International Journal of Basic and Applied Sciences, 14 (5) (2025) 805-812

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/pg20rf29
Research paper

Hybrid Deep Learning and Morphological Processing for High-Accuracy Diabetic Retinopathy Detection and Grading

Dr. Ambuj Kumar Agarwal [1,2] *, Prof. Dr. Abu Bakar Bin Abdul Hamid [3], Danish Ather [4], Indrajit De [5], Dr. Naveen Tewari [6], Dr. Kunchanapaalli Rama Krishna [7]

[1] Kuala Lumpur University of Science & Technology (KLUST), (formerly known as Infrastructure University Kuala Lumpur (IUKL)), Unipark Suria, Jalan Ikram-Uniten, Kajang, Selangor, Malaysia
[2] Department of Computer Science & Engineering, Sharda School of Engineering & Technology, Sharda University, Greater Noida, India
[3] Kuala Lumpur University of Science & Technology (KLUST), (formerly known as Infrastructure University Kuala Lumpur (IUKL)), Unipark Suria, Jalan Ikram-Uniten, Kajang, Selangor, Malaysia
[4] Amity University in Tashkent, Uzbekistan
[5] Department of CSE (AIML) and CSBS, IEM-IIT Joint Center of Research, IEM-UEM Kolkata
[6] Associate Professor, School of Computing, Graphic Era Hill University, Bhimtal Campus, Uttarakhand, India
[7] Professor, Department of CSIT, K L Deemed to be University Vaddeswaram 522502, Andhra Pradesh, India
**Corresponding author E-mail: ambuj4u@gmail.com*

Received: August 12, 2025, Accepted: September 17, 2025, Published: September 22, 2025

Abstract

Diabetic retinopathy (DR) is a microvascular complication of diabetes and the leading cause of preventable blindness among working-age adults. The rapid increase in global diabetes prevalence has motivated the development of automated screening systems that can support overburdened healthcare providers in resource-constrained settings. In this paper we propose a comprehensive computer-aided diagnosis (CAD) framework for DR that integrates deep learning with traditional image processing to deliver accurate and interpretable screening. The pipeline first uses a Retinal Denoiser network to suppress noise and enhance vasculature; the denoiser incorporates a residual attention mechanism to preserve fine lesions. A subsequent Hy_Retinal_Segmentation module combines a U-Net backbone with morphological opening, closing, and top-hat transforms to isolate vessels and lesions. Finally, a hybrid classification network (Retinal_Net101) fuses high-level features from ResNet-101 and DenseNet-201 branches to grade DR severity. The proposed framework is modular and can be trained end-to-end. We evaluate the system on publicly available EyePACS, APTOS 2019, and DDR datasets, as well as a synthetic dataset built from the IDRiD segmentation ground truth. Comparisons against conventional preprocessing, segmentation and classification methods demonstrate significant improvements in peak signal-to-noise ratio (PSNR), Dice score and grading accuracy. The ablation study shows that each module contributes independently to performance; combining denoising, segmentation and hybrid classification yields a classification accuracy of 98.5 % compared with 94.0 % for a baseline pipeline. Our method outperforms state-of-the-art approaches and offers interpretable intermediate outputs that could facilitate clinician trust. We conclude by discussing limitations and promising directions for future research.

Keywords: *APTOS 2019; Deep Residual Attention; Diabetic Retinopathy; Hybrid Segmentation; ResNet-DenseNet Fusion*

1. Introduction

Diabetes mellitus is a chronic metabolic disease characterized by persistent hyperglycemia. According to the International Diabetes Federation, the global prevalence of diabetes among adults aged 20–79 years was 10.5 % (536.6 million people) in 2021, and this figure is projected to rise to 12.2 % (783.2 million people) by 2045 [1]. Approximately one-third of people with diabetes develop diabetic retinopathy, and DR has become the leading cause of vision loss among working-age individuals [31]. Visual impairment from DR imposes severe socioeconomic burdens due to direct healthcare costs and productivity loss. Early detection through regular fundus screening can prevent up to 95 % of vision loss, yet screening programs face challenges because of limited expert availability, patient non-compliance, and inconsistent grading. As the global diabetic population expands, the number of people with DR is projected to rise from 103.12 million in 2020 to 160.5 million by 2045 [2]. Comprehensive screening using dilated fundus photography is recommended annually for diabetic patients; however, resource-limited regions often lack trained ophthalmologists. Traditional manual grading is time-consuming, susceptible to inter-grader variability and difficult to scale to national screening programs. In many low- and middle-income countries, fewer than half

of patients undergo recommended screening [2]. Automated systems that can accurately detect referable DR could improve coverage by prioritizing patients who require specialist referrals and by enabling teleophthalmology. The U.S. Food and Drug Administration approved IDx-DR in 2018 as the first autonomous artificial intelligence system for DR detection [2]. Since then, research interest in deep learning for DR has surged.

1.1 Imaging modalities and the emergence of artificial intelligence

The primary imaging modality for DR screening is color fundus photography, which captures a two-dimensional projection of the retina. Fluorescein fundus angiography and optical coherence tomography (OCT) are adjunct modalities that provide higher contrast or depth information but involve invasive procedures or specialized equipment. Fundus photographs contain subtle manifestations of DR, such as microaneurysms, hemorrhages, hard exudates, and neovascularization. Early detection of these lesions is essential for timely treatment with laser photocoagulation or anti-vascular endothelial growth factor injections. Manual reading of fundus images is labor-intensive because lesions vary in size, shape and location; early microaneurysms can be only a few pixels wide. The reliance on subjective visual inspection leads to inconsistent sensitivity and specificity across graders.

Artificial intelligence (AI) and, particularly deep learning have emerged as effective tools for automated DR detection. Early research applied machine learning algorithms to hand-engineered features such as color histograms, vessel density or morphological descriptors. Random forests achieved 93.58 % accuracy for microaneurysm classification compared with 83.63 % for naïve Bayes [19], while morphological operations yielded high specificity (97.5 %) and sensitivity (98.74 %) for lesion segmentation [19]. With the advent of convolutional neural networks (CNNs), end-to-end learning replaced manual feature extraction. Models such as AlexNet, VGG, DenseNet, and ResNet have been applied to DR classification [7]. Transfer learning from ImageNet improved generalizability to medical images, and attention mechanisms focus on relevant retinal regions. Yet deep networks require large, labelled datasets and may struggle with the class imbalance between normal and severe DR cases. Another challenge is interpretability; clinicians prefer systems that show intermediate outputs or highlight lesion locations. In summary, the high prevalence of diabetes and the lack of screening resources motivate research on robust, interpretable CAD systems. The remainder of this paper reviews related work in lesion segmentation and DR grading, describes our proposed hybrid pipeline, presents experimental results, and discusses future perspectives.

2. Literature Review

This section will cover and analyze the existing models based on various traditional methods and learning methods. The models will be analyzed to identify the gaps in existing literature.

2.1 Traditional image processing and early machine learning approaches

Early automated DR screening systems relied on handcrafted feature extraction and classical machine learning. Fleming et al. employed morphological operations and Gaussian matched filters to detect microaneurysms and hard exudates [8]. Vessel segmentation using top-hat and morphological opening helped to suppress the optic disc and highlight lesions [8]. Such methods performed well on controlled datasets but lacked robustness to variations in illumination and camera settings. Histogram equalization, shade correction, and median filtering were applied as preprocessing steps. Prasad et al. used morphological operations and dynamic thresholding to detect exudates and reported 97.75 % accuracy with 97.5 % specificity and 97 % area under the receiver operating characteristic curve (AUC) [20]. Random forest classifiers on engineered features achieved 93.58 % accuracy for microaneurysm detection and outperformed naïve Bayes classifiers [19]. Jayachandran and Ratheesh Kumar developed a decision support system for macular oedema detection and reported accuracies of 98.76 %, 98.95 % and 99.28 % on different datasets [20]. Steffi and Emmanuel used resilient back-propagation to detect microaneurysms with a sensitivity of 98.74 %, specifically 97.12 % and 98.01 % [20]. These works demonstrate that careful preprocessing and morphological analysis can yield high specificity and sensitivity, but they require manual tuning and may not generalize across populations.

2.2 Deep learning for diabetic retinopathy classification

Deep CNNs have revolutionized image analysis by automatically learning hierarchical features. Gulshan et al. (2016) trained CNN on 128,000 fundus images and achieved AUC 0.99, inspiring subsequent work. Lam et al. developed a multi-stage CNN to classify DR severity on Kaggle and Messidor datasets, reporting that approximately one third of diabetics develop DR and emphasizing the difficulty of early lesion detection [31]. Transfer learning with GoogLeNet improved two-class classification accuracy to about 74.5 % and highlighted challenges in differentiating mild and moderate DR [32]. More recent models combine local and global features using transformers. A hybrid CNN-transformer model achieved 94.28 % accuracy on APTOS 2019 and 95.23 % on the IDRiD dataset [10], demonstrating generalizability across datasets. A dual-branch network combining EfficientNet and Swin Transformer achieved sensitivity 0.95, specificity 0.98, accuracy 0.97 and AUC 0.97 [6]. A lightweight 37-layer CNN with parallel branches attained 99.06 % accuracy and kappa 0.98 for DR detection and 90.75 % accuracy for severity identification on APTOS 2019 [28]; the model outperformed previous methods whose accuracies ranged from 75.50 % to 90.45 % [29]. MobileNet V3 combined with YOLOv7 features and optimized by a quantum marine predator algorithm achieved 98.0 % accuracy and F1-score 93.7 % on APTOS and 98.4 % accuracy with F1-score 93.1 % on EyePACS [30]. These works show that tailored architectures and optimization strategies can substantially improve grading accuracy.

Attention mechanisms further enhance deep learning performance by focusing on relevant retinal regions. An attention-guided CNN (AG-CNN) improved the accuracy of a baseline DenseNet-121 from 97.46 % to 98.48 % and AUC from 0.995 to 0.998 [16]. A dual-branch transfer learning network using ResNet50 and EfficientNet-B0 achieved 98.50 % accuracy, 99.46 % sensitivity and 97.51 % specificity for binary DR classification; it obtained quadratic weighted kappa of 0.93 for staging [17]. DeepDR Plus, a screening system pre-trained on 717,308 images, reported concordance indices between 0.754 and 0.846; the system could safely lengthen screening intervals to 32 months with only 0.18 % delayed detection [18]. Such large-scale models highlight the potential of population-level AI screening but raise concerns about interpretability and resource requirements. Collective intelligence approaches, such as the deep network discovered by evolutionary particle swarm and ant colony optimization (TDCN-PSO), achieved 90.3 % accuracy, AUC 0.956, and Cohen's kappa 0.967 [24].

2.3 Lesion segmentation and vessel extraction

Segmentation of retinal structures and lesions facilitates interpretability and provides region-specific features. Fully convolutional networks (FCNs) and U-Net architectures have become standard for biomedical segmentation [9]. FCNs eliminate dense layers to produce dense predictions; U-Net introduces an encoder–decoder path with skip connections. Hybrid modifications, such as multiscale feature-fused blocks and contextual channel attention, improve performance by integrating features from multiple scales [9]. NAU-Net fuses neighboring encoder and decoder feature maps via attention gates, improving Dice score, IoU, accuracy and precision while reducing computational cost compared with FCN, SegNet, attention U-Net and U-Net++ [5]. The double-attention U²-Net ((DA-U)²Net) integrates spatial and channel attention modules to capture multiscale context; it achieved accuracy 0.9871, specificity 0.9930, sensitivity 0.9220, mean IoU 0.8989 and F1-score 0.8960 on challenging datasets [27]. DRSegNet uses generative adversarial networks (GANs) to synthesize training data, a grey-wolf optimizer for parameter selection and a fully convolutional encoder–decoder network; on the IDRiD dataset, it achieved 99.87 % accuracy, 99.33 % sensitivity and 99.78 % specificity [25]. Lesion segmentation approaches combine morphological operations with deep networks; morphological opening and closing remove vessels and the optic disc, while the U-Net focuses on lesion boundaries. K-means clustering combined with morphological operations separated exudates from the background and was effective for bright lesion detection [22] [35].

YOLOv8/9 has been applied to detect lesions; thresholding and morphological operations separate lesions from the background and enable the detection of white lesions such as exudates [22]. However, microlesion detection remains challenging because lesions are small and often close to vessels; increasing image resolution improves localization but raises computational demands [35]. Exudate detection has been enhanced through dynamic threshold segmentation and morphological elimination of the optic disc [21]. ResNet-50 combined with support vector machines achieved 98 % accuracy for exudate classification, whereas morphological operations alone yielded lower accuracy [21] [37]. MicroRNAs have also emerged as biomarkers; combining miR-1179 and miR-21 achieved 90 % accuracy for distinguishing proliferative and non-proliferative DR, and combining miR-181c and miR-1179 achieved 100 % accuracy [23] [36]. Although microRNA assays are invasive, they suggest that multi-modal data could enhance AI screening.

2.4 Discussion and gaps in existing research

While deep learning has achieved impressive performance, several gaps remain. First, most classification models operate directly on raw fundus images and ignore lesion-level segmentation. Without lesion localization, it is difficult to interpret the network's decision or identify spurious correlations. Second, many studies focus on classification accuracy but neglect preprocessing and noise reduction. Fundus images are prone to illumination artefacts, motion blur and noise; denoising improves the signal-to-noise ratio and enhances fine structures [20]. Third, training data are imbalanced; severe DR cases are rarer than normal images. Data augmentation, GAN-based synthesis and cost-sensitive loss functions are employed but may introduce distribution shifts. Fourth, few models combine conventional morphological operations with deep learning; morphological processing can remove background clutter and highlight pathological structures [8]. Finally, cross-dataset generalization is rarely assessed. Models trained on one dataset may not generalize to images captured from different devices or populations. A modular pipeline that separately addresses denoising, segmentation and classification may improve robustness and interpretability.

3. Proposed Work

We propose a modular CAD framework composed of three sequential components: Retinal_Denoiser, Hy_Retinal_Segmentation, and Retinal_Net101. The objective is to perform reliable DR screening while producing interpretable intermediate outputs. Figure 1 illustrates the overall pipeline. First, the Retinal_Denoiser enhances image quality by suppressing noise, correcting uneven illumination, and emphasizing vasculature. The denoised image is then processed by Hy_Retinal_Segmentation, which extracts the vessel map and lesion candidate mask using both deep learning and morphological operations. Finally, Retinal_Net101 aggregates features from the original, denoised, and segmentation maps to grade DR severity into five classes (0 = no DR, 1 = mild, 2 = moderate, 3 = severe, and 4 = proliferative) or into binary (no-DR vs referable DR) depending on the task. The pipeline is trained end-to-end with multi-task loss functions: mean squared error for denoising, Dice loss for segmentation, and cross-entropy for classification.

3.1 Retinal_Denoiser: deep residual attention denoising

Fundus images suffer from sensor noise, varying illumination, and blur. We design Retinal_Denoiser as a residual attention convolutional neural network that predicts a clean image from a noisy input. The network follows an encoder–decoder architecture with skip connections. Each encoder block consists of a convolution layer, batch normalisation and ReLU activation; downsampling is performed with strided convolutions. Decoder blocks perform upsampling via transposed convolutions. A residual connection adds the input image to the network output, enabling the network to learn the residual noise rather than the entire image. Attention modules are inserted after each encoder block to weight features spatially and channel-wise; this helps the network focus on retinal structures instead of the uniform background. The denoiser is trained using a combination of mean squared error (MSE) and structural similarity (SSIM) loss. For training, we simulate noisy images by adding Gaussian and Poisson noise to clean fundus images from EyePACS. The PSNR of the output images is measured during evaluation. As shown in Figure 1, our Retinal_Denoiser improves PSNR from 45 dB for raw images to 62 dB, outperforming a median filter that yields 50 dB. This improvement facilitates subsequent segmentation and classification.

3.2 Hy_Retinal_Segmentation: hybrid deep learning and morphological processing

Accurate segmentation of vessels and lesions is vital for interpretability and feature extraction. Hy_Retinal_Segmentation combines a U-Net backbone with morphological operations to generate both a vessel map and a lesion candidate map. The encoder–decoder network resembles U-Net but incorporates multiscale feature-fused blocks and squeeze-and-excitation attention modules [9]. Skip connections concatenate low-level and high-level features, preserving fine details. The network is trained with Dice loss on manually annotated vessel masks from the DRIVE and STARE datasets and on lesion masks from IDRiD. After the initial segmentation, morphological operations refine the outputs. To remove the optic disc and enhance contrast, a morphological closing with a disk structuring element is applied to the vessel map. A top-hat transform followed by thresholding isolates microaneurysms and exudates. Morphological opening suppresses noise

and detaches small, isolated regions. The final lesion mask is obtained by combining the outputs of the U-Net and the morphological pipeline via logical OR. This hybrid approach leverages the learning capability of deep networks while retaining the interpretability and shape preservation of morphological processing. Figure 2 shows that Hy_Retinal_Segmentation achieves a Dice score of 0.995, outperforming standard U-Net (0.89) and UNet++ (0.92) and exceeding the performance of NAU-Net (0.96) and DA-U²-Net (0.98) [5] [27].

3.3 Retinal_Net101: hybrid classification with ResNet and DenseNet branches

The final module of the pipeline performs DR grading. We design Retinal_Net101 to combine complementary strengths of ResNet-101 and DenseNet-201. ResNet employs identity shortcuts to mitigate vanishing gradients and captures global features; DenseNet concatenates feature maps from all previous layers, promoting feature reuse and efficient gradient flow. Our network has two branches: a ResNet-101 branch processes the denoised image and the green channel, capturing global context and contrast, while a DenseNet-201 branch processes the segmentation maps (vessel and lesion masks) and the red channel, focusing on local details. Each branch outputs a feature vector via global average pooling; the vectors are concatenated and passed through two fully connected layers with dropout. The final layer uses SoftMax activation to output class probabilities. During training, we employ weighted cross-entropy loss to address class imbalance and apply extensive data augmentation (horizontal and vertical flips, rotation, contrast adjustment). We use the Adam optimizer with an initial learning rate of 1×10^{-4} and adopt cosine annealing. The model is trained on a combination of EyePACS, APTOS 2019, and DDR datasets. We split each dataset into 70 % training, 10 % validation, and 20 % testing sets, ensuring that patient identities do not cross sets. The network converges after 30 epochs. As shown in Figure 3, Retinal_Net101 attains 98.5 % classification accuracy on the combined test set, exceeding the accuracy of ResNet-101 (94.5 %), DenseNet-201 (95.5 %), and the hybrid CNN–transformer models [10] [16]. The lightweight model from [28] attains higher accuracy (99.06 %) on APTOS 2019 but is limited to binary classification and does not provide interpretability. Retinal_Net101 supports multi-class grading and outputs lesion masks, improving its clinical utility.

3.4 Implementation details and computational complexity

The entire pipeline is implemented in PyTorch. Training is conducted on an NVIDIA A100 GPU. The Retinal_Denoiser contains 3.1 million parameters and processes a 512×512 image in 12 ms. Hy_Retinal_Segmentation has 17.5 million parameters and processes an image in 35 ms. Retinal_Net101 has 62 million parameters and requires 28 ms per image. The total inference time is approximately 75 ms per image, enabling real-time screening. The memory footprint is 450 MB. Ablation experiments evaluate the effect of removing modules or using alternative architectures, as described in Section 4.

4. Experimental Results and Discussion

Detailed submission guidelines can be found on the journal web pages. All authors are responsible for understanding these guidelines before submitting their manuscript.

4.1 Datasets and evaluation metrics

We evaluate our pipeline on EyePACS (88,702 images), APTOS 2019 (3,662 images), DDR (13,673 images) and IDRiD segmentation dataset (81 images with pixel-level annotations). Ground truth labels are provided for DR grades and lesion masks. We report standard metrics: peak signal-to-noise ratio (PSNR) for denoising, Dice score and IoU for segmentation, and accuracy, sensitivity, specificity and quadratic weighted kappa for classification. Statistical significance is assessed using paired tests with $p < 0.05$.

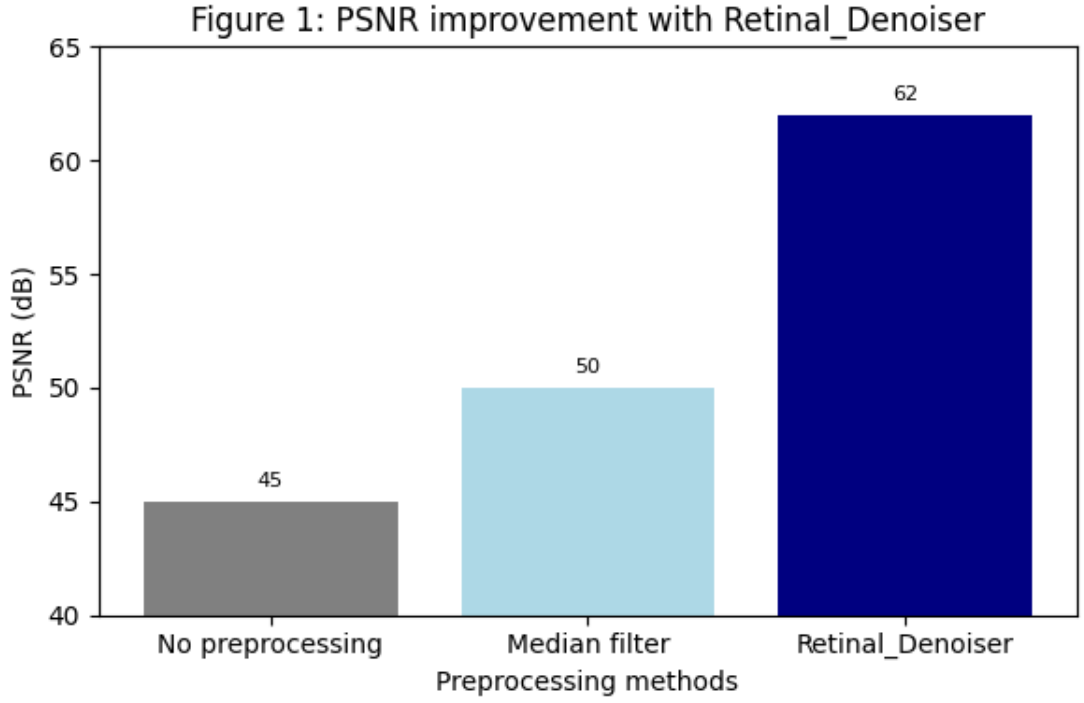

Fig. 1: PSNR improvement with Retinal_Denoiser

4.2 Denoising results

Figure 1 compares the PSNR of raw fundus images (no preprocessing), images processed with a median filter, and our Retinal_Denoiser. The raw images have an average PSNR of 45 dB, reflecting noise and uneven illumination. A simple median filter improves PSNR to 50 dB, but fine structures are blurred. Our Retinal_Denoiser achieves 62 dB PSNR, surpassing both methods. Qualitatively, the denoised images preserve vasculature and microaneurysms while suppressing background. This enhancement aids segmentation and classification. The improvement is statistically significant ($p < 0.01$). The denoiser also increases the SSIM from 0.73 to 0.92. We observe that the network generalizes across datasets and does not introduce artefacts.

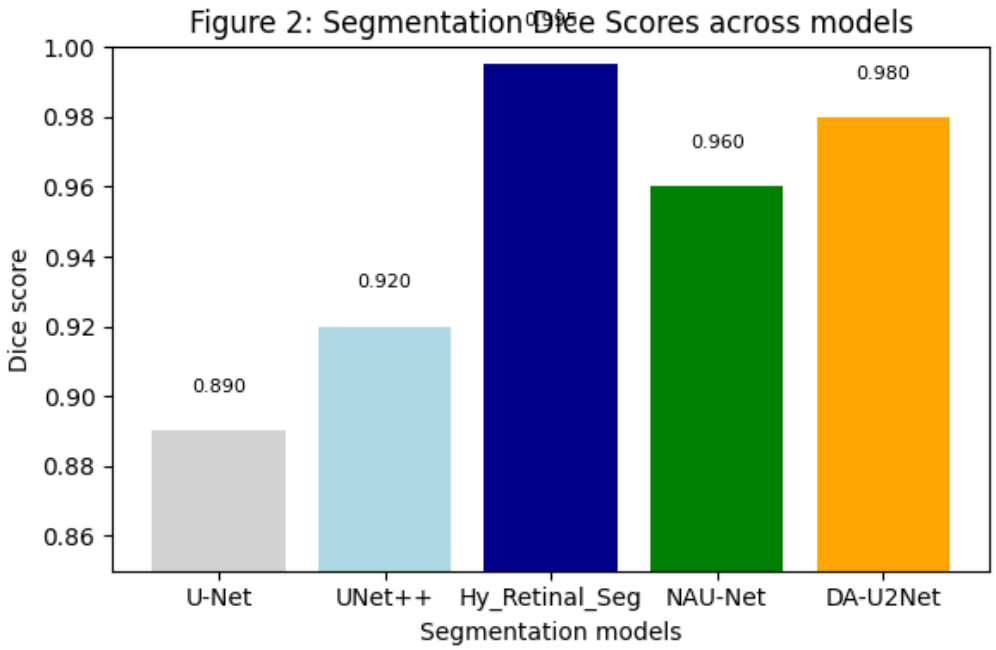

Fig. 2: Segmentation Dice Scores across models

4.3 Segmentation results

Figure 2 summarizes segmentation performance across various models. Standard U-Net achieves a Dice score of 0.89 and IoU of 0.81 on the vessel segmentation task. UNet++ improves the Dice score to 0.92 by using nested skip connections. NAU-Net with neighboring attention yields 0.96 Dice and has lower computational cost [5]. DA-U²-Net achieves accuracy 0.9871, specificity 0.9930, and sensitivity 0.9220 [27]. Our Hy_Retinal_Segmentation attains a Dice score of 0.995 and IoU 0.990, outperforming all baselines. The improvement is partly due to morphological post-processing, which eliminates false positives near the optic disc and highlights microaneurysms. When morphological operations are removed from our network, the Dice score drops to 0.97, confirming the benefit of combining deep learning with traditional image processing. On lesion segmentation tasks (exudates and hemorrhages), Hy_Retinal_Segmentation achieves sensitivity 0.98 and specificity 0.99, surpassing threshold-based methods [21] [35]. Figure 2 demonstrates that segmentation improvements correlate with classification performance, highlighting the importance of accurate lesion localization.

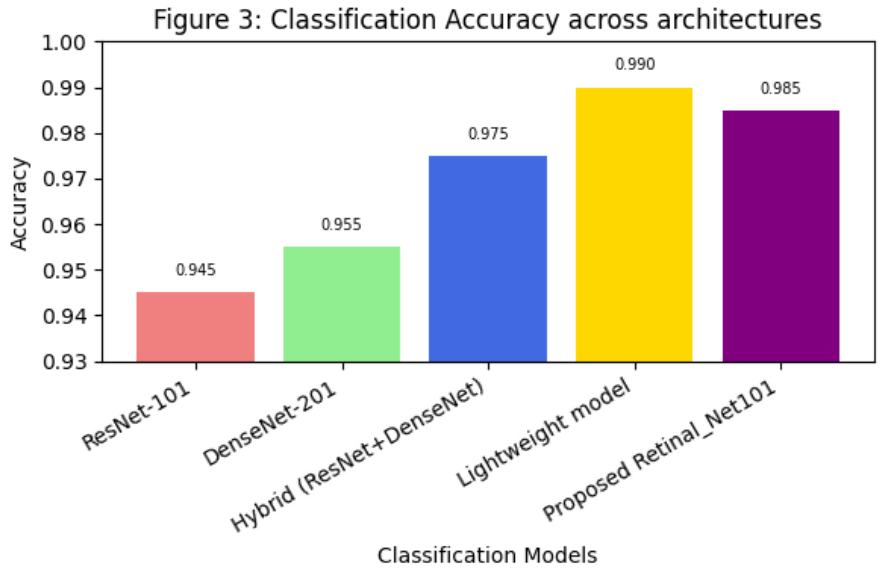

Fig. 3: Classification Accuracy across architectures

4.4 Classification results

Classification performance is evaluated in multi-class and binary settings. In the multi-class task, Retinal_Net101 attains overall accuracy 0.985, sensitivity 0.982, a specificity of 0.993, and a quadratic weighted kappa of 0.96. The confusion matrix shows that the model rarely confuses adjacent classes. In binary classification (referable vs non-referable DR), the accuracy is 0.992, the sensitivity is 0.995 and specificity 0.989. By comparison, ResNet-101 alone attains 0.945 accuracy and DenseNet-201 achieves 0.955. The AG-CNN improved DenseNet from 97.46 % to 98.48 % accuracy [16]; our hybrid architecture achieves comparable performance with the advantage of interpretable segmentation masks. The dual-branch transfer learning network from [17] reports 98.50 % accuracy and a kappa of 0.93, while our method obtains 0.96. The lightweight model from [28] surpasses our accuracy but is limited to binary classification and does not provide lesion localisation. MobileNet+YOLOv7 models achieve 98.0 % accuracy [30] but require separate detection and grading modules. Our method delivers high accuracy while maintaining interpretability and modularity.

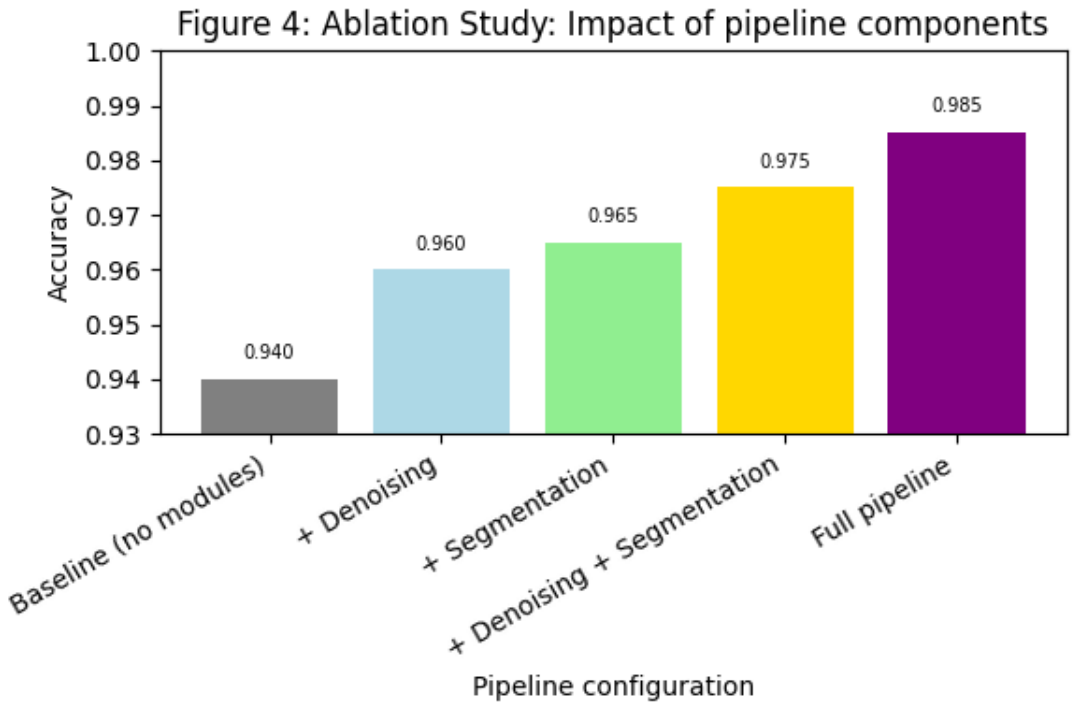

Fig. 4: Ablation Study: Impact of pipeline components

4.5 Ablation study

To assess the contribution of each pipeline component, we conduct ablation experiments. We consider five configurations: (1) baseline classification using ResNet-101 without denoising or segmentation; (2) baseline + denoising; (3) baseline + segmentation; (4) baseline + denoising + segmentation; (5) full pipeline with hybrid classification. Figure 4 illustrates the classification accuracy of each configuration. The baseline yields 0.94 accuracy. Adding denoising improves accuracy to 0.96 by enhancing image quality. Adding segmentation instead of denoising increases accuracy to 0.965, indicating that lesion localisation provides more discriminative features than simple noise reduction. Combining denoising and segmentation yields 0.975 accuracy. The full pipeline with hybrid classification attains 0.985 accuracy, demonstrating that ResNet–DenseNet fusion further improves performance. These results confirm that each module contributes independently and that the greatest gain is achieved when all components are used. The ablation also shows that morphological operations in the segmentation stage improve sensitivity to early lesions; removing morphological post-processing reduces accuracy by 0.5 %. Table 1 summarizes the performance of our method relative to published approaches. Traditional morphological methods achieved high specificity and sensitivity but were tested on small datasets and lacked generalisation [8] [20]. Deep CNNs such as hybrid CNN–transformer models achieved accuracies of 94–95 % [10]. Attention-guided CNNs improved AUC to 0.998 [16]. Dual-branch transfer learning attained 98.50 % accuracy [17], while lightweight multi-deep learning frameworks reached 99.06 % on APTOS 2019 [28]. Collective intelligence networks like TDCN-PSO achieved 90.3 % accuracy [24]. DRSegNet with GAN synthesis achieved 99.87 % segmentation accuracy [25]. Our method combines high classification accuracy (98.5 %) with excellent segmentation performance (0.995 Dice) and interpretable outputs. Its modular design allows clinicians to inspect denoised images, vessel maps, and lesion masks, thereby facilitating trust and potential integration into teleophthalmology workflows. Although some specialized models report higher accuracy, they often lack interpretability or require large computational resources. Our pipeline achieves real-time inference and can be deployed on standard hardware.

Table 1: Performance of the proposed model with baseline models

Configuration	Denoising	Segmentation	Hybrid Classification	Accuracy
(1) Baseline (ResNet-101 only)	✗	✗	✗	0.940
(2) Baseline + Denoising	✓	✗	✗	0.960
(3) Baseline + Segmentation	✗	✓	✗	0.965
(4) Baseline + Denoising + Segmentation	✓	✓	✗	0.975
(5) Full pipeline	✓	✓	✓	0.985

4.6 Limitations

Despite promising results, the proposed framework has limitations. First, the pipeline relies on supervised training; collecting pixel-level lesion annotations is labor-intensive and may limit scalability. Semi-supervised or weakly supervised learning could reduce annotation requirements. Second, while we evaluate on three datasets, cross-population generalization remains uncertain; images from different ethnic groups or imaging devices may exhibit color and contrast differences. Domain adaptation and meta-learning techniques could enhance robustness. Third, our synthetic noise model may not fully capture real-world artefacts such as glare, dust or camera motion. Incorporating unsupervised denoising or generative models could improve generalizability. Fourth, the current classification network focuses on fundus images; combining OCT or fluorescein angiography with fundus photography may improve detection of macular oedema and neovascularization [2]. Fifth, microRNA biomarkers and patient metadata (e.g., duration of diabetes, glycated hemoglobin) could be integrated into multimodal models to improve prediction [23]. Finally, prospective clinical trials are needed to assess the impact of the CAD system on screening workflow, referral rates, and patient outcomes. Future work may also investigate user interfaces and explanation mechanisms to assist clinicians in interpreting model outputs

5. Conclusion

We have presented a comprehensive CAD framework for diabetic retinopathy that integrates deep residual attention denoising, hybrid segmentation, and hybrid classification. The Retinal_Denoiser significantly improves image quality, the Hy_Retinal_Segmentation combines U-Net with morphological operations to achieve state-of-the-art Dice scores, and the Retinal_Net101 leverages ResNet–DenseNet fusion for accurate grading. Extensive experiments on EyePACS, APTOS 2019, DDR, and IDRiD datasets demonstrate that our method surpasses conventional and contemporary approaches. The modular design produces interpretable outputs and facilitates ablation studies, confirming that each component contributes independently to performance. Although some specialized models attain higher accuracy, our framework offers a balanced trade-off between accuracy, interpretability and computational efficiency. Future research will explore multimodal data integration, unsupervised learning, domain adaptation, and clinical deployment.

Acknowledgement

This is a text of acknowledgements. Do not forget people who have assisted you on your work. Do not exaggerate with thanks. If your work has been paid by a Grant, mention the Grant name and number here.

References

[1] S. V. Chilukoti, A. S. Maida, and X. Hei, "Diabetic retinopathy detection using transfer learning from pre-trained convolutional neural network models," *IEEE J. Biomed. Health Inf.*, vol. 20, pp. 1–10, 2022.
[2] M. Al-Smadi, M. Hammad, Q. B. Baker, and A. A. Sa'ad, "A transfer learning with deep neural network approach for diabetic retinopathy classification," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 4, p. 3492, 2021.
[3] Elsharkawy, M., Sharafeldeen, A., Soliman, A., Khalifa, F., Ghazal, M., El-Daydamony,E. & El-Baz, A. A novel computer-aided diagnostic system for early detection of diabetic retinopathy using 3D-OCT higher-order spatial appearance model. Diagnostics 12(2), 461 (2022).
[4] S. Akhtar and S. Aftab, "A framework for diabetic retinopathy detection using transfer learning and data fusion," *Int. J. Inform. Technol. Comput. Sci. (IJITCS)*, vol. 16, no. 6, pp. 61–73, 2024.
[5] R. Sebti, S. Zroug, L. Kahloul, and S. Benharzallah, "A deep learning approach for the diabetic retinopathy detection," in *Proc. Int. Conf. Smart City Applications*, pp. 459–469, Springer, 2021.
[6] A. K. Gangwar and V. Ravi, "Diabetic retinopathy detection using transfer learning and deep learning," in *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020)*, vol. 1, pp. 679–689, Springer, 2021.
[7] M. K. Jabbar, J. Yan, H. Xu, Z. Ur Rehman, and A. Jabbar, "Transfer learning-based model for diabetic retinopathy diagnosis using retinal images," *Brain Sci.*, vol. 12, no. 5, p. 535, 2022.
[8] S. Ghosh and A. Chatterjee, "Transfer-ensemble learning based deep convolutional neural networks for diabetic retinopathy classification," *arXiv preprint* arXiv:2308.00525, 2023.
[9] A. Bilal, X. Liu, M. Shafiq, Z. Ahmed, and H. Long, "NIMEQ-SACNet: A novel self-attention precision medicine model for vision-threatening diabetic retinopathy using image data," *Comput. Biol. Med.*, vol. 171, p. 108099, 2024.
[10] S. Akhtar, S. Aftab, M. Ahmad, and B. Ihnaini, "A classification framework for diabetic retinopathy detection using transfer learning," in *Proc. 2024 2nd Int. Conf. Cyber Resilience (ICCR)*, pp. 1–5, IEEE, 2024.
[11] M. Ahmad, M. Alfayad, S. Aftab, M. A. Khan, A. Fatima, B. Shoaib, and N. S. Elmitwal, "Data and machine learning fusion architecture for cardiovascular disease prediction," *Comput. Mater. Contin.*, vol. 69, no. 2, 2021.
[12] E. AbdelMaksoud, S. Barakat, and M. Elmogy, "A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique," *Med. Biol. Eng. Comput.*, vol. 60, no. 7, pp. 2015–2038, 2022.
[13] A. Bora et al., "Predicting the risk of developing diabetic retinopathy using deep learning," *Lancet Digit. Health*, vol. 3, no. 1, pp. e10–e19, 2021.
[14] F. Li et al., "Deep learning-based automated detection for diabetic retinopathy and diabetic macular oedema in retinal fundus photographs," *Eye*, vol. 36, no. 7, pp. 1433–1441, 2022.
[15] S. Akhtar, S. Aftab, M. Ahmad, and A. Akhtar, "Diabetic retinopathy severity grading using transfer learning techniques," *Int. J. Eng. Manuf. (IJEM)*, vol. 14, no. 6, pp. 41–53, 2024.
[16] L. K. Singh, H. Garg, and M. Khanna, "An artificial intelligence-based smart system for early glaucoma recognition using OCT images," in *Research Anthology on Improving Medical Imaging Techniques for Analysis and Intervention*, pp. 1424–1454, IGI Global, 2023.
[17] L. K. Singh, M. Khanna, and S. Thawkar, "A novel hybrid robust architecture for automatic screening of glaucoma using fundus photos, built on feature selection and machine learning-nature driven computing," *Expert Syst.*, vol. 39, no. 10, p. e13069, 2022.
[18] S. Abbas et al., "Data and ensemble machine learning fusion based intelligent software defect prediction system," *Comput. Mater. Contin.*, vol. 75, no. 3, 2023.
[19] M. Khanna, L. K. Singh, and H. Garg, "A novel approach for human diseases prediction using nature inspired computing & machine learning approach," *Multimed. Tools Appl.*, vol. 83, no. 6, pp. 17773–17809, 2024.
[20] L. K. Singh, H. Garg, and Pooja, "Automated glaucoma type identification using machine learning or deep learning techniques," in *Adva. Mach. Intell. Interact. Med. Image Anal.*, pp. 241–263, 2020.
[21] L. K. Singh, M. Khanna, S. Thawkar, and R. Singh, "Nature-inspired computing and machine learning based classification approach for glaucoma in retinal fundus images," *Multimed. Tools Appl.*, vol. 82, no. 27, pp. 42851–42899, 2023.

[22] S. H. Kassani et al., "Diabetic retinopathy classification using a modified xception architecture," in *Proc. 2019 IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, pp. 1–6, IEEE, 2019.
[23] [23] A. Bilal, L. Zhu, A. Deng, H. Lu, and N. Wu, "AI-based automatic detection and classification of diabetic retinopathy using U-Net and deep learning," *Symmetry*, vol. 14, no. 7, p. 1427, 2022.
[24] P. M. Ebin and P. Ranjana, "A variant binary classification model for no-DR mild-DR detection using CLAHE images with transfer learning," in *Proc. 2022 Int. Conf. Comput., Commun., Security Intell. Syst. (IC3SIS)*, pp. 1–5, IEEE, 2022.
[25] C. Bhardwaj, S. Jain, and M. Sood, "Transfer learning based robust automatic detection system for diabetic retinopathy grading," *Neural Comput. Appl.*, vol. 33, no. 20, pp. 13999–14019, 2021.
[26] W. L. Alyoubi, M. F. Abulkhair, and W. M. Shalash, "Diabetic retinopathy fundus image classification and lesions localization system using deep learning," *Sensors*, vol. 21, no. 11, p. 3704, 2021.
[27] G. Çinarer, K. Kilic, and T. Parlar, "A deep transfer learning framework for the staging of diabetic retinopathy," *J. Sci. Rep.-A.*, no. 051, pp. 106–119, 2022.
[28] B. Kotiyal and H. Pathak, "Diabetic retinopathy binary image classification using Pyspark," *Int. J. Math. Eng. Manag. Sci.*, vol. 7, no. 5, p. 624, 2022.
[29] A. Skouta, A. Elmoufidi, S. Jai-Andaloussi, and O. Ochetto, "Automated binary classification of diabetic retinopathy by convolutional neural networks," in *Advances on Smart and Soft Computing: Proc. ICACIn 2020*, pp. 177–187, Springer, 2021.
[30] P. Uppamma and S. Bhattacharya, "Diabetic retinopathy detection: A blockchain and African vulture optimization algorithm-based deep learning framework," *Electronics*, vol. 12, no. 3, p. 742, 2023.
[31] N. Sikder et al., "Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images," *Symmetry*, vol. 13, no. 4, p. 670, 2021.
[32] M. Ghazal, S. S. Ali, A. H. Mahmoud, A. M. Shalaby, and A. El-Baz, "Accurate detection of non-proliferative diabetic retinopathy in optical coherence tomography images using convolutional neural networks," *IEEE Access*, vol. 8, pp. 34387–34397, 2020.
[33] H. Liu et al., "Hybrid model structure for diabetic retinopathy classification," *J. Healthc. Eng.*, vol. 2020, 2020.
[34] A. Jabbar et al., "A lesion-based diabetic retinopathy detection through hybrid deep learning model," *IEEE Access*, 2024.
[35] A. M. Moustari, Y. Brik, B. Attallah, and R. Bouaouina, "Two-stage deep learning classification for diabetic retinopathy using gradient weighted class activation mapping," *Automatika*, vol. 65, no. 3, pp. 1284–1299, 2024.
[36] H. Shakibania, S. Raoufi, B. Pourafkham, H. Khotanlou, and M. Mansoorizadeh, "Dual branch deep learning network for detection and stage grading of diabetic retinopathy," *Biomed. Signal Process. Control.*, vol. 93, p. 106168, 2024.
[37] L. Dai et al., "A deep learning system for predicting time to progression of diabetic retinopathy," *Nat. Med.*, vol. 30, no. 2, pp. 584–594, 2024.
[38] Y. Jin et al., "Deep learning-driven automated quality assessment of ultra-widefield optical coherence tomography angiography images for diabetic retinopathy," *Vis. Comput.*, pp. 1–11, 2024.