used the reported held-out partitions of ISLES 2017, ISLES 2018, and AISD, comprising 575 cases across MRI, CTP, and NCCT. The aggregate evaluation reported Dice of 0.913, IoU of 0.842, Precision of 0.926, Recall of 0.901, and HD95 of 4.21 mm. Component ablation and qualitative refinement trajectories support the contribution of the encoder-decoder and actor-critic modules. Because case-level statistical outputs, an independent external cohort, reward-weight sensitivity logs, and deployment profiling were not available in the retained study record, these findings are interpreted as descriptive evidence of feasibility rather than proof of universal superiority or clinical readiness. External multi-center validation, statistical testing from retained case-level predictions, and measured efficiency benchmarking remain necessary.

KEYWORDS: Stroke lesion segmentation; actor-critic reinforcement learning; iterative mask refinement; uncertainty; magnetic resonance imaging; computed tomography perfusion; non-contrast computed tomography

1 Introduction

Stroke is a leading cause of mortality and long-term disability, and imaging-based lesion delineation supports treatment planning, prognosis estimation, and rehabilitation assessment. The segmentation problem is heterogeneous rather than modality-neutral: ISLES 2017 is MRI-based, ISLES 2018 uses CT perfusion information with a DWI-derived reference, and AISD contains NCCT scans annotated with DWI support. Across these settings, lesions may have irregular shapes, heterogeneous intensities, blurred boundaries, severe foreground-background imbalance, and substantial inter-patient variability. Deep learning has improved feature extraction and end-to-end optimization [1], while convolutional encoder-decoder

Copyright © 2026 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

networks capture local context effectively [2]. However, small lesions, low-contrast tissue, and complex contours remain difficult [3,4].

Recent developments have introduced transformer-based architectures and hybrid convolution-transformer networks to enhance long-range contextual modeling and improve feature representation in medical image segmentation. Self-attention mechanisms enable global dependency modeling and have demonstrated superior performance compared with conventional convolutional models in numerous biomedical imaging applications [5]. Nevertheless, most existing segmentation networks rely on a single forward inference process, in which prediction errors cannot be corrected after the initial segmentation is generated. As a result, false-positive regions, missed lesions, and inaccurate lesion boundaries frequently remain in the final prediction. Although post-processing algorithms can alleviate some of these limitations, they are generally based on predefined heuristics and lack adaptive decision-making capabilities. Reinforcement learning provides a promising alternative by enabling an intelligent agent to sequentially optimize decisions according to interactions with the environment and reward-driven learning [6]. Unlike conventional segmentation models, reinforcement learning can formulate lesion refinement as a sequential decision-making process, allowing the model to progressively improve segmentation quality through adaptive actions. Recent studies have demonstrated the effectiveness of reinforcement learning in image analysis, medical image registration, and clinical decision support, suggesting its potential for adaptive segmentation refinement [7].

Supervised segmentation and reinforcement learning optimize different problems. A supervised network learns a direct mapping from an image to a reference mask by minimizing a differentiable loss over labeled examples; its inference is normally a single forward pass. In contrast, reinforcement learning defines a state, an action, a transition, and a return. The ACRL-Net policy observes the current mask and uncertainty, selects a corrective action, receives a reward that summarizes the change in segmentation quality during training, and may stop when further refinement is not selected. Reinforcement learning therefore supplements rather than replaces supervised representation learning in this framework.

Motivated by these observations, this paper proposes ACRL-Net for adaptive stroke-lesion refinement. The novelty is not reinforcement learning for segmentation in isolation. It is the integration of (i) a multi-scale stroke-segmentation backbone, (ii) an uncertainty-aware state containing the current mask and encoder features, (iii) a finite set of region- and boundary-level correction actions, and (iv) actor-critic value learning that selects a short sequence of post-prediction refinements. The critic guides policy learning but is not used during inference. The study evaluates whether this combination improves the reported aggregate overlap and boundary metrics relative to the included within-study baselines.

The study does not seek to establish ACRL-Net as universally best for all stroke types, modalities, or clinical environments. Its evidence is limited to retrospective public benchmarks and the retained aggregate records. Accordingly, claims in the revised manuscript are restricted to the reported evaluation, and prospective real-world deployment remains outside the demonstrated scope.

2 Related Works

Stroke-lesion segmentation spans clinically distinct tasks. Acute ischemic lesion segmentation commonly uses DWI/ADC MRI, perfusion imaging, or NCCT/CTP, whereas chronic post-stroke delineation frequently relies on structural T1-weighted MRI. Hemorrhagic-lesion segmentation on CT has different intensity patterns and clinical objectives, and generic tumor or organ segmentation should not be treated as a direct stroke comparator. Public resources reflect this diversity: ISLES benchmarks acute ischemic settings [8,9], AISD provides NCCT with DWI-supported annotations [10], and ATLAS v2.0 provides a large

chronic post-stroke T1-weighted MRI benchmark [11]. Recent stroke-specific work includes transformer [12], attention [13], frequency-domain [14], CNN-ViT [15], and clinically validated ensemble approaches [16–18].

Convolutional encoder-decoder networks remain strong baselines because they preserve local texture and spatial detail, while transformer and hybrid architectures model longer-range dependencies. Recent residual and dual-attention U-Net design [19] demonstrates how supervised architectural changes can improve boundary representation. Tensor residual attention combined with superpixel features has also been explored for efficient brain-tumor diagnosis [20]. These two studies are relevant architectural references, but their tasks and datasets differ from stroke-lesion segmentation; they are therefore discussed as methodological context rather than inserted as numerically comparable stroke baselines. Likewise, CSNet [21] and a lightweight asymmetric U-Net for acute ischemic stroke in CT/CTP [22] show that task-specific convolutional design and modality-aware preprocessing remain important.

Reinforcement learning-based medical image segmentation includes interactive click-guided refinement, pixel-level multi-agent decisions, organ-specific anatomical mapping, and value- or policy-based contour correction. RCI-Seg uses deep reinforcement learning in an interactive biomedical segmentation setting [7]. PixelDRL-MG treats pixels as agents and uses an asynchronous advantage actor-critic strategy to generate masks iteratively [23]. Other recent studies apply reinforcement learning to intracerebral hemorrhage segmentation [24] and pancreas subregion/duct mapping [25]. These methods establish that reinforcement learning for segmentation and refinement is not new. Their state design, action granularity, supervision, target anatomy, and inference interaction differ from the region- and boundary-level, uncertainty-aware stroke refinement used in ACRL-Net.

Despite the growing application of reinforcement learning to medical image segmentation, an important methodological gap remains in the development of a reproducibly specified, stroke-focused refinement framework that can correct an initial segmentation through a limited sequence of adaptive decisions. Existing reinforcement learning paradigms differ substantially in their state representations, action granularity, learning strategies, target anatomy, and degree of user interaction, making it important to distinguish the proposed approach from reinforcement learning for segmentation in general. Table 1 summarizes the methodological positioning of ACRL-Net relative to representative reinforcement learning-based segmentation paradigms, including interactive region correction, pixel-level multi-agent segmentation, hemorrhage segmentation, and anatomical localization. In contrast to these approaches, ACRL-Net combines the source image, current segmentation mask, multi-scale encoder features, and uncertainty information within an actor-critic framework that performs region- and boundary-level post-prediction corrections and supports explicit termination after a maximum of four refinement steps. Accordingly, the contribution of ACRL-Net lies in integrating a multi-scale initial predictor with uncertainty-aware sequential decision making for stroke-specific lesion refinement rather than introducing reinforcement learning for medical image segmentation itself. The comparison in Table 1 is intended to clarify these methodological distinctions and does not imply that the listed approaches were evaluated under identical datasets, experimental protocols, or benchmark conditions.

Table 1: Positioning of ACRL-Net relative to representative reinforcement learning segmentation paradigms.

Method/Task	State or Input	Action Granularity	Learning Signal	Relation to ACRL-Net
RCI-Seg [7]	Image plus interactive clicks	Interactive region correction	DRL-guided interaction	Requires user interaction; not stroke-specific

(Continued)

Table 1 (continued)

Method/Task	State or Input	Action Granularity	Learning Signal	Relation to ACRL-Net
PixelDRL-MG [23]	Image and evolving pixel states	Pixel-level multi-agent actions	Asynchronous advantage actor-critic	Automatic pixel-wise generation rather than post-decoder regional correction
ICH RL segmentation [24]	Brain CT features	Volumetric hemorrhage segmentation	RL-enhanced UNet-Transformer	Hemorrhagic CT target; different pathology
Pancreas anatomical maps [25]	CT anatomy and learned maps	Organ/subregion mapping	RL-based anatomical localization	Different anatomy and output objective
ACRL-Net	Image, current mask, encoder features, uncertainty	Region and boundary actions; explicit stop	Actor-critic return over up to four refinements	Stroke-focused post-prediction refinement

3 Materials and Methods

The proposed methodology integrates deep feature learning with actor-critic reinforcement learning to achieve adaptive and accurate brain stroke lesion segmentation. The overall framework combines hierarchical feature extraction, progressive lesion reconstruction, sequential policy-guided refinement, and reward-driven optimization into a unified end-to-end architecture. This section describes the proposed network architecture, the actor-critic reinforcement learning strategy, the adaptive lesion refinement algorithm, the experimental datasets, and the evaluation methodology used to validate the effectiveness of the proposed ACRL-Net.

3.1 Overall Framework

The proposed Actor-Critic Reinforcement Learning-Guided Adaptive Lesion Refinement Network (ACRL-Net) formulates stroke-lesion mask correction as a short sequential decision process after an initial encoder-decoder prediction. The framework combines deep feature extraction, segmentation decoding, and actor-critic policy learning. Fig. 1 depicts the intended workflow from MRI/CT image acquisition through retrospective quantitative evaluation. Its deployment and clinical-decision-support blocks are conceptual extensions and were not evaluated in the reported experiments.

Fig. 1 depicts six intended stages: data input, modality-specific preprocessing, reinforcement-learning-guided segmentation, output generation, quantitative evaluation, and a conceptual deployment pathway. During training, the policy interacts with the segmentation environment through action selection and reward feedback. The reported experiments evaluate the lesion-mask output and segmentation metrics only; lesion severity, stroke type, treatment planning, progress monitoring, and clinical decision support shown in the schematic were not tested and are not used to support the claims.

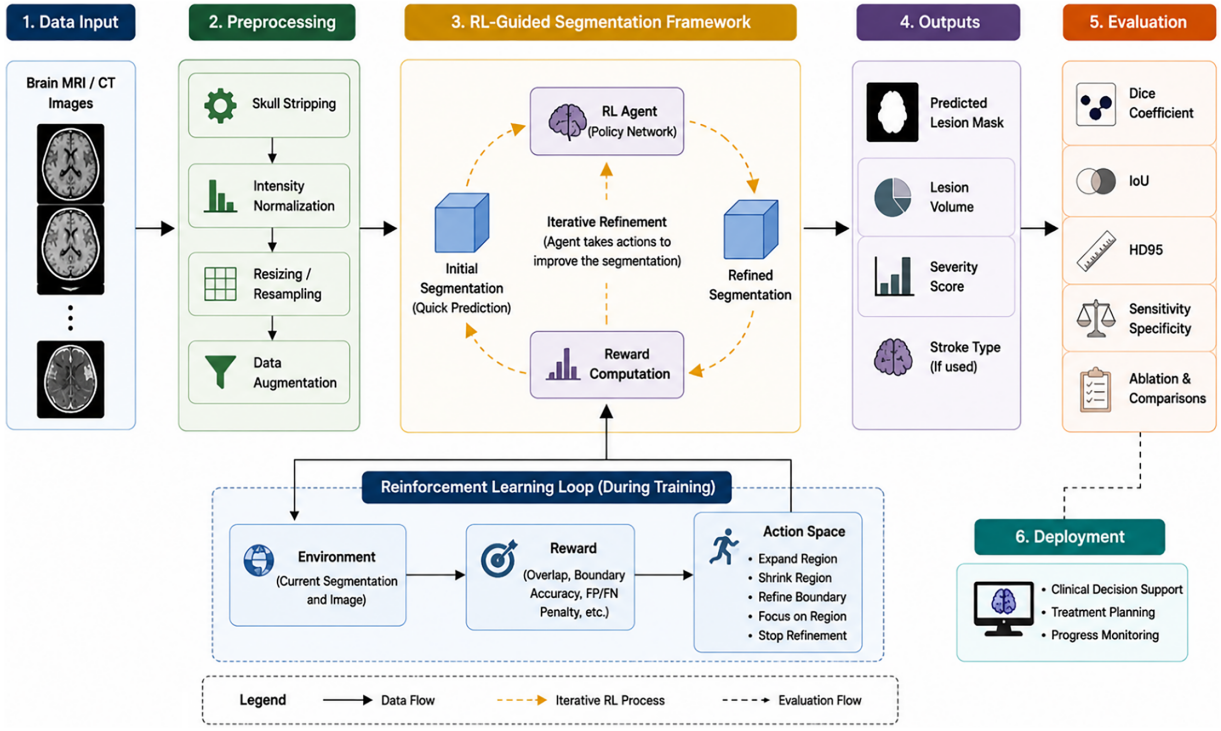


Figure 1: Overall flowchart of the study.

Given an input brain image volume

$$X \in \mathbb{R}^{H \times W \times D} \tag{1}$$

the objective is to estimate the final lesion mask

$$M_T = f(X) \tag{2}$$

where $f(\cdot)$ represents ACRL-Net after T refinement iterations. Unlike a single-pass segmentation network, the policy may update the lesion mask until Stop is selected or the refinement cap is reached. The reported output is the refined lesion mask.

3.2 RL-Guided Segmentation Architecture

The overall architecture of the proposed network consists of four interconnected components: a multi-scale feature encoder, an actor-critic reinforcement learning agent, a segmentation decoder, and a reward optimization module. Fig. 2 illustrates the interaction among these components, where encoder features provide rich semantic representations, the reinforcement learning agent determines refinement actions, and the decoder reconstructs the final lesion segmentation. The reward computation module evaluates each refinement step and provides feedback for policy optimization.

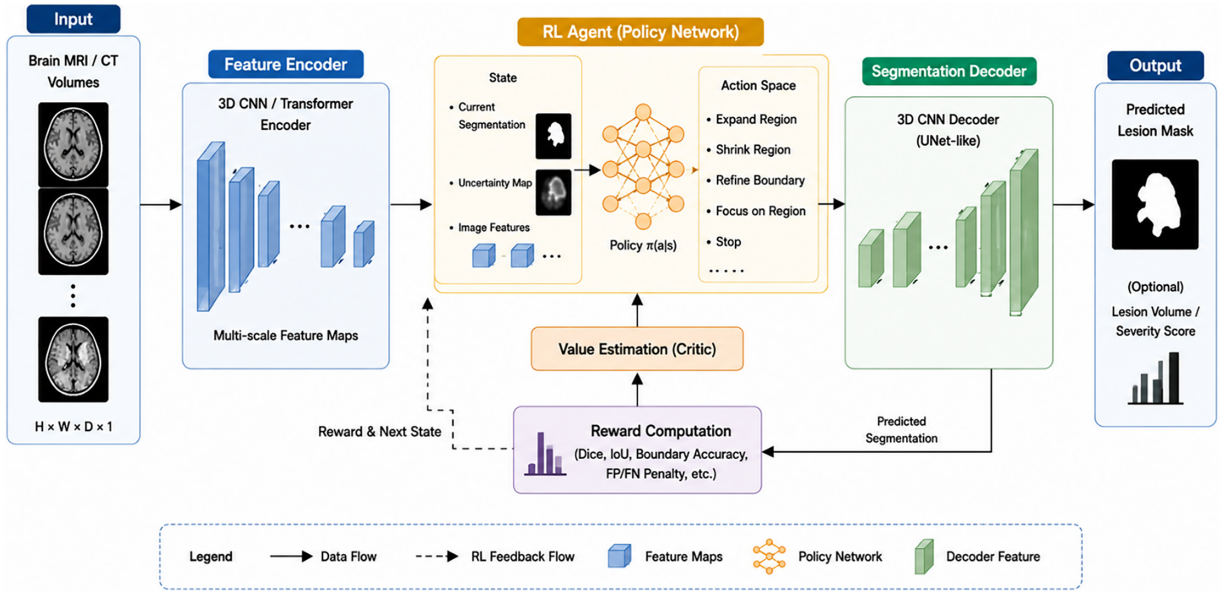


Figure 2: Main architecture of RL-guided segmentation network.

The segmentation process begins by extracting hierarchical image representations using the feature encoder. Let

$$F = \{F_1, F_2, F_3, F_4, F_B\} \tag{3}$$

denote the multi-scale feature maps generated by the encoder, where F_B represents the bottleneck representation. These features are subsequently provided to both the segmentation decoder and the reinforcement learning agent, allowing global semantic information and local structural details to jointly guide lesion refinement.

Instead of directly accepting the decoder prediction as the final segmentation, the proposed architecture formulates refinement as a sequential decision-making process. At each iteration, the reinforcement learning agent observes the current segmentation state

$$s_t = \{X, M_t, F, U_t\} \tag{4}$$

where M_t denotes the current lesion mask and U_t represents the uncertainty map estimated from the segmentation network. Based on this state, the policy network predicts an optimal refinement action, while the critic network evaluates its long-term benefit.

3.3 Multi-Scale Feature Encoder and Segmentation Decoder

The proposed network employs an encoder-decoder architecture to simultaneously extract hierarchical semantic features and reconstruct accurate brain stroke lesion segmentation maps. This design enables effective preservation of multi-scale contextual information while progressively recovering fine anatomical boundaries during the decoding process. Fig. 3 depicts the overall encoder-decoder architecture adopted in the proposed framework. Specifically, Fig. 3a illustrates the multi-scale feature encoder responsible for hierarchical feature extraction, whereas Fig. 3b presents the adaptive segmentation decoder that progressively reconstructs the final lesion mask using the encoded feature representations.

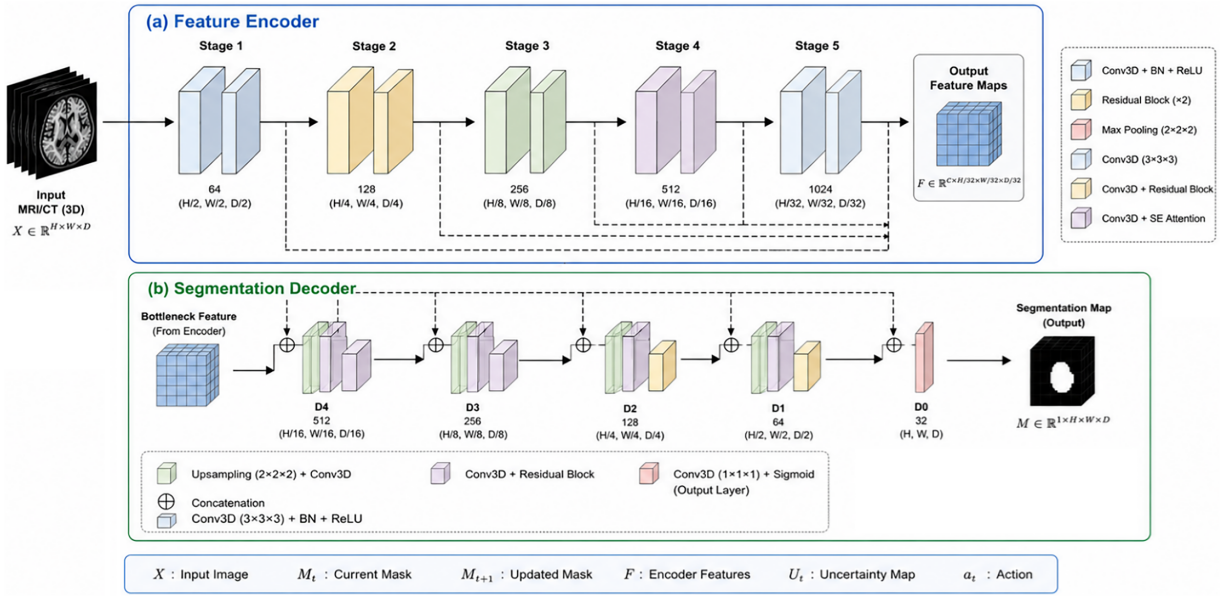


Figure 3: Encoder and decoder architecture.

The feature encoder adopts a five-stage hierarchical architecture composed of three-dimensional convolutional layers, residual learning blocks, squeeze-and-excitation attention, and progressive downsampling operations. Each encoder stage extracts increasingly abstract semantic representations while reducing spatial resolution. Consequently, shallow layers preserve fine anatomical structures, whereas deeper layers capture high-level contextual information necessary for lesion localization.

The encoder operation can be expressed as

$$F_i = \mathcal{E}_i(F_{i-1}) \tag{5}$$

where

$$F_0 = X \tag{6}$$

and

$$i = 1, \ldots, 5 \tag{7}$$

here, $\mathcal{E}_i(\cdot)$ denotes the transformation performed by the i-th encoder stage.

The segmentation decoder progressively reconstructs high-resolution lesion masks through successive upsampling and feature fusion. Skip connections combine encoder representations with decoder features, enabling recovery of fine structural information that would otherwise be lost during downsampling. Fig. 3b illustrates the decoder hierarchy together with concatenation operations and progressive feature refinement.

The decoder reconstruction process is formulated as

$$D_i = \mathcal{D}_i(D_{i+1}, F_i) \tag{8}$$

where D_i denotes the decoder feature map at level i. The final segmentation mask is obtained through a sigmoid activation

$$M_0 = \sigma(D_0) \tag{9}$$

where M_0 serves as the initial prediction before reinforcement learning refinement.

3.4 Actor-Critic Reinforcement Learning Agent

The adaptive lesion refinement mechanism is realized using an actor-critic reinforcement learning framework. Fig. 4 illustrates the complete policy network architecture, including state encoding, actor network, critic network, action space, and optimization losses. Listing 1 summarizes the complete training procedure employed by the proposed algorithm.

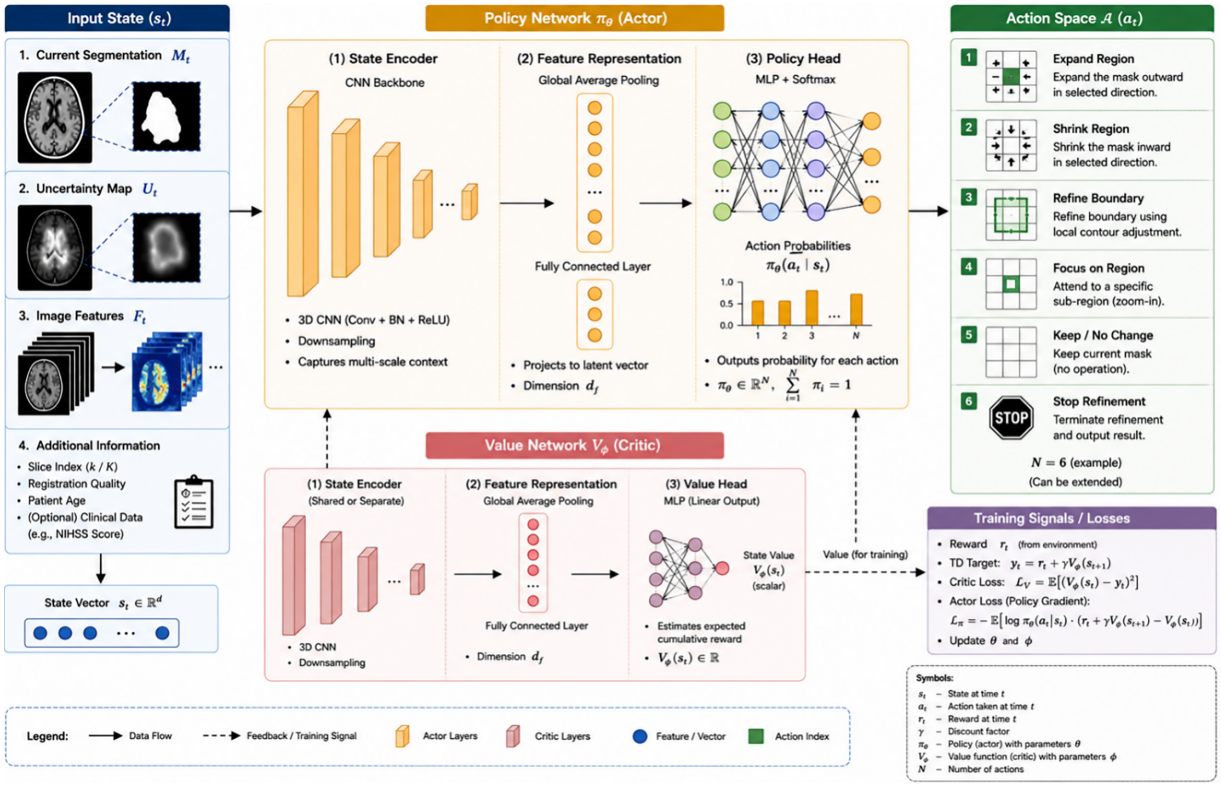


Figure 4: Architecture of the RL agent (policy network).

The reported state contains four inputs: the source image X, current binary mask M_t, multi-scale encoder features F, and uncertainty map U_t. X and F preserve anatomy and modality-specific evidence; M_t identifies which prediction is being corrected; and U_t directs the policy toward ambiguous regions rather than already confident background. U_t is derived from the decoder probability map and is updated after each correction.

$$s_t = \{X, M_t, F, U_t\} \tag{10}$$

The actor network estimates the probability distribution over all refinement actions according to

$$\pi_\theta(a_t \mid s_t) = \text{Softmax}\,(f_\theta(s_t)) \tag{11}$$

where f_θ denotes the actor network parameterized by θ.

During training, actions are sampled from the learned policy

$$a_t \sim \pi_\theta(a \mid s_t) \tag{12}$$

The policy outputs one of six discrete actions: expand region, shrink region, refine boundary, focus on an uncertain region, keep the mask unchanged, or stop. These actions do not operate as independent per-pixel agents. The manuscript and figures support a region- or boundary-level interpretation in which candidate voxels are guided by the decoder probability and uncertainty maps; the resulting abstract transition is denoted by Eq. (13). Exact structuring elements, probability/uncertainty thresholds, and connected-component rules were not retained and are therefore not claimed as code-equivalent settings.

$$M_{t+1} = g(M_t, a_t) \tag{13}$$

where $g(\cdot)$ denotes the adaptive refinement operator.

After executing an action, the environment transitions to the next state

$$s_{t+1} = \{X, M_{t+1}, F, U_{t+1}\} \tag{14}$$

The critic network estimates the expected cumulative reward associated with the current state according to

$$V_\phi(s_t) = \text{Critic}(s_t) \tag{15}$$

where ϕ represents the critic parameters. Fig. 4 demonstrates that the critic is used only during training to estimate long-term return and stabilize policy learning; the critic is omitted from the inference graph.

Operational Interpretation of the Refinement Actions

The actor samples refinement actions during training and selects the action with the highest predicted probability during deterministic inference. The refinement operator g is implemented as a deterministic mask transformation and is therefore not differentiated through during optimization. Policy gradients update the actor according to the sampled action and its temporal-difference advantage, whereas the critic is optimized using the corresponding value-estimation error. Consequently, the proposed framework combines differentiable components, including the encoder, decoder, actor, and critic, with discrete operations governing mask transitions and episode termination. Table 2 summarizes the operational interpretation, granularity, and inference purpose of the discrete actions available to the policy. Specifically, the Expand and Shrink actions modify the predicted lesion extent to address under-segmentation and false-positive regions, respectively, while Refine Boundary targets ambiguous contour regions to improve local boundary agreement. Focus on Region directs subsequent correction toward high-uncertainty areas, whereas Keep/No Change preserves the current mask without terminating the refinement process. Finally, the Stop action terminates the refinement sequence and returns the current mask as the final prediction. This structured action space enables the policy to perform targeted corrections according to the segmentation state rather than applying a uniform post-processing operation to all predicted regions.

Inference begins with the decoder prediction M_0. The actor then applies at most four refinement steps. Refinement ends when the actor selects Stop or when the hard cap $T = 4$ is reached; Keep preserves the current mask but does not terminate the episode. The decoder and feature encoder provide the initial probability/features, the actor is evaluated once per refinement state, and no ground-truth mask or critic output is required during inference.

Table 2: Operational interpretation and granularity of the discrete refinement actions.

Action	Granularity and Mask Transition	Purpose at Inference
Expand region	Region/outer boundary intent; add uncertainty-gated candidates (kernel and threshold not retained)	Recover under-segmented lesion extent
Shrink region	Region/outer boundary intent; remove low-support candidates (kernel and threshold not retained)	Suppress false-positive extent
Refine boundary	Contour-band intent; reclassify ambiguous contour candidates from decoder evidence (band width not retained)	Improve local contour agreement
Focus on region	High-uncertainty region-of-interest intent (component rule not retained)	Concentrate the next correction on a difficult subregion
Keep/no change	Whole mask; $M_{t+1} = M_t$	Avoid an unnecessary correction without terminating
Stop	Terminal action; output current mask	End refinement early

3.5 Reward Function and Training Objective

The proposed reinforcement learning framework employs a reward function specifically designed for medical image segmentation. Rather than optimizing only overlap-based metrics, the reward simultaneously considers segmentation accuracy, boundary quality, false-positive suppression, and false-negative reduction. Listing 1 incorporates this reward computation within every refinement iteration, allowing the policy network to learn clinically meaningful correction strategies.

The reward is defined as

$$\begin{aligned} r_t \; &= \lambda_1 \left[\mathrm{Dice}\left(M_{t+1}, Y\right) - \mathrm{Dice}\left(M_t, Y\right)\right] \\ &+ \lambda_2 \left[\mathrm{IoU}\left(M_{t+1}, Y\right) - \mathrm{IoU}\left(M_t, Y\right)\right] \\ &- \lambda_3 \,\mathrm{FP}\left(M_{t+1}, Y\right) - \lambda_4 \,\mathrm{FN}\left(M_{t+1}, Y\right) \\ &- \lambda_5 \,\mathrm{HD95}\left(M_{t+1}, Y\right) \end{aligned} \tag{16}$$

where Y denotes the ground-truth lesion mask and λ_1,..., λ_5 weight overlap improvement, false-positive and false-negative control, and boundary distance. Dice and IoU enter as changes between consecutive masks, whereas FP, FN, and HD95 are penalties. The plotted reward is treated as a dimensionless weighted optimization objective rather than a physical or clinical unit; its absolute scale depends on the coefficients and any term normalization used in the original implementation. Those coefficients and normalization details were not retained in the study record and are therefore not reconstructed as measured settings.

Dice and IoU are mathematically correlated for a single binary confusion matrix, with IoU = Dice/(2 – Dice). Their simultaneous inclusion therefore does not provide independent information. In the original reward design, the two incremental terms were intended to emphasize different portions of the low-vs. high-overlap range, but this can overweight overlap unless λ_2 is constrained. The revised manuscript identifies this redundancy explicitly and recommends a single overlap term, or a documented coefficient sensitivity analysis, in future replication.

The actor-critic framework is optimized using temporal-difference learning. The critic minimizes the value estimation error according to

$$L_{\text{Critic}} = \left(r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t)\right)^2 \tag{17}$$

while the actor updates the policy using

$$L_{\text{Actor}} = -\log \pi_\theta(a_t \mid s_t)\left(r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t)\right) \tag{18}$$

The corresponding temporal-difference advantage is computed as

$$A_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t) \tag{19}$$

which measures the improvement achieved by the selected action relative to the critic's prediction.

Listing 1: Actor-critic adaptive segmentation algorithm

Input:

Brain image volume X

Initial segmentation mask M_0

Encoder feature maps F

Ground-truth mask Y during training

Maximum refinement steps T

Output:

Refined lesion mask M_t

1: Initialize segmentation state:

$s_0 = \{X, M_0, F, U_0\}$

2: for t =0 to $T-1$ do

3: Estimate action probabilities using policy network:

$\pi_\theta(a_t \mid s_t) = Softmax(f_\theta(s_t))$

4: Select refinement action:

$a_t \sim \pi_\theta(a \mid s_t)$

5: Apply selected action to current mask:

$M_{t+1} = g(M_t, a_t)$

where $a_t \in \{$expand lesion region,

shrink false-positive region,

refine boundary,

focus on uncertain region,

keep unchanged,

stop refinement$\}$

6: Compute reward:

$r_t = \lambda_1[Dice(M_{t+1}, Y) - Dice(M_t, Y)]$

$+ \lambda_2[IoU(M_{t+1}, Y) - IoU(M_t, Y)]$

$- \lambda_3 FP(M_{t+1}, Y) - \lambda_4 FN(M_{t+1}, Y)$

$- \lambda_5 HD95(M_{t+1}, Y)$

7: Update state:

$s_{t+1} = \{X, M_{t+1}, F, U_{t+1}\}$

8: Estimate state value:

$V_\phi(s_t) = Critic(s_t)$

9: Update actor and critic using:

$L_{Critic} = \left(r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t)\right)^2$

$L_{Actor} = -\log \pi_\theta(a_t \mid s_t)\left(r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t)\right)$

$A_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t)$

10: if at = stop refinement then

break

end if

11: end for

12: Return final refined segmentation mask M_t

Listing 1 demonstrates that refinement continues until the policy selects Stop or the maximum number of refinement iterations is reached. The actor-critic objective optimizes the expected episode return rather than back-propagating through the discrete mask operator.

3.6 Reproducibility and Hyperparameter Record

Table 3 separates settings that can be verified from the retained manuscript and figures from settings not preserved in the archived study record. Hyperparameter selection was described as validation-fold based, but the optimizer, actor/critic learning rates, batch size, discount factor, reward coefficients, and operator thresholds cannot be recovered reliably without the original code, checkpoints, or experiment log. No values are imputed. A complete reproducibility release should report these fields and repeat reward-weight perturbations (for example, 0.5×, 1×, and 2× one coefficient at a time) on the same validation folds.

Table 3: Reproducibility status of architecture and training settings.

Setting	Reported or Recoverable Value
Input size	256 × 256 (2D) or 128 × 128 × 128 patches (3D)
Encoder channels	64, 128, 256, 512, 1024
Training episodes	500
Maximum refinements	$T = 4$
Reward moving average	Window = 10
Model selection	Five-fold validation within development data
Optimizer and actor/critic learning rates	Not retained
Batch size and discount factor	Not retained
Reward coefficients	Not retained
Mask-operator thresholds	Not retained

4 Dataset

ACRL-Net was evaluated on three public stroke-imaging resources with materially different acquisition characteristics: ISLES 2017 [8], ISLES 2018 [26], and AISD [10]. ISLES 2017 is an MRI benchmark. ISLES 2018 provides acute CT perfusion studies and perfusion maps with DWI-derived infarct-core references. AISD contains 397 NCCT scans with lesions manually contoured using subsequent DWI as the reference standard. Table 4 corrects the earlier MRI-only description and preserves the study-specific train/validation/test counts reported in the original manuscript. The archived study record does not include split files or a random seed, so these counts cannot be audited against the official release partitions.

Table 4: Dataset statistics, modalities, and evaluation partitions.

Dataset	Input Modality	Cases	Unit	Reference Annotation	Development	Validation	Test	Evaluation Role
ISLES 2017	Multi-spectral MRI	75	3D studies	Expert lesion/outcome masks	50	10	15	Development and held-out evaluation
ISLES 2018	CTP and perfusion maps	103	CTP studies	DWI-derived infarct-core masks	72	10	21	Study-reported modality evaluation
AISD	NCCT (DWI reference)	397	3D scans processed slice-wise	Manual NCCT contours supported by DWI	278	40	79	Study-reported robustness evaluation
Total	MRI and CT	575	Patient-level studies	Expert/manual masks	400	60	115	Study-reported retrospective evaluation

ISLES 2017 contains multi-spectral MRI volumes and expert stroke-lesion/outcome annotations [8]. ISLES 2018 instead targets infarct-core estimation from CT perfusion data acquired in acute stroke, with DWI-derived reference masks [26]; it should not be described as an MRI segmentation cohort. AISD contains 397 NCCT scans acquired within 24 h of symptom onset, with lesions delineated on NCCT using DWI obtained within the subsequent 24 h as reference [10]. These modality differences motivate modality-specific input handling and prevent a naive pooling interpretation. Table 4 reports the original manuscript's study-specific partitions (50/10/15, 72/10/21, and 278/40/79). These do not reproduce the official ISLES 2018 release organization of 63 development and 40 challenge-test cases or the AISD organization of 345 development and 52 test cases. Because the actual split lists were not supplied, conclusions are limited to the manuscript-reported retrospective evaluation rather than an official challenge-test claim.

Table 5 summarizes the modality-specific dataset characteristics and preprocessing protocol adopted in this study, including input modalities, reference annotations, normalization and augmentation procedures, input dimensions, evaluation metrics, validation strategy, and leakage-control considerations. The table emphasizes that preprocessing was adapted to the characteristics of each imaging modality while maintaining a consistent binary stroke-lesion segmentation objective across the evaluated datasets.

Table 5: Modality-specific dataset and preprocessing description.

Property	Description
Input modalities	ISLES 2017 multi-spectral MRI; ISLES 2018 CTP/perfusion maps; AISD NCCT
Output	Binary stroke-lesion or infarct-core mask, according to dataset definition
Annotation type	Expert voxel-wise/pixel-wise reference masks
Image preprocessing	Brain-region extraction where applicable, modality-specific intensity normalization/windowing, resampling, and cropping
Data augmentation	Random rotation, flipping, scaling, elastic deformation, and Gaussian noise where physically appropriate for the modality
Input size	256 × 256 (2D) or 128 × 128 × 128 (3D patches)
Segmentation classes	Background and stroke lesion
Evaluation metrics	Case-wise Dice and IoU, pooled Precision/Recall/F1, and HD95
Validation	Five-fold cross-validation was reported; exact fold assignments and selection rule were not retained
Leakage control	Patient/study is the intended partition unit; split files are required to verify that slices/patches never cross subsets

The datasets share a binary lesion-segmentation objective but not a common image distribution. The revised protocol therefore specifies modality-specific preprocessing: MRI brain extraction and sequence-wise normalization, CTP perfusion-map alignment and scaling, and NCCT brain-region cropping and intensity windowing, followed by the reported resampling and spatial cropping. The patient/study is the intended partition unit so that slices or patches from one subject do not cross subsets. However, the source package contains neither preprocessing code nor patient identifiers and split lists; exact execution and leakage control therefore remain to be verified in a reproducibility release.

5 Evaluation Metrics

The performance of the proposed ACRL-Net was quantitatively evaluated using widely adopted segmentation metrics that jointly assess region overlap, boundary accuracy, and classification performance. Since brain stroke lesion segmentation is characterized by highly imbalanced foreground and background regions, relying on a single evaluation criterion may provide an incomplete assessment of model performance. Therefore, six complementary metrics, namely Dice coefficient, Intersection over Union (IoU), Precision, Recall, Fl-score, and the 95th percentile Hausdorff Distance (HD95), were employed to comprehensively evaluate segmentation quality.

The Dice coefficient measures the degree of overlap between the predicted lesion mask M and the ground-truth annotation Y [27]. A higher Dice value indicates greater segmentation accuracy and better agreement with expert annotations. The Dice coefficient is defined as

$$\text{Dice} = \frac{2 \mid M \cap Y \mid}{\mid M \mid + \mid Y \mid} = \frac{2TP}{2TP + FP + FN} \tag{20}$$

where TP, FP, and FN denote the numbers of true-positive, false-positive, and false-negative pixels, respectively. Closely related to Dice, the Intersection over Union evaluates the ratio between the overlapping region and the union of the predicted and reference lesion masks [28]. It is formulated as

$$\mathrm{IoU} = \frac{|M \cap Y|}{|M \cup Y|} = \frac{TP}{TP + FP + FN} \tag{21}$$

Both Dice and IoU emphasize spatial overlap and are particularly suitable for evaluating semantic segmentation algorithms because they directly reflect the agreement between predicted and manually annotated lesions.

Pixel-wise classification performance was further assessed using Precision, Recall, and the F1-score. Precision evaluates the proportion of correctly predicted lesion pixels among all predicted lesion pixels, whereas Recall measures the ability of the model to detect all lesion regions present in the reference annotation [27]. Their mathematical definitions are

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{22}$$

and

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{23}$$

The F1-score represents the harmonic mean of Precision and Recall, providing a balanced assessment of false-positive and false-negative errors. It is computed as

$$F_1 = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{24}$$

Dice and the binary F1-score are algebraically identical when they are computed from the same confusion matrix. In this manuscript, Dice is reported as a mean of case-wise overlap scores, whereas Precision, Recall, and F1 are pooled over evaluated pixels/voxels. These metrics are particularly important for brain stroke lesion segmentation because inaccurate lesion detection may either overlook clinically significant infarct regions or incorrectly classify healthy tissues as lesions.

In addition to overlap-based measures, boundary accuracy was evaluated using the 95th percentile Hausdorff Distance (HD95), which quantifies the maximum boundary discrepancy between the predicted segmentation and the ground-truth contour while reducing sensitivity to extreme outliers [29]. Let B_M and B_Y denote the boundary point sets of the predicted and reference lesion masks, respectively. The directed Hausdorff distance is defined as

$$d(B_M, B_Y) = \max_{x \in B_M} \min_{y \in B_Y} \| x - y \|_2 \tag{25}$$

where $\| \cdot \|_2$ represents the Euclidean distance. The HD95 metric is subsequently computed as

$$\mathrm{HD95} = P_{95}\left(\max\left\{d(B_M, B_Y), d(B_Y, B_M)\right\}\right) \tag{26}$$

where $P_{95}(\cdot)$ denotes the 95th percentile operator. Unlike overlap-based metrics, HD95 directly evaluates boundary localization accuracy and is therefore highly suitable for assessing lesion contour refinement. Since the proposed ACRL-Net performs iterative boundary optimization through actor-critic reinforcement learning, HD95 serves as an essential indicator for measuring the effectiveness of the adaptive lesion

refinement strategy. Collectively, these complementary evaluation metrics provide a comprehensive and clinically meaningful assessment of segmentation performance by simultaneously measuring region overlap, classification accuracy, and boundary precision.

6 Results

This section reports the available aggregate quantitative results, qualitative examples, iterative refinement trajectories, policy behavior, and component ablation. The results describe the reported retrospective evaluation; they are not presented as statistically significant superiority because case-level prediction records were unavailable for paired testing.

6.1 Training Convergence

Fig. 5 reports episode return, actor loss, and critic loss for the retained training run. These curves are used as optimization diagnostics: they can show whether that run stabilizes, but they do not by themselves establish policy optimality, robustness across random seeds or hyperparameters, or generalization to new clinical data.

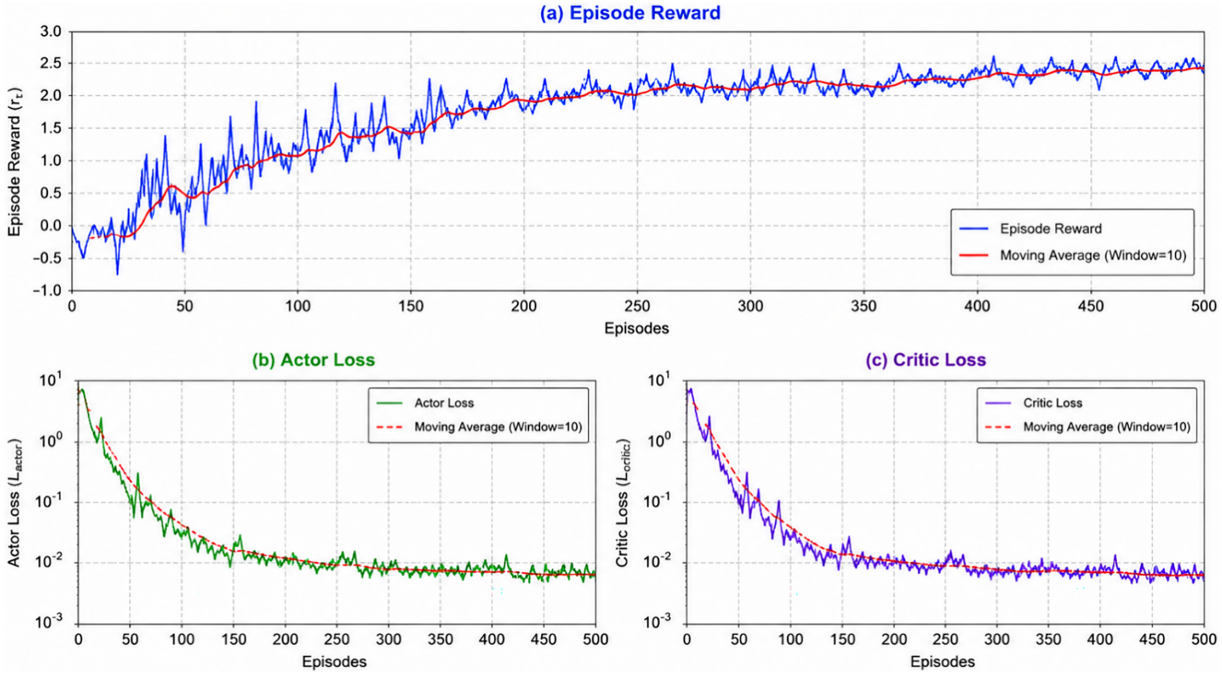


Figure 5: Training convergence (reward, actor loss, critic loss).

Fig. 5a shows the cumulative episode return, not a physical clinical quantity. The plotted value is the dimensionless sum of per-step weighted rewards over a refinement episode. Early negative values indicate that exploratory actions initially produced more weighted penalties than improvements. The moving average begins to flatten after roughly 250 episodes and ends near 2.4, which is consistent with stabilization under the chosen reward scaling; the absolute value 2.4 has no transferable meaning across coefficient choices. Training stability is interpreted jointly from the reward trend and the actor/critic loss trends, rather than from a predetermined target reward.

Fig. 5b illustrates that actor loss decreases during early policy optimization and stabilizes near 10^{-2}. Fig. 5c shows a comparable decline in critic loss toward 10^{-2}. These curves are consistent with

optimization stability for the reported run, but they do not establish robustness to different reward weights, random seeds, or learning rates.

6.2 Quantitative Comparison

Fig. 6 and Table 6 compare ACRL-Net with the included within-study CNN, transformer, and reinforcement-learning baselines under the reported protocol described in Section 4. Fig. 6a shows that ACRL-Net records Dice = 0.913 in the reported aggregate evaluation, compared with 0.712 for U-Net and 0.799 for Swin-UNETR. These correspond to descriptive absolute differences of 20.1 and 11.4 percentage points, respectively. Fig. 6b reproduces the source table's IoU entries of 0.422 for Swin-UNETR and 0.409 for RL-UNet. The original surrounding prose instead stated 0.822 and 0.809; those alternatives exceed the corresponding Dice values and were therefore removed from the claims. Because per-case outputs are unavailable, the table entries are retained rather than replaced by another inferred value, and the discrepancy requires author verification before submission.

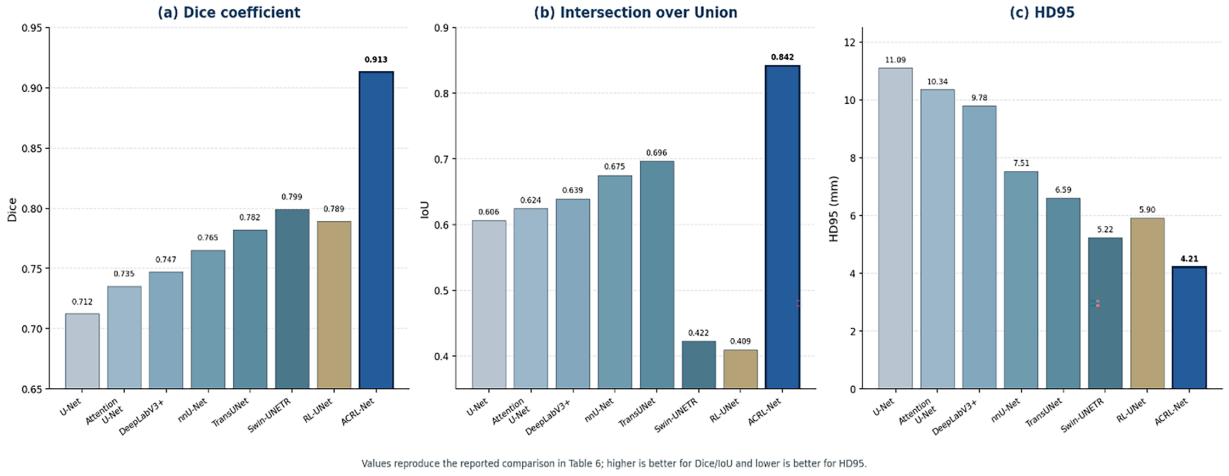


Figure 6: Quantitative segmentation performance (Dice, IoU, HD95).

Table 6: Reported within-study quantitative comparison of stroke-lesion segmentation methods.

Method	Backbone	Dice ↑	IoU ↑	Precision ↑	Recall ↑	F1-Score ↑	HD95 (mm) ↓
U-Net	CNN	0.712	0.606	0.831	0.795	0.812	11.09
Attention U-Net	CNN + Attention	0.735	0.624	0.849	0.823	0.836	10.34
DeepLabV3+	CNN + ASPP	0.747	0.639	0.861	0.834	0.847	9.78
nnU-Net	Self-configuring CNN	0.765	0.675	0.881	0.850	0.865	7.51
TransUNet	CNN + Transformer	0.782	0.696	0.896	0.869	0.882	6.59
Swin-UNETR	Swin Transformer	0.799	0.422	0.913	0.886	0.899	5.22

(Continued)

Table 6 (continued)

Method	Backbone	Dice ↑	IoU ↑	Precision ↑	Recall ↑	F1-Score ↑	HD95 (mm) ↓
RL-Unet	Value-based Reinforcement Learning	0.789	0.409	0.904	0.875	0.889	5.90
Proposed ACRL-Net	Actor-Critic Reinforcement Learning	0.913	0.842	0.926	0.901	0.913	4.21

Fig. 6c reports HD95 = 4.21 mm for ACRL-Net, compared with 11.09 mm for U-Net, 5.90 mm for RL-UNet, and 5.22 mm for Swin-UNETR. Table 6 also reports pooled Precision = 0.926, Recall = 0.901, and F1 = 0.913. Together, the reported aggregate values indicate improved overlap and boundary localization under the reported protocol, while the absence of case-level distributions limits inferential conclusions.

Table 6 presents the reported within-study comparison. ACRL-Net has the highest reported aggregate Dice and IoU and the lowest reported HD95 in this table. Relative to Swin-UNETR, the descriptive Dice difference is 11.4 percentage points and the HD95 reduction is 19.3%. Because Dice is case-averaged while F1 is pooled, their values are not required to coincide. The table supports a reported within-protocol performance difference but not a claim of audited fairness, universal superiority, or statistical significance.

Table 7 contextualizes ACRL-Net against recent stroke-lesion segmentation approaches evaluated on different datasets and protocols. The reported results demonstrate substantial variability among contemporary architectures, with Dice scores ranging from 0.590 to 0.8894 for existing methods. Attention-based, transformer-based, frequency-domain, and hybrid CNN-ViT architectures generally improve feature representation, although their performance remains dependent on dataset characteristics and evaluation settings. ACRL-Net reports a Dice coefficient of 0.913, IoU of 0.842, and HD95 of 4.21 mm, indicating favorable overlap and boundary localization within its evaluation protocol. Its distinguishing characteristic is the integration of actor-critic reinforcement learning with adaptive lesion refinement. However, these numerical values should be interpreted contextually rather than as a direct ranking because the studies employ different datasets, imaging modalities, preprocessing procedures, and experimental protocols.

Table 7: Contextual comparison with recent stroke-lesion segmentation studies (different datasets and protocols).

Ref.	Study/Method	Dataset	Main Technique	Dice ↑	IoU ↑	HD95 (mm) ↓
Ganie and Pacal, 2026 [10]	GBA-Net	ISLES 2024	Gated bottleneck and attention-driven architecture	0.7376	0.6364	10.94
Ahmed et al., 2024 [12]	Transformer-Augmented MRI Analysis	ATLAS 2022, ISLES 2015	CNN + Deformable Transformer	0.820	NR	NR

(Continued)

Table 7 (continued)

Ref.	Study/Method	Dataset	Main Technique	Dice ↑	IoU ↑	HD95 (mm) ↓
Pacal et al., 2025 [13]	FasterNet and Attention-Gated UNet	ISLES 2022	UNet + Attention	0.8676	0.7584	NR
Feng et al., 2026 [14]	FCSG-Net	ISLES 2022	Frequency-domain Cross-scale Gated Network	0.7284	NR	NR
Aras et al., 2026 [15]	MAGNUS	ISLES 2022	Multi-Attention Guided Network via CNN-ViT Fusion	0.833	0.717	NR
Ince et al., 2025 [16]	ConvNeXtV2 blocks in U-Net	ISLES 2022	ConvNeXtV2 blocks with GRN-based MLP into the standard U-Net architecture	0.8894	0.8015	NR
de la Rosa et al., 2025 [17]	DeepISLES	ISLES 2022	ensemble algorithm	0.73	NR	NR
Rahman et al., 2025 [18]	DenseNet121	ISLES 2022	DenseNet121 encoder with SelfONN CSCA U-Net decoder	0.8749	0.8350	NR
Alirr, 2025 [30]	Attention ResUnet	ISLES 2018	Attention ResUnet	0.59	NR	NR
Proposed	ACRL-Net	ISLES 2017, ISLES 2018, AISD	Actor-Critic Reinforcement Learning + Adaptive Lesion Refinement	0.913	0.842	4.21

Note: NR = not reproted.

6.3 Qualitative Lesion Segmentation Results

Qualitative visual assessment was performed to investigate the segmentation quality produced by different methods under various lesion characteristics and imaging conditions. Fig. 7 illustrates segmentation comparisons across representative cases from the ISLES 2017, ISLES 2018, and AISD datasets. The visual comparison includes the input modalities, expert annotations, and segmentation results obtained by seven competing methods together with the proposed ACRL-Net. False-positive regions and missed lesions are highlighted to facilitate visual assessment of segmentation quality.

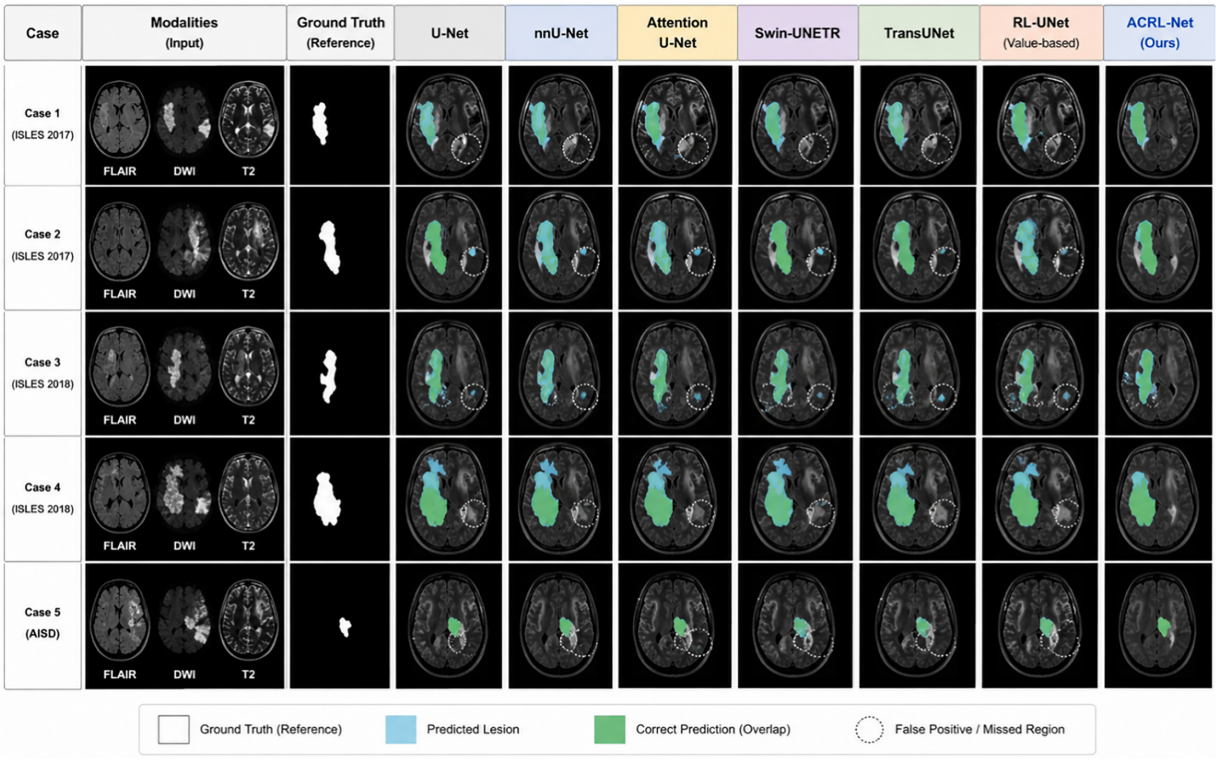


Figure 7: Qualitative lesion segmentation results.

In the selected Fig. 7 panels, several baseline masks visibly extend beyond or fall short of the expert contour, particularly near irregular boundaries and small components. These panel-level observations illustrate possible error patterns; they are not a dataset-wide comparison and do not establish that one architecture is consistently better without case-level quantitative analysis.

In the displayed examples, ACRL-Net produces masks closer to the expert annotations, with fewer isolated predictions and smoother contours than the illustrated baselines. These examples are qualitative and do not replace dataset-wide lesion-size or location-stratified analysis.

6.4 Adaptive Lesion Refinement

Fig. 8 visualizes the initial prediction and four successive mask states for three selected cases. The panels are used to illustrate the reported refinement sequence rather than to claim an optimal mask at the final step. Because the underlying action log and composite reward components were not retained, the displayed changes cannot be attributed quantitatively to individual action types.

Across the displayed cases, the initial mask contains visible boundary mismatches, while later masks show larger overlap with the reference and fewer isolated disagreements. The accompanying Dice, IoU, and HD95 trajectories describe these three examples only; they do not establish monotonic improvement for every case in the evaluated datasets.

The component trajectories in Fig. 8 show increasing Dice/IoU and decreasing HD95 over four refinements for three illustrative cases. For example, Case 1 changes from Dice = 0.42 at iteration 0 to Dice = 0.89 at iteration 4, while HD95 changes from 42.7 to 8.9 mm. The figure displays the segmentation mask and the available overlap/boundary components after each step. Composite per-step reward values cannot be

reconstructed because the FP/FN components and reward coefficients were not retained; inferring them from Dice, IoU, and HD95 alone would be invalid.

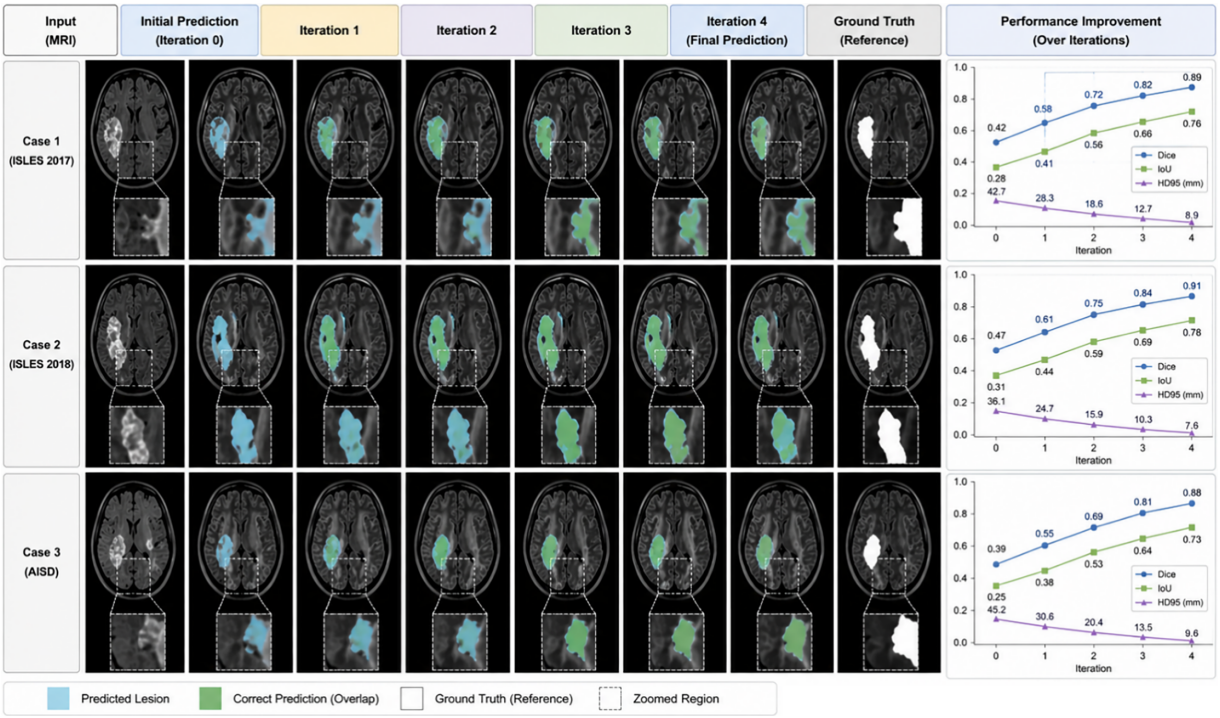


Figure 8: Adaptive lesion refinement process.

6.5 Policy Behavior Analysis

Fig. 9 summarizes five nonterminal refinement actions: expand, shrink, refine boundary, smooth/focus, and keep. Stop is a sixth terminal action in the policy but is excluded from the normalized nonterminal action percentages because an episode ends when it is selected. The plotted distributions therefore describe the composition of actions taken before termination.

In the displayed action summary, expansion decreases from approximately 35% in the first iteration to 10% in the last, while refine-boundary increases from approximately 25% to 38%. Smooth/focus actions also increase in later iterations. These frequencies describe the retained visualization; without per-case action logs or counterfactual analysis, they do not establish why an action was selected or that the policy objective changed causally.

The action distributions shift from expansion in early iterations toward boundary refinement and keep actions later in the episode. This pattern is consistent with coarse-to-fine correction in the displayed run, but it is descriptive and does not by itself demonstrate optimality of the learned policy.

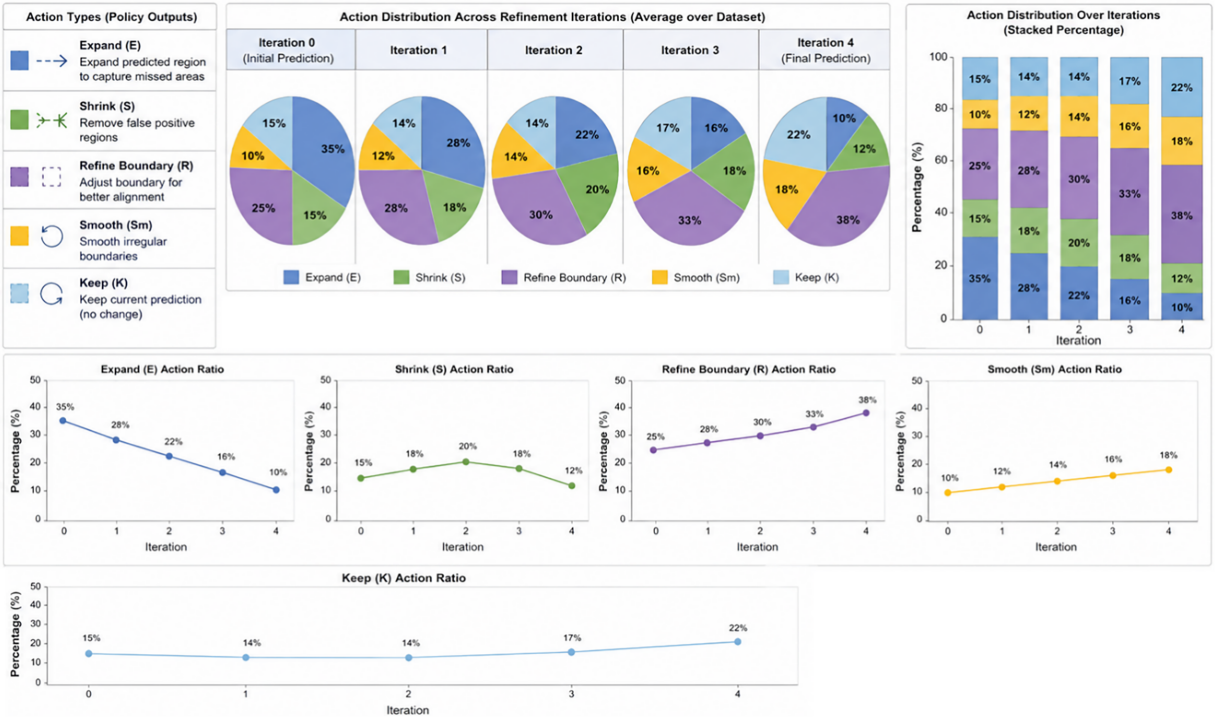


Figure 9: Policy action distribution during refinement.

6.6 Clinical Case Studies

Fig. 10 presents three selected cases as qualitative error overlays. Each panel includes the source image, expert annotation, initial mask, refined mask, and residual false-positive/false-negative regions. The figure illustrates possible behavior under different lesion appearances; it is not an external clinical validation set and cannot establish practical applicability by itself.

In the displayed panels, the initial masks cover much of the main lesion but retain local boundary disagreement and isolated residual errors. The refined masks appear closer to the expert contours in these examples, and the overlays show where disagreement remains. This visual evidence is descriptive; it does not demonstrate elimination of artifacts or a dataset-wide performance effect.

These examples suggest that the policy can refine lesions with different sizes and locations, but three selected cases cannot establish generalization to routine clinical practice. They should be interpreted alongside the failure modes and validation limitations in Section 6.8.

6.7 Ablation Analysis

Table 8 reports a progressive component ablation for the retained aggregate evaluation: the configuration is expanded from the ACRL encoder-decoder backbone through encoder, decoder, actor, critic, and multi-step refinement additions. Because only this cumulative path is available, the table shows associations with each successive configuration but does not isolate every component independently or quantify interactions as a full factorial ablation would.

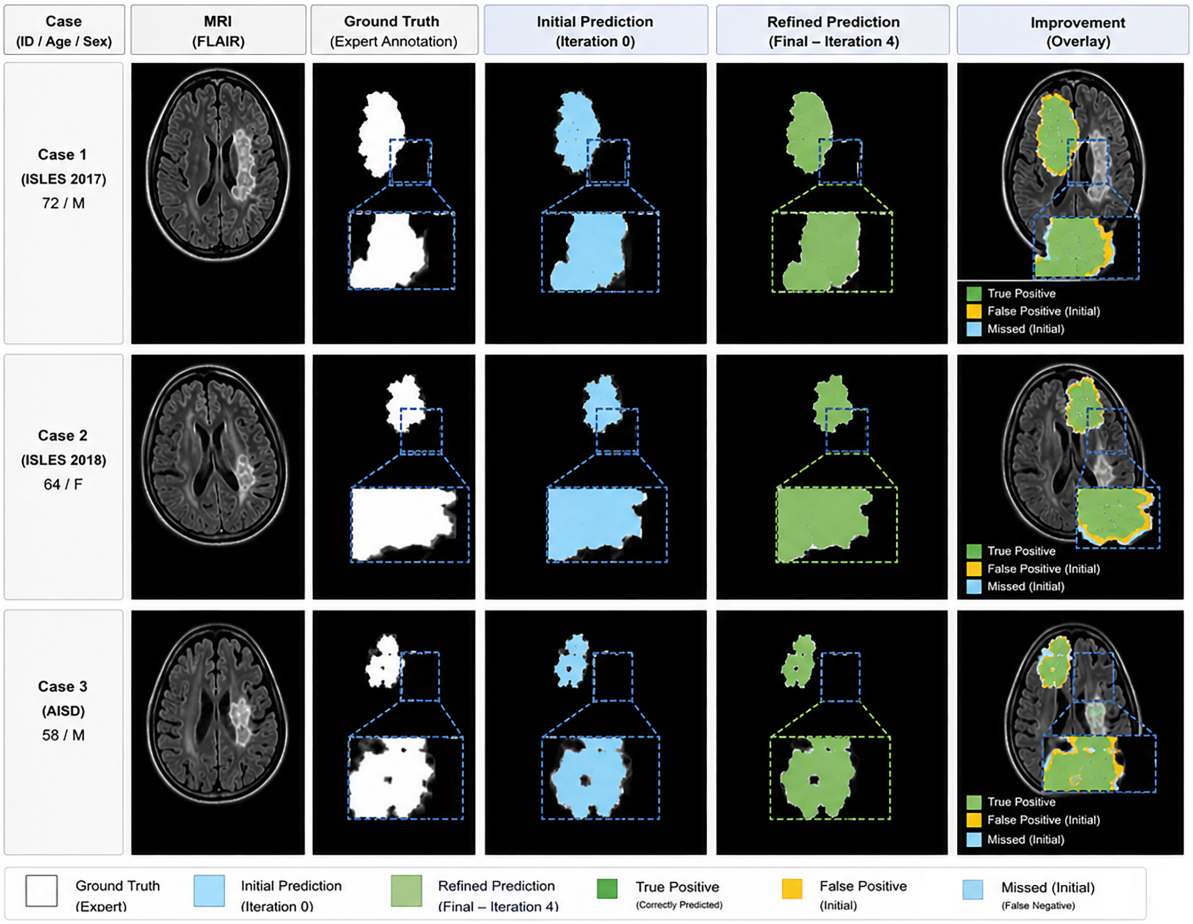


Figure 10: Clinical case studies and residual error overlays.

Table 8: Component-wise ablation study of ACRL-Net.

Configuration	Feature Encoder	Segmen-tation Decoder	Actor-Critic RL Agent	Adaptive Lesion Refinement	Dice ↑	IoU ↑	Preci-sion ↑	Recall ↑	HD95 (mm) ↓
ACRL backbone (encoder-decoder; no RL)	Baseline	Baseline	None	None	0.865	0.775	0.881	0.850	7.51
+ Improved Feature Encoder	Improved	Baseline	None	None	0.883	0.798	0.897	0.871	6.48
+ Adaptive Decoder	Improved	Adaptive	None	Decoder-only	0.894	0.814	0.909	0.885	5.82
+ Actor Network Only	Improved	Adaptive	Actor only	Policy refinement	0.901	0.824	0.915	0.892	5.23
+ Actor-Critic RL Agent	Improved	Adaptive	Actor + critic	Actor-critic refinement	0.908	0.835	0.922	0.896	4.67
Proposed ACRL-Net	Improved	Adaptive	Actor + critic	Multi-step refinement	0.913	0.842	0.926	0.901	4.21

The first row of Table 8 is the ACRL encoder-decoder backbone without actor-critic refinement. This clarification resolves the apparent metric mismatch between the two tables. Within the progressive ablation, adding the improved encoder, adaptive decoder, actor, critic, and multi-step refinement is associated with successive descriptive improvements in Dice and HD95.

The complete configuration reports Dice = 0.913, IoU = 0.842, Precision = 0.926, Recall = 0.901, and HD95 = 4.21 mm. Table 8 supports the contribution of the major architectural blocks under a progressive ablation. It is not a full factorial study and does not test each reward term or coefficient separately; no reward-weight robustness claim is made.

6.8 Clinical Interpretation, Failure Modes, and Computational Scope

The qualitative panels suggest that refinement is most useful when the initial predictor captures the main lesion but leaves boundary under-segmentation or isolated false-positive regions. Large, connected, higher-contrast lesions provide a stable region on which expansion, shrinkage, and boundary actions can operate. Small or disconnected lesions are more vulnerable: shrink can remove a true small component, expansion can merge a lesion with adjacent structures, and an early Stop decision can preserve a missed lesion. Lesions near ventricles, cortex, skull-adjacent artifacts, or low-contrast NCCT tissue may also confuse the uncertainty map.

No lesion-volume, anatomical-location, or scanner-stratified metric vectors were included in the retained study record, so quantitative subgroup analysis is not reported. The method has not been tested on an independent prospective or local multi-center cohort outside the datasets used for model development and evaluation. Real-world generalization therefore remains unproven, especially under protocol shift, motion, registration error, incomplete modalities, and rare very small lesions.

The actor-critic design adds iterative cost after the initial segmentation. The critic contributes to training but not inference. If C_E, C_D, C_A, and C_g denote encoder, decoder, actor-state, and mask-operator costs, the worst-case inference cost is $C_E + C_D + 4(C_A + C_g)$. Parameter count, FLOPs, wall-clock latency, memory, and device-specific throughput were not logged in the retained study record and cannot be reconstructed without code or a checkpoint. Table 9 reports only the analytically defensible execution structure.

Table 9: Analytical inference structure and availability of efficiency measurements.

Component	Inference Use	Analytical Contribution	Measured Value Available?
Feature encoder	Once	C_E; produces shared multi-scale features	No parameter/FLOP record
Initial decoder	Once	C_D; produces M_0 and probability map	No latency record
Actor/state encoder	Up to four times	Sum from $t = 0$ to $T - 1$ of C_A	No latency record
Discrete mask operator	Up to four times	C_g, approximately linear in evaluated voxels	No profiling record
Critic	Not used	Zero inference contribution	Training-only
Worst-case total	$T = 4$	$C_E + C_D + 4\left(C_A + C_g\right)$	Requires code/checkpoint benchmarking

7 Discussion

The reported aggregate results indicate that sequential post-prediction correction can complement a supervised encoder-decoder. The plausible mechanism is targeted reuse of the current mask, uncertainty, and shared features: expansion addresses under-segmentation, shrink suppresses false-positive extent, and boundary actions concentrate on ambiguous contours. The qualitative trajectories are consistent with this mechanism, although they do not establish causality or statistical significance without retained case-level outputs.

The contribution should be interpreted narrowly. Reinforcement learning, actor-critic optimization, and iterative medical-image refinement already exist [7,23–25]. ACRL-Net combines them in a stroke-focused pipeline with an uncertainty-aware state, explicit region/boundary actions, and a finite stopping policy after an initial multi-scale prediction. This positioning distinguishes architectural integration from a claim that reinforcement learning-based refinement is entirely new.

Table 8 associates each progressive architectural addition with improved reported aggregate metrics, while Table 6 provides the reported within-study baseline comparison. However, the ablation is not full factorial, and no retained experiment isolates every reward term, reward coefficient, learning rate, or stopping rule. The contribution of those design choices therefore remains uncertain.

Several limitations constrain interpretation. First, up to four actor evaluations and discrete mask updates increase inference cost, but parameter count, FLOPs, latency, memory, and throughput were not available. Second, per-case and per-fold metric vectors were not retained, preventing confidence intervals, paired Wilcoxon tests, Friedman analysis, and quantitative lesion-size/location stratification. The source prose and Table 6 also conflict for two baseline IoU values; the table values are retained and the discrepancy requires verification from the original predictions. Third, the selected optimizer settings, reward coefficients, and mask thresholds were not recoverable, and empirical reward-weight sensitivity was not available. Dice and IoU are correlated reward terms and may double-weight overlap. Fourth, the evaluation does not constitute independent prospective multi-center validation, and the qualitative panels cannot establish clinical generalization. Finally, discrete region operators can erase small true lesions, merge neighboring structures, oscillate at uncertain boundaries, or stop prematurely. These issues must be addressed before clinical deployment.

8 Conclusion

This study presented ACRL-Net, an adaptive segmentation framework that integrates multi-scale stroke-lesion representation learning with uncertainty-aware actor-critic reinforcement learning to progressively improve lesion delineation. Unlike conventional segmentation approaches that produce a prediction through a single feed-forward process, ACRL-Net incorporates a structured refinement mechanism in which an initial segmentation is iteratively corrected through region- and boundary-oriented actions selected according to the current segmentation state and uncertainty information. Under the reported retrospective evaluation protocol, the complete framework achieved a Dice coefficient of 0.913, IoU of 0.842, Precision of 0.926, Recall of 0.901, and HD95 of 4.21 mm, demonstrating favorable overlap accuracy and boundary localization. The progressive ablation analysis further supported the complementary contributions of the improved feature encoder, adaptive segmentation decoder, actor-critic agent, and multi-step refinement strategy. Importantly, the critic is required only during training, whereas inference is performed using the learned actor and is restricted to no more than four refinement actions following the initial prediction, thereby limiting iterative processing. Nevertheless, the reported findings should be interpreted as evidence of feasibility within the evaluated retrospective setting rather than proof of universal superiority or immediate clinical readiness. Further research should therefore provide complete implementation details and

hyperparameters, release case-level predictions where permitted, investigate sensitivity to reward coefficients and stopping criteria, conduct paired statistical significance analyses, and report measured computational efficiency. Additional evaluation should characterize failure modes according to lesion size, anatomical location, imaging quality, and modality. Finally, prospective validation using independent multi-center cohorts is required to establish generalizability, robustness, and potential clinical utility across heterogeneous stroke imaging environments.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by the Science Committee of the Ministry of Higher Education and Science of the Republic of Kazakhstan within the framework of grant AP23489899.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Zhanar Azhibekova and Akbayan Aliyeva; methodology development: Zhanar Azhibekova, Daniyar Rakhymberdiyev and Meruert Dauletova; data collection and preprocessing: Akbayan Aliyeva and Daniyar Rakhymberdiyev; software implementation and experiments: Zhanar Azhibekova and Daniyar Rakhymberdiyev; analysis and interpretation of results: Zhanar Azhibekova, Akbayan Aliyeva, Daniyar Rakhymberdiyev and Meruert Dauletova; draft manuscript preparation: Zhanar Azhibekova and Akbayan Aliyeva; manuscript review and editing: Daniyar Rakhymberdiyev and Meruert Dauletova. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data openly available in a public repository. The data that support the findings of this study are openly available in [Ischemic Stroke Lesion Segmentation (ISLES) 2017 Dataset] at [https://www.isles-challenge.org/ISLES2017/]; [Ischemic Stroke Lesion Segmentation (ISLES) 2018 Dataset] at [https://www.isles-challenge.org/ISLES2018/]; [Acute ischemic stroke dataset (AISD) Dataset] at [https://github.com/griffinliang/aisd].

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Nomenclature

ACRL-Net	Actor-Critic Reinforcement Learning-Guided Adaptive Lesion Refinement Network
RL	Reinforcement Learning
CNN	Convolutional Neural Network
MRI	Magnetic Resonance Imaging
FLAIR	Fluid-Attenuated Inversion Recovery
DWI	Diffusion-Weighted Imaging
ADC	Apparent Diffusion Coefficient
IoU	Intersection over Union
HD95	95th Percentile Hausdorff Distance
TP	True Positive
FP	False Positive
FN	False Negative
SE	Squeeze-and-Excitation
Dice	Dice Similarity Coefficient
AISD	Acute Ischemic Stroke Dataset
ISLES	Ischemic Stroke Lesion Segmentation Challenge

References

1. Kumar RR, Priyadarshi R. Denoising and segmentation in medical image analysis: a comprehensive review on machine learning and deep learning approaches. Multimed Tools Appl. 2025;84(12):10817–75. doi:10.1007/s11042-024-19313-6.
2. Zhou Y, Yang X, Yin J, Liu S. Research on multi-scale feature fusion network algorithm based on brain tumor medical image classification. Comput Mater Contin. 2024;79(3):5313–33. doi:10.32604/cmc.2024.052060.
3. Shangguan H, Zhao H, Zhang X, Yang J, Sun Y, Li Z, et al. A multi-stage progressive feature fusion network for brain stroke lesion segmentation. Biomed Signal Process Control. 2026;112(1):108581. doi:10.1016/j.bspc.2025.108581.
4. Meng W, Liu S, Wang H. AFC-Unet: attention-fused full-scale CNN-transformer unet for medical image segmentation. Biomed Signal Process Control. 2025;99(7):106839. doi:10.1016/j.bspc.2024.106839.
5. Agarwal R, Chowdhury A, Chatterjee RK, Chel H, Murmu C, Murmu N, et al. Deep quasi-recurrent self-attention with dual encoder-decoder in biomedical CT image segmentation. IEEE J Biomed Health Inform. 2024;28(12):7195–205. doi:10.1109/JBHI.2024.3447689.
6. Thakur P, Talwandi NS. Deep reinforcement learning in healthcare and bio-medical applications. In: Proceedings of the 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT); 2024 Feb 9–10; Greater Noida, India. doi:10.1109/ic2pct60090.2024.10486549.
7. Tian Z, He Y, Sun L, Li Y, Du S. RCI-Seg: robust click-based interactive segmentation framework with deep reinforcement learning for biomedical images. Neurocomputing. 2024;601(7):128184. doi:10.1016/j.neucom.2024.128184.
8. Winzeck S, Hakim A, McKinley R, Pinto JAADSR, Alves V, Silva C, et al. ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. Front Neurol. 2018;9:679. doi:10.3389/fneur.2018.00679.
9. Hernandez Petzsche MR, de la Rosa E, Hanning U, Wiest R, Valenzuela W, Reyes M, et al. ISLES 2022: a multi-center magnetic resonance imaging stroke lesion segmentation dataset. Sci Data. 2022;9(1):762. doi:10.1038/s41597-022-01875-5.
10. Ganie S, Pacal I. GBA-NET: a gated bottleneck and attention-driven architecture for robust ischemic stroke segmentation across CT and DWI. Appl Comput Math. 2026;25:52–78. doi:10.30546/1683-6154.25.1.2026.52.
11. Liew SL, Tavenner BP, Donnelly MR, Zavaliangos-Petropulu A, Jeong JN, Barisano G, et al. A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. Sci Data. 2022;9(1):320. doi:10.1038/s41597-022-01401-7.
12. Ahmed R, Al Shehhi A, Werghi N, Seghier ML. Segmentation of stroke lesions using transformers-augmented MRI analysis. Hum Brain Mapp. 2024;45(11):e26803. doi:10.1002/hbm.26803.
13. Pacal I, Algarni A, Bayram B, Ince S. FA-UNet: a FasterNet and attention-gated hybrid network for precise ischemic stroke segmentation. J Integr Neurosci. 2025;24(10):40100. doi:10.31083/JIN40100.
14. Feng Y, Xie P, Chen S, Zhu Y, Wang X, Wu Z, et al. FCSG-Net: frequency-domain cross-scale gated network for stroke segmentation. Biomed Signal Process Control. 2026;113(2):109242. doi:10.1016/j.bspc.2025.109242.
15. Aras E, Kayikçioglu T, Aras S, Merd N. MAGNUS: multi-attention guided network for unified segmentation via CNN-ViT fusion. IEEE Access. 2026;14:12775–96. doi:10.1109/access.2026.3656667.
16. Ince S, Kunduracioglu I, Algarni A, Bayram B, Pacal I. Deep learning for cerebral vascular occlusion segmentation: a novel ConvNeXtV2 and GRN-integrated U-Net framework for diffusion-weighted imaging. Neuroscience. 2025;574(22):42–53. doi:10.1016/j.neuroscience.2025.04.010.
17. de la Rosa E, Reyes M, Liew SL, Hutton A, Wiest R, Kaesmacher J, et al. DeepISLES: a clinically validated ischemic stroke segmentation model from the ISLES'22 challenge. Nat Commun. 2025;16(1):7357. doi:10.1038/s41467-025-62373-x.
18. Rahman A, Chowdhury MEH, Ibne Wadud MS, Sarmun R, Mushtak A, Zoghoul SB, et al. Deep learning-driven segmentation of ischemic stroke lesions using multi-channel MRI. Biomed Signal Process Control. 2025;105(20):107676. doi:10.1016/j.bspc.2025.107676.
19. Xiao L, Song J, Xie X, Fan C. Enhanced medical image segmentation using U-Net with residual connections and dual attention mechanism. Eng Appl Artif Intell. 2025;153(5):110794. doi:10.1016/j.engappai.2025.110794.

20. Xiao L, Zhou B, Xie X, Fan C. Efficient brain tumor diagnosis networks by fusing tensor residual attention and superpixel map features. Eng Appl Artif Intell. 2026;181(11):115333. doi:10.1016/j.engappai.2026.115333.
21. Kumar A, Upadhyay N, Ghosal P, Chowdhury T, Das D, Mukherjee A, et al. CSNet: a new DeepNet framework for ischemic stroke lesion segmentation. Comput Methods Programs Biomed. 2020;193(1):105524. doi:10.1016/j.cmpb.2020.105524.
22. Kumar A, Ghosal P, Kundu SS, Mukherjee A, Nandi D. A lightweight asymmetric U-Net framework for acute ischemic stroke lesion segmentation in CT and CTP images. Comput Methods Programs Biomed. 2022;226(1):107157. doi:10.1016/j.cmpb.2022.107157.
23. Liu Y, Yuan D, Xu Z, Zhan Y, Zhang H, Lu J, et al. Pixel level deep reinforcement learning for accurate and robust medical image segmentation. Sci Rep. 2025;15(1):8213. doi:10.1038/s41598-025-92117-2.
24. Ullah S, Gabin K, Lee D, Tufail HZ, Lee K, Park S, et al. 3D segmentation of intracerebral hemorrhage in brain CT using enhanced UNet transformer via reinforcement learning. In: Proceedings of the Medical Imaging with Deep Learning 2026; 2026 Jul 8–10; Taipei, Taiwan.
25. Amiri S, Vrtovec T, Mustafaev T, Deufel CL, Thomsen HS, Sillesen MH, et al. Reinforcement learning-based anatomical maps for pancreas subregion and duct segmentation. Med Phys. 2024;51(10):7378–92. doi:10.1002/mp.17300.
26. ISLES Challenge Organizing Committee. Ischemic stroke lesion segmentation challenge 2018 [Internet]. [cited 2026 Aug 4]. Available from: https://www.isles-challenge.org/ISLES2018/.
27. Tursynova A, Omarov B, Tukenova N, Salgozha I, Khaaval O, Ramazanov R, et al. Deep learning-enabled brain stroke classification on computed tomography images. Comput Mater Contin. 2023;75(1):1431–46. doi:10.32604/cmc.2023.034400.
28. Samak ZA. Multi-type stroke lesion segmentation: comparison of single-stage and hierarchical approach. Med Biol Eng Comput. 2025;63(4):975–86. doi:10.1007/s11517-024-03243-4.
29. Guzzi L, Zuluaga MA, Taiello R, Lareyre F, Di Lorenzo G, Goffart S, et al. Regional Hausdorff distance losses for medical image segmentation. In: Machine learning in medical imaging. Berlin/Heidelberg, Germany: Springer; 2026. p. 63–73. doi:10.1007/978-3-032-09513-8_7.
30. Alirr OI. Ischemic stroke lesion core segmentation from CT perfusion scans using attention ResUnet deep learning. J Imaging Inform Med. 2025;38(6):3507–16. doi:10.1007/s10278-025-01407-8.