EDA-UNet++: EfficientNet-B4 Dual-Attention U-Net++ for Lesion-Wise DR Segmentation

Akansha Gupta*, Arjun Singh Rawat, and Gunjan Rehani

Department of Computer Science & Engineering, National Institute of Technology Delhi, New Delhi, India
*242210003@nitdelhi.ac.in, arjunsinghrawat@nitdelhi.ac.in, gunjan.rehani@nitdelhi.ac.in

Abstract. For automated screening and clinical decision support, accurate segmentation of diabetic retinopathy (DR) lesions is crucial. However, this is still difficult because of lesion variability, low contrast structures, and semantic discrepancies between encoder and decoder representations in convolutional networks. In order to promote accurate retinal lesion delineation, this study presents EDA-UNet++, a dual-attention improved U-Net++ framework. Three essential elements are integrated in the suggested architecture: dense nested skip pathways for progressive semantic alignment, an EfficientNet-B4 encoder for enhanced feature preservation during downsampling, and spatial–channel dual attention refinement to highlight lesion-relevant responses while reducing noise. All baseline models, including U-Net, U-Net++, and Attention U-Net, used the same encoders to guarantee a fair comparison. Experimental assessment on the DDR dataset shows better macro and lesion-wise segmentation performance, with the best validation Dice score among comparison models and significant improvements for exudates and soft exudates. Stable generalization with no loss in performance is confirmed by cross-dataset evaluation on IDRiD. According to computational research, significant gains are made with little extra overhead, confirming their applicability for extensive clinical screening applications.

Keywords: Multi-scale lesion detection, UNet-family, EfficientNet-B4, nested skips, dual attention mechanisms, cross-dataset validation

1 Introduction

Diabetic retinopathy is one of the most common complications of diabetes and remains a major cause of vision impairment worldwide. As diabetes prevalence continues to increase, accurate identification of retinal lesions such as microaneurysms, hemorrhages, exudates, and soft exudates becomes increasingly important. This is vital for ensuring timely interventions, assessing risk levels, and developing tailored treatment strategies. Reliable lesion segmentation supports

* Corresponding author: 242210003@nitdelhi.ac.in

A. Agnihotri et al. (eds.), *Proceedings of the Conference on Bridging Engineering Disciplines with AI and Machine Learning (BEDAIML 2026)*, Advances in Intelligent Systems Research 209,
https://doi.org/10.2991/978-94-6239-697-5_22

automated screening systems and assists clinicians in monitoring disease progression and evaluating treatment outcomes, ultimately reducing the burden of vision loss.

U-Net and its variants are widely adopted in medical image segmentation because the encoder–decoder design with skip connections helps retain spatial details during feature reconstruction [4,16,8]. Although these models are excellent, they struggle with fine-grained DR lesion segmentation. However, repeated down-sampling operations within the encoder gradually reduce spatial resolution. As a result, subtle lesions with low contrast may lose important visual cues before reaching the decoder stage, as a result of this, which makes it difficult to accurately recover tiny structures [1,2,3,4,5].In the conventional U-Net architecture, skip connections merge shallow encoder features with deeper decoder representations that differ in semantic abstraction. Integration and dissemination of features are not as effective.

Even while there has been improvement, there are still significant gaps. U-Net++ reduces the difference in meaning between what the encoder and decoder produce while maintaining structure by integrating features at different scales via dense nested skip connections [17,18]. Attention gates in Attention U-Net improve spatial localization, especially for microaneurysms [21]. For efficient comparison, further techniques include boundary-focused refinements, contextual fusion, multi-resolution calibration, and simplified baselines (fu2022, wang2021, avazov2024, lu2022, halfunet). However, U-Net++ relies on straightforward feature concatenation in its nested connections without mechanisms to selectively suppress irrelevant or redundant features, potentially propagating noise to the decoder [19]. Attention U-Net focuses mainly on spatial “where” attention but amplifies all feature channels equally, overlooking semantic mismatches and failing to prioritize lesion specific channels [11]. Reviews highlight recurring limitations in handling small scale objects, maintaining structural fidelity, and achieving consistent performance across U-Net derivatives [8].

To overcome these limitations, we propose the EDA-UNet++ architecture, a fresh architecture that smartly integrates three key innovations to enhance fine-grained DR lesion segmentation. Initially, our goal is to minimize the feature degradation caused by downsampling. We utilize the EfficientNet-B4 encoder, recognized for its compound scaling and capability to maintain medical image details even with some resolution loss [23,24,25]. Secondly, the use of dense nested skip connections, drawing inspiration from cutting-edge fusion techniques, boosts semantic alignment and strengthens the interactions between the encoder and decoder [18]. In a dual attention mechanism, spatial attention boosts localization and clarifies boundaries, while channel attention focuses on key representations, minimizes noise, and emphasizes relevant cues for lesions during decoding [12,13,15].

The identical EfficientNet-B4 encoder is applied to the baselines (U-Net, U-Net++, Attention U-Net) to highlight design advantages for a fair assessment. During comprehensive DDR testing, EDA-UNet++ surpassed baseline models

in segmentation metrics like the Dice score and IoU. The power and breadth of applicability of IDRiD are validated through cross-dataset testing.

2 Related Work

Recent developments in CNNs and U-Net architectures focus on keeping spatial information intact with pooling-free designs that maintain resolution while extracting features [1].Lightweight convolutional operations combined with depth-wise separable filters have been explored to reduce computational cost while preserving segmentation accuracy [6]. The dynamic modification of feature map weighting by residual connections and channel recalibration procedures adds to the complexity of contextual modeling [7]. The complexity of contextual modeling is increased by the dynamic adjustment of feature map weighting by residual connections and channel recalibration techniques [7]. Many extensions of the U-Net architecture focus on dense feature aggregation and redesigned skip pathways to improve multi-scale contextual representation [8]. Nevertheless, semantic inconsistencies between encoder and decoder representations remain a key challenge, the prioritization of adaptive features, and the effective management of fine-grained lesions. Attention-based segmentation techniques leverage dual cross attention to enhance the collaboration of features in spatial contexts [12]. Employing spatial-channel dual attention in transformer-based approaches enhances localization, channel discrimination, and long-range context modeling significantly. This allows for precise boundary delineation [15]. Conversely, these options can be costly to operate, challenging to integrate into current pipelines, and still not fine-tuned for tasks that require distinguishing multiple lesions in the retina.

Comparative investigations in MRI reveal that different designs perform differently, although benchmarking tests on U-Net variations demonstrate that they perform well for stroke lesion segmentation [16]. These experiments demonstrate the effectiveness of cross-modality adaptation and encoder-decoder synergy. However, reliable recovery of small details is hindered by susceptibility to downsampling, persisting semantic gaps, and intrinsic structural restrictions.

Because U-Net concentrates on tiny lesions, it is highly effective at detecting microaneurysms in domain-specific DR segmentation, demonstrating the significance of attention for early detection [21]. However, there is little architectural variation and the scope is restricted to single lesion scenarios.

Adding EfficientNet to encoders makes feature extraction better, makes representations deeper, and works well with multi-scale attention in U-Net frameworks [23,24,25]. But just upgrading the encoder isn't enough because it doesn't have better skip paths or joint spatial-channel prioritization.

Previous research has looked at efficiency, attention, context, and encoders separately, without putting them all together in a single framework for semantic alignment, careful lesion preservation, adaptive prioritization, and strong encoder-decoder integration. This is why we want to make EDA-UNet++.

3 Methodology

3.1 Hardware and Software Stack

For each experiment, an NVIDIA A100 GPU running the PyTorch framework with automatic mixed precision enabled was used. This setup maintained repeatable and consistent numbers while guaranteeing training success.

3.2 Dataset

The study separated lesions associated with diabetic retinopathy using two publicly available retinal fundus imaging datasets, DDR and IDRiD. The DDR dataset consists of 757 pictures. They are as follows: 225 for testing, 149 for validation, and 383 for training. We labeled these images using five classes: microaneurysms, exudates, soft exudates, hemorrhages, and background. To increase validation reliability, a study-specific split of 43 training, 11 validation, and 27 test images was employed rather than splitting the IDRiD dataset into 54 training and 27 test images. Numerous lesion types found in these datasets are crucial for multiclass segmentation.

3.3 Preprocessing & Augmentation Pipeline

The training patches were cropped to 768×768 pixels, and the photos were scaled to a typical input resolution of 1024 × 1024 pixels. The average and variability statistics from ImageNet were followed by normalization to match the backbone of the EfficientNet-B4 encoder. This made sure that the distribution of features was the same across all models. We only did data augmentation during training. It included changing the brightness and contrast at random, flipping the data horizontally and vertically, rotating it, and using contrast-limited adaptive histogram equalization to make the model more general.

To better capture small lesions and address severe class imbalance, a patch-based lesion-centered sampling strategy was adopted, a patch-based lesion-centered sampling technique was employed. Lesion-centered crops had a 0.85 probability of being selected, and sliding window extraction with a 0.5 overlap ratio was used to create 6000 samples each epoch. While maintaining the same level of information, our approach accorded uncommon diseased regions greater weight.

3.4 Training Configuration and Hyperparameters

We went with the AdamW optimizer and learning rate set to 2×10^{-4} and a CosineAnnealingWarmRestarts scheduler (T_0 = 5 × steps per epoch, T_{mult} = 2). During 110–140 epochs, the batch size was 4 and the gradient clipping was set to 1.0. To save time, mixed precision was turned on. Class weights were used to give more weight to lesion categories that weren't well represented. For example, background lesions got 1.2, microaneurysms got 1.4, exudates got 3.0, soft exudates got 1.0, and hemorrhages got 1.4. This helped guide optimization

toward clinically important features. For fair comparison, all evaluated models (U-Net, U-Net++, Attention U-Net, and EDA-UNet++) were trained using the same dataset splits, preprocessing pipeline, input resolution, augmentation strategy, optimizer settings, learning-rate schedule, and evaluation metrics.

3.5 Loss Functions and Evaluation Metrics

We used a variety of loss functions generated from raw logits to train the model in order to overcome issues with overlapping areas, class imbalances, and challenging pixels. This included Dice loss (weight 0.3), Tversky loss (α = 0.3, β = 0.7; weight 0.3), Focal Tversky loss (γ = 1.33; weight 0.2), and weighted Cross Entropy loss (weight 0.2).By adding class-specific weights for background (1.2), microaneurysms (1.4), exudates (3.0), soft exudates (1.0), and hemorrhages (1.4), the Cross Entropy component lessened the focus on the more common background while enhancing signals from less common lesions. In addition to macro averages across lesion classes, performance was evaluated using per-lesion IoU, Dice, Precision, Recall, and F1 scores.

3.6 Proposed EDA-UNet++ Architecture

Figure 1 illustrates the proposed EDA-UNet++ architecture. The model follows a nested encoder–decoder topology inspired by U-Net++, but introduces three structural refinements to enhance fine-grained lesion recovery.

First, feature extraction is performed using an EfficientNet-B4 backbone, which generates multi-scale hierarchical representations denoted as $X^{i,0}$ across encoder depths. This encoder preserves discriminative low-level cues while maintaining computational efficiency through compound scaling.

Second, dense nested skip pathways propagate intermediate feature maps ($X^{i,j}$) across decoder stages. These pathways progressively reduce semantic disparity between encoder and decoder representations by enabling multi-resolution aggregation and refinement before fusion. This staged alignment improves structural continuity and stabilizes feature integration, which is different from regular direct concatenation. Also, dual-attention modules (DA_k) have been added to skip fusion nodes. To find areas and decide which traits are the most important, each module uses both spatial attention and channel attention. This technique is excellent for minor retinal issues because it prevents pointless activations and enhances responses that are crucial to the issue.

The segmentation maps are reassembled by the decoder after several iterations of upsampling and some convolutional fine-tuning. By combining characteristics that match representations that are given more attention, this is accomplished. Finally, the design combines adaptive attention calibration, encoder strengthening, and semantic alignment into a single system that efficiently classifies lesions into groups.

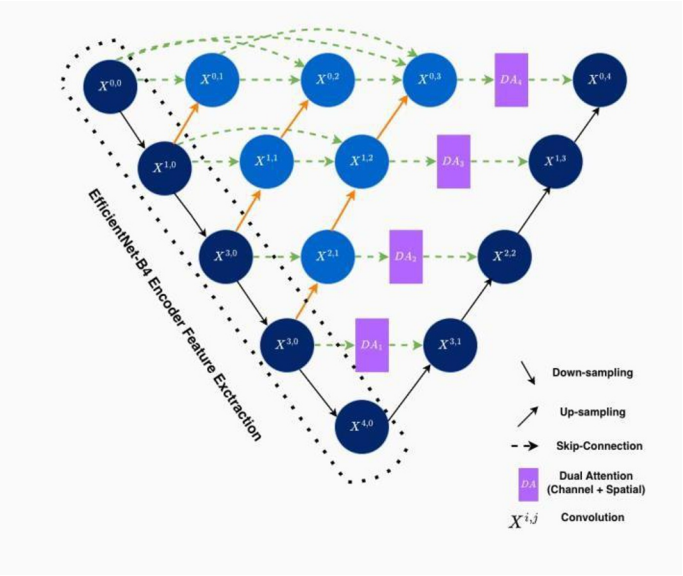


Fig. 1. Proposed EDA-UNet++ architecture showing EfficientNet-B4 encoder, nested skip pathways, and dual-attention refinement modules

4 Results and Discussion

This part of the paper has quantitative, qualitative, and computational tests that show in detail how well our proposed framework works. We used different segmentation models to test our method on the DDR dataset, and then we confirmed it on the IDRiD dataset to get information from both sets of data. The objective is to evaluate segmentation accuracy, generalization capability, and computational efficiency under consistent experimental conditions.

4.1 Macro Performance on DDR (Dice and IoU)

Macro level how well it segments on the DDR dataset was evaluated with class-averaged Dice and IoU scores for all lesion types. The comparison validation and test results for the assessed architectures are compiled in Fig. 2. Stable relative ranking among models and the validity of macro-level evaluation for overall segmentation behavior were confirmed by consistent trends found in both metrics.

Beyond benchmarking performance, this comparison also illustrates the effect of individual architectural components. U-Net serves as the basic encoder–decoder baseline, U-Net++ reflects the effect of nested skip pathways for improved semantic alignment, and Attention U-Net highlights the contribution of attention-based refinement. The proposed EDA-UNet++ combines these strengths

with an EfficientNet-B4 encoder and spatial–channel dual attention. The progressive improvement across these architectures indicates that encoder strengthening and dual-attention refinement jointly contribute to the observed gains in lesion-wise segmentation performance. The proposed **EDA-UNet++** achieved the highest macro performance on the validation split, obtaining approximately 0.44 Dice and 0.29 IoU, outperforming baseline architectures featuring UNet family. This improvement suggests that the integration of an EfficientNet based encoder together with dual attention refinement enhances feature discrimination while preserving lesion relevant spatial structures during decoding. Although a modest performance drop was observed on the test split for all models reflecting domain variability and distribution shift. EDA-UNet++ maintained competitive generalization and remained among the top performing architectures.

Across competing architectures, classical U-Net variants exhibited stable macro scores, while attention enhanced designs yielded incremental gains. The consistent Dice–IoU trends and relatively narrow validation–test performance gaps indicate balanced optimization without evidence of overfitting. Overall, the observed improvements highlight the benefit of encoder strengthening and attention guided refinement in enhancing global segmentation consistency, motivating the subsequent lesion wise performance analysis.

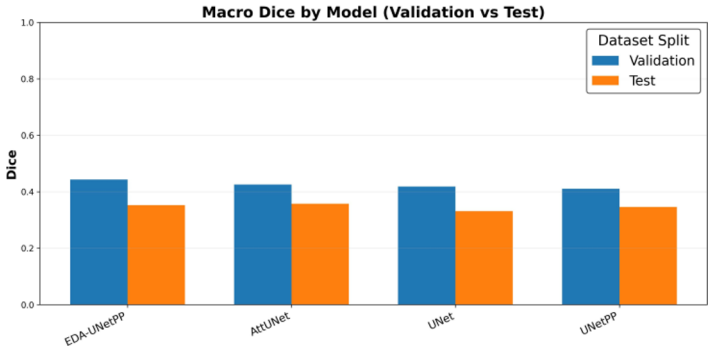
4.2 Per Lesion Performance on DDR

Although macro averaged metrics provide an overall indication of segmentation quality, diabetic retinopathy lesions exhibit substantial variability in scale, texture, and contrast. Consequently, lesion wise evaluation is critical for understanding architectural behavior. Fig. 3 visualizes per lesion Dice comparisons across Microaneurysms (MA), Exudates (EX), Soft Exudates (SE), and Hemorrhages (HE), while Tables 1 and 2 report the complete quantitative metrics (IoU, Dice, Precision, Recall, and F1) for the validation and test splits.

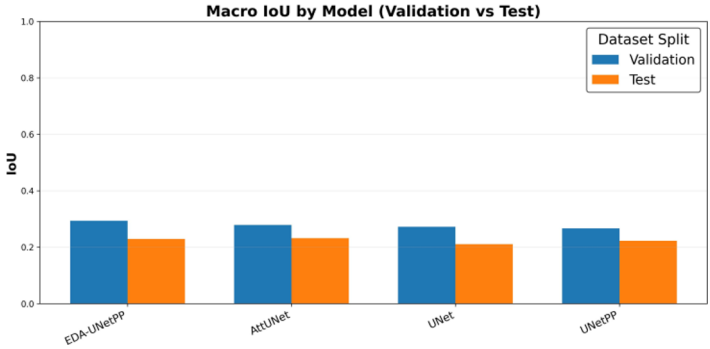
In terms of numbers, EDA-UNet++ outperforms UNet (0.41), UNet++ (0.41), and Attention UNet (0.42) with the best mean Dice on DDR validation (0.44). With an overall mean Dice of 0.35 and the highest EX Dice (0.57) on the DDR test split, EDA-UNet++ maintains its competitiveness with the strongest baselines and validates steady generalization trends.

Across all lesion categories, consistent ranking trends are observed among the four evaluated architectures. The proposed EDA-UNet++ demonstrates improved representation of medium scale lesions and maintains competitive generalization performance, indicating that encoder strengthening and dual attention refinement effectively enhance discriminative feature fusion.

Microaneurysms (MA): MA segmentation is still the hardest because it has low contrast and takes up very little space. The small range of dice values for all models shows that they are sensitive to resolution loss and encoder downsampling. The proposed EDA-UNet++ performs just as well as baseline architectures on both splits, which shows that the stronger encoder keeps fine-grained lesion cues without lowering detection sensitivity.



(a) Macro Dice comparison



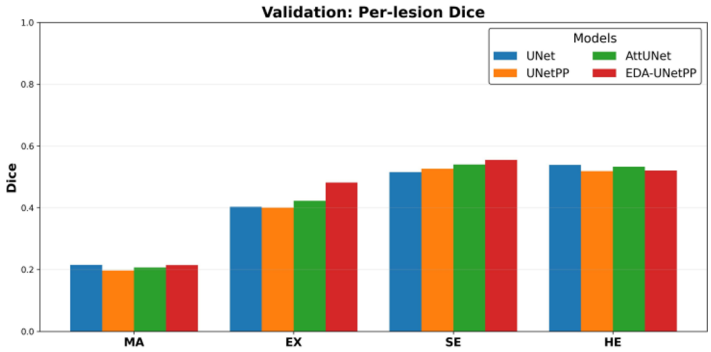
(b) Macro IoU comparison

Fig. 2. Macro-level segmentation performance on DDR showing averaged Dice and IoU across lesion classes

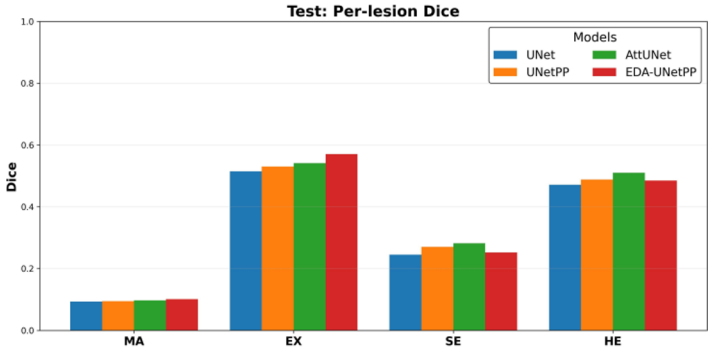
Exudates (EX): EX segmentation shows the best things about EDA-UNet++. It gets the top Dice and IoU scores on on both the validation and test sets because it can find boundaries and choose features well. Channel-spatial dual attention clearly boosts important signals while lowering distractions. Balanced Precision Recall cuts down on false positives and keeps good lesion coverage.

Soft Exudates (SE): Nested skip refinement enables effective multi-scale feature fusion, which enhances EDA-UNet++'s robust validation performance in soft exudate segmentation. The gaps between architectures in the test set get smaller, which shows that they are attuned to shifts in how data is spread. This shows how hard it is to find the borders of lesions that are spread out and have unclear textures.

Hemorrhages (HE): When it comes to dividing up a hemorrhage, all models do pretty much the same thing. These growths often look like big, dark spots with sharp edges. In this case, the models don't matter; what matters is the shape of the injury. They all get about the same score on the Dice test.



(a) Validation per-lesion Dice



(b) Test per-lesion Dice

Fig. 3. Comparing DDR validation and test splits for UNet, UNet++, Attention U-Net, and the suggested EDA-UNet++ per lesion dice

Overall Interpretation: At the macro level, we see the same trends as at the lesion level: architectural improvements help lesions with more distinct structures, like exudates and soft exudates, the most. Microaneurysms, on the other hand, are still limited by how well they can be seen. Consistent performance across validation and test sets shows that model rankings are stable and the evaluation process is strong. These results lead us to the next step: a qualitative look at how to improve boundaries and deal with failure modes in tough cases.

4.3 Cross-Dataset Evaluation on IDRiD

EDA-UNet++ generalization across DDR and IDRiD datasets was evaluated across datasets. Macro Dice was 0.44 on DDR validation, with microaneurysms lagging behind due to strong performance on soft exudates and hemorrhages. With a robust exudate dice of 0.57, the test dice dropped to 0.35.

With a macro Dice of 0.38 from IDRiD validation, which excelled in hemorrhages and exudates, the test Dice was 0.37, while the exudate Dice increased to 0.60 and the soft exudates to 0.41.

Table 1. DDR validation results for the four evaluated models

Model	Lesion	IoU	Dice	Prec.	Rec.	F1
U-Net	MA	0.12	0.21	0.15	0.33	0.21
	EX	0.25	0.40	0.26	0.81	0.40
	SE	0.34	0.51	0.46	0.58	0.51
	HE	0.36	0.53	0.45	0.66	0.53
	Mean	0.27	0.41	0.33	0.60	0.41
U-Net++	MA	0.10	0.19	0.12	0.46	0.19
	EX	0.25	0.40	0.26	0.84	0.40
	SE	0.35	0.52	0.45	0.62	0.52
	HE	0.35	0.51	0.44	0.61	0.51
	Mean	0.26	0.41	0.32	0.63	0.41
Attention U-Net	MA	0.11	0.20	0.14	0.39	0.20
	EX	0.26	0.42	0.28	0.80	0.42
	SE	0.36	0.53	0.52	0.56	0.53
	HE	0.36	0.53	0.41	0.72	0.53
	Mean	0.27	0.42	0.34	0.62	0.42
EDA-UNet++	MA	0.12	0.21	0.14	0.41	0.21
	EX	0.31	0.48	0.35	0.76	0.48
	SE	0.38	0.55	0.46	0.68	0.55
	HE	0.35	0.52	0.44	0.62	0.52
	Mean	0.29	0.44	0.35	0.62	0.44

As summarized in Table 3**,** the lesion-wise Dice scores across DDR and IDRiD demonstrate consistent performance trends, confirming stable cross-dataset generalization.

Microaneurysms proved challenging everywhere due to their extremely small spatial footprint. Encoder strengthening, nested semantic alignment, and dual attention refinement all work together to improve cross domain robustness appropriate for clinical screening settings, as evidenced by the comparatively small macro Dice shifts across datasets that verify the lack of DDR-specific overfitting.

4.4 Qualitative Analysis Across Datasets

Qualitative inspection complements quantitative results by assessing boundary fidelity and localization across lesion categories. In Fig. 4, representative DDR

Table 2. DDR test results for the four evaluated models

Model	Lesion	IoU	Dice	Prec.	Rec.	F1
U-Net	MA	0.04	0.09	0.05	0.34	0.09
	EX	0.34	0.51	0.40	0.71	0.51
	SE	0.13	0.24	0.17	0.41	0.24
	HE	0.30	0.47	0.48	0.45	0.47
	Mean	0.21	0.33	0.27	0.48	0.33
U-Net++	MA	0.04	0.09	0.05	0.47	0.09
	EX	0.36	0.53	0.40	0.75	0.53
	SE	0.15	0.27	0.18	0.52	0.27
	HE	0.32	0.48	0.52	0.45	0.48
	Mean	0.22	0.34	0.29	0.55	0.34
Attention U-Net	MA	0.05	0.09	0.05	0.43	0.09
	EX	0.37	0.54	0.44	0.70	0.54
	SE	0.16	0.28	0.22	0.37	0.28
	HE	0.34	0.51	0.50	0.51	0.51
	Mean	0.23	0.35	0.30	0.50	0.35
EDA-UNet++	MA	0.05	0.10	0.05	0.43	0.10
	EX	0.39	0.57	0.48	0.69	0.57
	SE	0.14	0.25	0.17	0.45	0.25
	HE	0.32	0.48	**0.56**	0.42	0.48
	Mean	0.22	0.35	**0.31**	0.50	0.35

and IDRiD predictions are displayed. Despite detections that match annotations, microaneurysms are still challenging because of their small size and low contrast. Strong structural agreement is seen in exudates, with salient region focus being improved by attention mechanisms. Through nested skip refinement, soft exudates exhibit seamless boundary recovery while preserving shape uniformity across domains. Hemorrhages show high overlap and stable edge continuity. Overall, EDA-UNet++ achieves consistent localization and boundary coherence across datasets, confirming robust cross-domain generalization and validating the combined benefits of encoder strengthening, semantic alignment, and dual attention refinement.

Table 3. Cross-dataset lesion-wise Dice performance of EDA-UNet++

Dataset	Split	MA	EX	SE	HE	Macro
DDR	Val.	0.21	0.48	0.55	0.52	0.44
DDR	Test	0.10	0.57	0.25	0.48	0.35
IDRiD	Val.	0.16	0.58	0.24	0.53	0.38
IDRiD	Test	0.12	0.60	0.41	0.35	0.37

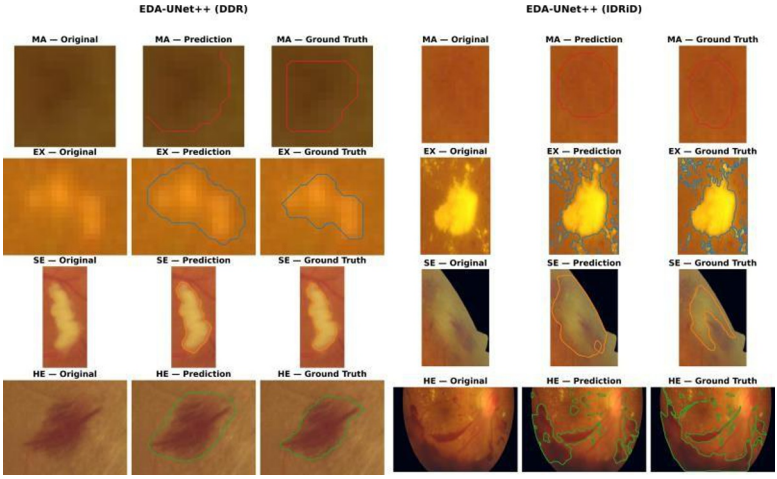


Fig. 4. Qualitative segmentation comparison of EDA-UNet++ across DDR and IDRiD datasets

4.5 Computational Efficiency Analysis

Computational efficiency was assessed using parameter count, total inference time, per-image latency, and environmental impact metrics (Table 4). Baseline models (U-Net, U-Net++, and Attention U-Net) maintain comparable sizes between 20.22M and 20.81M parameters, whereas EDA-UNet++ increases capacity to 27.24M due to EfficientNet-based encoding and dual-attention refinement.

Despite the higher parameter count, inference runtime remains nearly identical across models. Validation processing time ranges between 16.44–16.96 s (approximately 0.11 s per image), while test inference time lies between 19.17–19.51 s (approximately 0.085 s per image), indicating negligible latency overhead.

Energy consumption across models remains low and comparable. Validation energy usage varies between 0.00114–0.00118 kWh, with corresponding CO_2 emissions around 0.00054–0.00056 kg. Test energy consumption ranges between 0.00133–0.00135 kWh, with emissions near 0.00063–0.00064 kg. Importantly, EDA-UNet++ achieves the highest validation mDice (0.44) and competitive test mDice (0.35) without increasing energy or carbon footprint relative to baseline models.

These results demonstrate that the proposed architecture preserves an effective efficiency–accuracy–sustainability balance, supporting its suitability for large-scale retinal screening deployment.

Table 4. Computational efficiency and environmental impact comparison

Model	Split	Params (M)	Time (s)	Images	Time/img (s)	kWh	CO_2 (kg)	mDice
UNet	Validation	20.22	16.96	149	0.1138	0.00117	0.00055	0.41
UNet	Test	20.22	19.51	225	0.0867	0.00135	0.00064	0.33
UNet++	Validation	20.81	16.44	149	0.1104	0.00114	0.00054	0.41
UNet++	Test	20.81	19.29	225	0.0858	0.00134	0.00063	0.34
AttUNet	Validation	20.30	16.94	149	0.1138	0.00117	0.00055	0.42
AttUNet	Test	20.30	19.17	225	0.0852	0.00133	0.00063	0.35
EDA-UNet++	Validation	27.24	16.92	149	0.1136	0.00117	0.00055	0.44
EDA-UNet++	Test	27.24	19.42	225	0.0863	0.00134	0.00064	0.35

5 Conclusion and Future Work

This work presented an analysis comparing deep learning architectures for lesion-wise diabetic retinopathy segmentation and introduced the proposed EDA-UNet++ model to address structural limitations of U-Net derivatives. To improve adaptive feature prioritization, decrease encoder–decoder semantic mismatch, and preserve fine lesion details, the architecture combined an EfficientNet-B4 encoder, dual spatial–channel attention, and nested semantic alignment via dense skip pathways.

Experiments on the DDR dataset showed improved macro and lesion-wise performance, especially for exudates and soft exudates, due to better boundary and contextual modeling. Cross-dataset evaluation on IDRiD confirmed stable generalization with minimal parameter and runtime overhead.

Sensitivity to dataset variability and microaneurysms are two issues that still exist for very small lesions. Future work will explore transformer-based hybrid encoders, adaptive multi-resolution strategies, and broader clinical validation to enhance robustness and deployment readiness.

References

1. M. V. Polyakova, “Image segmentation with a convolutional neural network without pooling layers in dermatological disease diagnostics systems,” Radioelectron. Comput. Sci. Control, no. 1, pp. 51–58, 2023. doi: 10.15588/1607-3274-2023-1-5.
2. M. Lv et al., “Small scale multi-object segmentation in mid-infrared image using the image timing features–Gaussian mixture model and convolutional-UNet,” Sensors, vol. 25, no. 11, p. 3440, 2025. doi: 10.3390/s25113440.
3. N. Mu et al., “S-Net: a multiple cross aggregation convolutional architecture for automatic segmentation of small/thin structures for cardiovascular applications,” Frontiers in Physiology, vol. 14, 2023. doi: 10.3389/fphys.2023.1209659.
4. R. Yousef et al., “U-Net-based models towards optimal MR brain image segmentation,” Diagnostics, vol. 13, no. 9, p. 1624, 2023. doi: 10.3390/diagnostics13091624.
5. H. Fu et al., “HMRNet: High and multi-resolution network with bidirectional feature calibration for brain structure segmentation in radiotherapy,” IEEE J. Biomed. Health Informat., vol. 26, no. 9, pp. 4519–4529, 2022.

6. H. Zunair and A. Ben Hamza, “Sharp U-Net: Depthwise convolutional network for biomedical image segmentation,” Comput. Biol. Med., vol. 136, p. 104699, 2021. doi: 10.1016/j.compbiomed.2021.104699.
7. J. Wang et al., “SAR-U-Net: Squeeze-and-excitation block and atrous spatial pyramid pooling based residual U-Net for automatic liver segmentation in computed tomography,” Comput. Methods Programs Biomed., vol. 208, p. 106268, 2021. doi: 10.1016/j.cmpb.2021.106268.
8. O. Abueed, Y. Wang, and M. Khasawneh, “A systematic review of U-Net optimizations: Advancing tumour segmentation in medical imaging,” IET Image Process., vol. 19, no. 1, 2025. doi: 10.1049/ipr2.70203.
9. E. J. Alwadee, X. Sun, Y. Qin, and F. C. Langbein, “LATUP-Net: A lightweight 3D attention U-Net with parallel convolutions for brain tumor segmentation,” Comput. Biol. Med., vol. 184, p. 109353, 2025. doi: 10.1016/j.compbiomed.2024.109353.
10. K. Avazov et al., “Dynamic focus on tumor boundaries: A lightweight U-Net for MRI brain tumor segmentation,” Bioengineering, vol. 11, no. 12, p. 1302, 2024. doi: 10.3390/bioengineering11121302.
11. A. Amer, T. Lambrou, and X. Ye, “MDA-Unet: A multi-scale dilated attention U-Net for medical image segmentation,” Appl. Sci., vol. 12, no. 7, p. 3676, 2022. doi: 10.3390/app12073676.
12. G. C. Ates, P. Mohan, and E. Celik, “Dual cross-attention for medical image segmentation,” Eng. Appl. Artif. Intell., vol. 125, p. 107139, 2023. doi: 10.1016/j.engappai.2023.107139.
13. H. Huang et al., “Channel prior convolutional attention for medical image segmentation,” Comput. Biol. Med., vol. 178, p. 108784, 2024. doi: 10.1016/j.compbiomed.2024.108784.
14. C. Su, X. Luo, S. Li, L. Chen, and J. Wang, “VMKLA-UNet: Vision Mamba with KAN linear attention U-Net,” Sci. Rep., vol. 15, p. 13258, 2025. doi: 10.1038/s41598-025-97397-2.
15. G. Sun et al., “DA-TransUNet: Integrating spatial and channel dual attention with transformer U-Net for medical image segmentation,” Front. Bioeng. Biotechnol., vol. 12, p. 1398237, 2024. doi: 10.3389/fbioe.2024.1398237.
16. S. İnce, I. Kunduracioglu, B. Bayram, and I. Pacal, “U-Net-based models for precise brain stroke segmentation,” Chaos Theory Appl., vol. 7, no. 1, pp. 50–60, 2025. doi: 10.51537/chaos.1605529.
17. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A nested U-Net architecture for medical image segmentation,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2018, pp. 3–7. doi: 10.1109/ICIP.2018.8451724.
18. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: Redesigning skip connections to exploit multiscale features in image segmentation,” IEEE Trans. Med. Imaging, vol. 39, no. 6, pp. 1856–1867, 2020. doi: 10.1109/TMI.2019.2959609.
19. Z. Li, H. Zhang, Z. Li, and Z. Ren, “Residual-Attention UNet++: A nested residual-attention U-Net for medical image segmentation,” Appl. Sci., vol. 12, no. 14, p. 7149, 2022. doi: 10.3390/app12147149.
20. H. Lu, Y. She, J. Tie, and S. Xu, “Half-UNet: A simplified U-Net architecture for medical image segmentation,” Front. Neuroinform., vol. 16, p. 911679, 2022. doi: 10.3389/fninf.2022.911679.
21. M. Z. Tahir et al., “Efficient microaneurysm segmentation in retinal images via a lightweight Attention U-Net for early DR diagnosis,” SLAS Technol., vol. 34, p. 100323, 2025. doi: 10.1016/j.slast.2025.100323.

22. C. Wan et al., "EAD-Net: A novel lesion segmentation method in diabetic retinopathy using neural networks," Dis. Markers, vol. 2021, p. 6482665, 2021. doi: 10.1155/2021/6482665.
23. S.-Y. Lin and C.-L. Lin, "Brain tumor segmentation using U-Net in conjunction with EfficientNet," PeerJ Comput. Sci., vol. 10, p. e1754, 2024. doi: 10.7717/peerj-cs.1754.
24. P. K. Tiwary, P. Johri, A. Katiyar, and M. K. Chhipa, "Deep learning-based MRI brain tumor segmentation with EfficientNet-enhanced UNet," IEEE Access, vol. 13, pp. 54920–54937, 2025. doi: 10.1109/ACCESS.2025.3554405.
25. P. R., J. P. P. M., and N. J. S., "Brain tumor segmentation using multi-scale attention U-Net with EfficientNetB4 encoder for enhanced MRI analysis," Sci. Rep., vol. 15, p. 9914, 2025. doi: 10.1038/s41598-025-94267-9.