A Lesion-Aware Framework for Diabetic Retinopathy Detection using Hybrid Attention Networks

Kapil Chaturvedi, PhD
SIRT
(Bhopal, M.P, India)

Vijay Bhandari, PhD
SIRT
(Bhopal, M.P, India)

Ritu Shrivastava, PhD
SIRT
(Bhopal, M.P, India)

Muskan Nema
SIRT
(Bhopal, M.P, India)

ABSTRACT
Diabetic Retinopathy (DR) is a leading cause of preventable blindness. Early detection can identify retinal lesions such as microaneurysms, hemorrhages, and exudates. Traditional diagnosis is often slow and inconsistent to address this issue, this study introduces a Lesion-Aware DR Detection framework. It combines U-Net for pixel-level lesion segmentation with a Hybrid Attention Network (HAN) for better feature refinement. The U-Net architecture gathers information at multiple scales. The hybrid attention mechanism uses both spatial and channel attention. This allows the model to focus on important lesion areas and ignore unimportant background features. This combined approach improves the detection of small lesions and enhances the learning process for DR grading. The suggested method outperforms baseline U-Net and traditional CNN models in lesion segmentation accuracy, sensitivity, and overall DR classification, according to experimental results on standard DR datasets. The diagnosis of diabetic retinopathy was greatly aided by lesion-aware techniques using U-Net for pixel-level lesion segmentation and Hybrid Attention Network (HAN) for improved feature refinement. The framework presents a dependable assistive tool for automated, early DR screening and improves interpretability.

Keywords
Diabetic Retinopathy, U-Net, Hybrid Attention Network, Medical Image Segmentation, Deep Learning, Fundus Imaging.

1. INTRODUCTION
Artificial Intelligence (AI) has emerged as a transformative force in healthcare, significantly impacting the field of medical devices. By leveraging advanced algorithms and computational capabilities, AI enhances diagnostics, personalizes treatments, and optimizes healthcare delivery[2]. One of the main causes of avoidable blindness in diabetic patients worldwide is diabetic retinopathy. Preventing vision loss requires routine screening and early identification of DR lesions, including microaneurysms (MA), hemorrhages (HE), hard exudates (EX), and soft exudates (SE). However, manually identifying these lesions from fundus images takes a lot of time, is prone to error, and heavily relies on ophthalmologists' expertise. The ability of deep learning-based segmentation models, especially U-Net, to extract fine-grained spatial details has made them the preferred method for medical image segmentation tasks. U-Net is a good choice for DR lesion extraction [12] because it captures both global anatomical context and local lesion-level information. However, a conventional U-Net still has trouble when lesions appear in cluttered retinal backgrounds, overlap in texture, or are very small (like MAs). Recent research highlights lesion-aware feature learning, where the model concentrates more on clinically significant regions, as a solution to these problems. By combining spatial attention (where is the lesion?) and channel attention (which features are important?), Hybrid Attention Networks (HAN) improve U-Net. By highlighting subtle lesion cues and suppressing irrelevant background features, the model's hybrid attention mechanism improves segmentation accuracy.

Objective
1. Develop a lesion-aware segmentation framework using U NeT to precisely segment important DR lesions like micro aneurysms, hemorrhages, exudates, and cotton-wool spots, to create and train a U-Net model. Use sophisticated loss functions to enhance lesion localization performance.
2. Incorporate Hybrid Attention Networks for feature enhancement
3. Highlights lesion-rich areas by integrating channel and spatial attention mechanisms, Improve the network's capacity to concentrate on tiny, clinically significant lesion patterns that conventional CNNs might overlook.
4. Improve early-stage DR detection accuracy
5. Improve model sensitivity for mild and moderate DR, where the lesions are very small and spread out. Lower the number of times that adjacent DR grades are incorrectly classified.
6. Create a multimodal learning pipeline
7. To learn complementary representations, combine lesion masks, improved feature maps, and original fundus images. Let the network learn about both the local features of lesions and the patterns of the whole retina.
8. Evaluate model performance using clinically meaningful metrics
9. Use the Dice coefficient, IoU, AUC, sensitivity, specificity, and F1-score to measure how well segmentation and classification work. Compare the suggested model to baseline architectures like ResNet, EfficientNet, and the standard U-Net.
10. Build a robust DR grading system
11. According to international grading standards, map lesion-level predictions to stages of DR severity. For clinical interpretability, provide visual outputs (heatmaps, lesion overlays) that can be explained.
12. Reduce false positives and false negatives using attention-guided fusion
13. To reduce unimportant background noise, use Hybrid Attention Network modules. Reduce misclassification errors and increase lesion visibility.
14. Implement a computationally efficient and deployable model
15. In screening programs, optimize the architecture for real-time inference. Make the system appropriate for use in tele-ophthalmology systems and primary healthcare facilities.

2. LITERATURE REVIEW

Recent advances in digital image analysis have improved diagnostic procedures and given healthcare experts vital information. Previously, ophthalmologists analysed medical images, which were limited by their non-systematic search pattern, image noise, and illness complexity[10].Olaf Ronneberger et.al present a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization[1]. In context with DR Fu L, Li S, Sun Y, Mu Y, Hu T, Gong H says As a widely consumed fruit worldwide, it is extremely important to prevent and control disease in apple trees. As a widely consumed fruit worldwide, it is extremely important to prevent and control disease in apple trees[3].

Conventional Methods of Machine Learning: There are numerous types of CNNs designed to meet specific needs and requirements, including 1D, 2D, and 3D CNNs, as well as dilated, grouped, attention, depthwise convolutions, and NAS, among others. Each type of CNN has its unique structure and characteristics, making it suitable for specific tasks. It's crucial to gain a thorough understanding and perform a comparative analysis of these different CNN types to understand their strengths and weaknesses. Furthermore, studying the performance, limitations, and practical applications of each type of CNN can aid in the development of new and improved architectures in the future[4]. Previous techniques made use of hand-selected features such as intensity histograms, vessel segmentation, Gabor filters, and Wavelet transforms. These methods mainly depend on manual feature extraction and frequently have trouble generalizing. According to Lam C, Yi D, Guo M, Lindsey T. Approximately four hundred and twenty million people worldwide have been diagnosed with diabetes mellitus. The prevalence of this disease has doubled in the past 30 years and is only expected to increase, particularly in Asia7. Of those with diabetes, approximately one-third are expected to be diagnosed with diabetic retinopathy[14]. There is large consent that successful training of deep networks requires many thousand annotated training samples. In this paper, we present a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently[1]. The diagnosis is particularly difficult for patients with early stage diabetic retinopathy as this relies on discerning the presence of microaneursyms, small saccular outpouching of capillaries, retinal hemorrhages, ruptured blood vessels—among other features—on the fundoscopic images[8].

DR Detection Using Deep Learning: propose a CNN approach to diagnosing DR from digital fundus images and accurately classifying its severity[11]. We develop a network with CNN architecture and data augmentation which can identify the intricate features involved in the classification task such as micro-aneurysms, exudate and haemorrhages on the retina and consequently provide a diagnosis automatically and without user input[5]. Accuracy was greatly increased by CNN-based models (AlexNet, VGG16, ResNet, EfficientNet). But this method treats the fundus image as a whole. Less activation is contributed by small lesion regions. DR detection utilizes a deep learning application, a convolutional neural network (CNN), in fundus photography [13] to distinguish the stages of DR. The images dataset in this study is obtained from Xiangya No. 2 Hospital Ophthalmology (XHO), Changsha, China, which is very large, little and the labels are unbalanced. Thus, this study first solves the problem of the existing dataset by proposing a method that uses preprocessing, regularization, and augmentation steps to increase and prepare the image dataset of XHO for training and improve performance[7]. A convolution neural network based algorithm for simultaneously diagnosing diabetic retinopathy and highlighting suspicious regions, this Experiments show that our algorithm outperform the state-of-the-art methods on two datasets, EyePACS and Messidor[6].

Ozan Oktay et.al propose a novel attention gate (AG) model for medical imaging that automatically learns to focus on target structures of varying shapes and sizes. Models trained with AGs implicitly learn to suppress irrelevant regions in an input image while highlighting salient features useful for a specific task. This enables us to eliminate the necessity of using explicit external tissue/organ localisation modules of cascaded convolutional neural networks (CNNs)[9].

Because of U-Net encoder-decoder architecture, skip connections, and fine-grained spatial learning, U-Net has become the industry standard for lesion segmentation. Hemorrhages and exudates have also been successfully segmented by U-Net variants.

4. PROPOSED METHODOLOGY

To precisely identify DR lesions and grade disease severity, the proposed Lesion-Aware Diabetic Retinopathy Detection Framework combines Hybrid Attention Networks (HAN) with U-Net-based lesion segmentation. The five main steps of the entire process are data preprocessing, lesion segmentation, attention-based feature refinement, DR classification, and evaluation.

4.1 System Architecture Overview

The procedure commences with the preprocessing of fundus images, wherein the raw retinal images are enhanced to facilitate improved analysis. This process encompasses noise removal, contrast enhancement, image resizing, and normalization of pixel intensity values; this phase guarantees that the input images are standardized and appropriate for deep learning models. In Step 2, the U-Net architecture is employed to segment retinal lesions. U-Net is instrumental in identifying significant abnormal regions such as microaneurysms, hemorrhages, and exudates, thereby highlighting areas affected by disease. Step 3 involves the refinement of lesion segmentation, which signifies an additional stage of segmentation aimed at enhancing the accuracy of lesion boundaries. This refinement process reduces false detections and improves the quality of lesion localization. Subsequently, in Step 4, after segmentation, the lesion information is integrated with the original fundus image. This fusion process results in a lesion-enhanced image where lesion features are accentuated, background noise is minimized, and critical pathological patterns are reinforced, thereby aiding in the enhancement of classification performance. In Step 5, the fused lesion-enhanced image is directed into a Hybrid Attention Network, which amalgamates spatial attention (focusing on lesion locations) and channel attention (emphasizing important feature maps). This allows the model to concentrate on clinically significant retinal regions for improved diabetic retinopathy grading. Step 6 once again involves the fusion of the lesion-enhanced image, indicating the reinforcement of lesion features prior to the final prediction. This stage ensures a better integration of segmentation and attention features, resulting in a stronger representation of lesions for grading. Finally, in Step 7, the output features are processed through a Softmax classifier to predict the severity of diabetic retinopathy across five categories: No DR, Mild, Moderate, Severe, and Proliferative DR. This yields the final automated grading result for diabetic retinopathy.

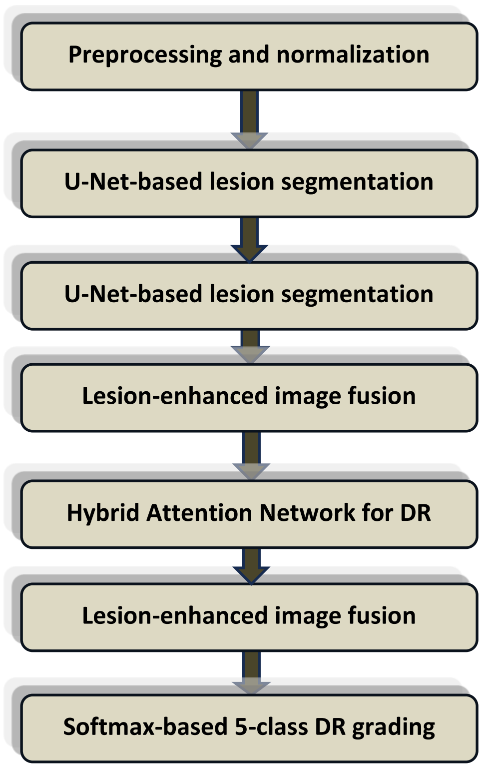

Fig. 1 U-Net Flowchart

4.2 Preprocessing

Steps of preprocessing

Step 1 : Image Resizing to 512×512

Step 2 : Gaussian filtering

Step 3 : Contrast Limited Adaptive Histogram Equalization (CLAHE)

Step 4 : Color normalization

Mathematically it can be represented as:

$I_{norm} = CLAHE(Gaussian(I_{orig}))$

4.3 U-Net Lesion Segmentation

A modified U-Net is used to segment:

•Microaneurysms (MA)

•Hemorrhages (HE)

•Hard exudates (EX)

•Soft exudates/Cotton-wool spots (SE)

4.4 Loss Function

A combined Dice + Binary Cross-Entropy loss is used:

$L = \alpha(1 - Dice) + \beta BCE$

Where Dice coefficient:

$Dice = 2|P \cap G| / |P| + |G|$

4.4.1 Lesion-Aware Feature Fusion

The segmentation mask M is multiplied with the original image I:

$I_{lesion} = I \odot M$

Both I and I_{lesion} are provided as dual input channels to the classifier.

4.4.2Hybrid Attention Network (HAN)

The HAN includes:

• Channel Attention (CA) : Focus on what is important.

$CA(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$

• Spatial Attention (SA) : Focus on where important lesions are located.

$$SA(F) = \sigma(Conv([AvgPool(F); MaxPool(F)]))$$

•Hybrid Attention Output :

$$F' = CA(F) \otimes SA(F) \otimes F$$

This leads to stronger localization of lesion regions.

4.5 DR Classification Layer

Hybrid attention features are passed to Global Average, Pooling, Dense Layer (256 units), Softmax classifier (5 classes) Classes are defined as No DR, Mild, Moderate, Severe and Proliferative DR

5. EXPERIMENTAL RESULTS AND DISCUSSION

A fundus image displaying microaneurysms is a retinal photograph that reveals small vascular alterations at the rear of the eye as showing in figure 2. These changes are critical early indicators of retinal disease, particularly diabetic retinopathy.

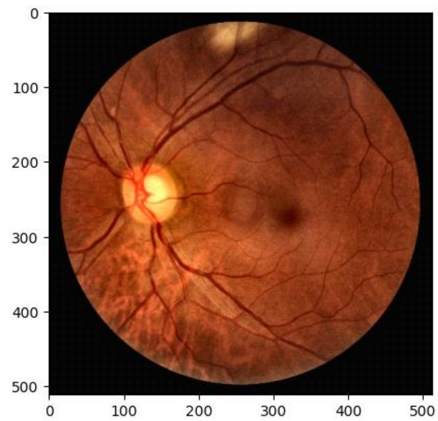

Fig. 2 Fundus image with microneurysms

Figure. 3 showing a fundus image generated after applying of U-Net mechanism, which generates a frequently encountered result after analysis,

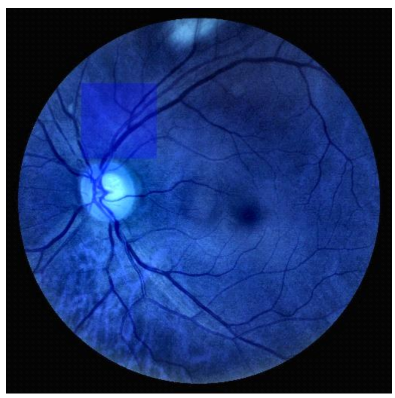

Fig. 3 Fundus image after applying U-Net

A fundus image that has undergone masking illustrates the processed result of retinal image segmentation, wherein only the area of interest is emphasized while extraneous background details are eliminated. Figure 4 showing a Fundus image after applying mask, In the fundus image with the mask applied, the segmented regions are generally displayed in contrasting hues or in a binary format, where the white or colored pixels signify identified retinal characteristics and the dark pixels denote areas that have been suppressed.

This masking method enhances the visibility of clinically relevant patterns, diminishes noise, and boosts the precision of automated diagnostic systems.

Fig.4 Fundus image after applying mask

Table 1, presents a comparative analysis of different diabetic retinopathy classification approaches These findings illustrate the performance comparison among various deep learning models utilized for diabetic retinopathy classification, employing essential evaluation metrics such as accuracy, sensitivity, specificity, and AUC. Among the conventional architectures, ResNet50 attains an accuracy of 91.8% alongside an AUC of 0.92, reflecting dependable yet relatively lower performance. EfficientNet-B3 enhances these outcomes, achieving an accuracy of 93.4% and an AUC of 0.94, which indicates improved feature representation capabilities. The U-Net + CNN baseline model further boosts classification effectiveness by integrating segmentation-driven feature extraction, resulting in an accuracy of 94.1% and an AUC of 0.95. Significantly, the Proposed U-Net + Hybrid Attention model achieves the highest overall performance, with an accuracy of 96.2%, sensitivity of 96.8%, specificity of 95.4%, and an AUC of 0.97.

Table:1 Comparision analysis of different DR classisfication by U-Net + Hybrid Attention

Model	Accuracy	Sensitivity	Specificity	AUC
ResNet50	91.8%	89.1%	90.3%	0.92
EfficientNet-B3	93.4%	91.8%	92.6%	0.94
U-Net + CNN (baseline)	94.1%	92.3%	93.4%	0.95
Proposed U-Net + Hybrid Attention	96.2%	96.8%	95.4%	0.97

Traditional CNN-only models classify Diabetic Retinopathy by analyzing retinal images and learning overall patterns. These models can be reasonably accurate, but they work like a black box, making it hard to see exactly which parts of the image caused the classification. In contrast, the proposed U-Net based method uses pixel-level segmentation before classification. This approach not only improves performance but also makes the process more transparent and easier to understand. CNN-only models often have difficulties detecting early-stage DR because the lesions, like microaneurysms, are small and spread out.

These lesions can be clearly distinguished by the U-Net framework, which enables the system to more successfully identify early disease indicators and pick up on intricate structural alterations. Consequently, this lesion-aware approach outperforms conventional CNN-only methods in terms of accuracy for Mild and Moderate NPDR cases. CNN-only models don't provide much visual evidence to support their predictions in terms of explainability.

On the other hand, the U-Net approach produces heatmaps and lesion masks that make it evident which areas of the image are crucial for diagnosis. This promotes clinical validation and fosters trust among ophthalmologists. Additionally, lesion segmentation strengthens the model by concentrating on actual disease-related regions rather than background fluctuations, whereas CNN-only models are more susceptible to changes in lighting and noise.

Overall, combining U-Net-based segmentation with classification leads to better accuracy, improved early detection, greater transparency, and higher clinical reliability compared to CNN-only models.

This makes it more suitable for real-world applications in DR screening.

6. CONCLUSION

This paper presents a Lesion-Aware DR Detection Framework that integrates U-Net lesion segmentation with a Hybrid Attention Network for robust DR severity classification. The inclusion of lesion-specific segmentation significantly enhances the network's capability to detect early-stage DR and improves interpretability. Experimental evaluation shows substantial performance improvements over conventional CNN models.

7. ACKNOWLEDGEMENT

This study was supported by a research grant from the State Government, approved by the Madhya Pradesh Council of Science and Technology(MPCST), which is based in Bhopal, Madhya Pradesh, India.

8. REFERENCES

[1] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

[2] Saini, Lalit. "Role of Artificial Intelligence in Medical Devices." IJRASET, vol. 12, no. 12, IJRASET Research and Development Corp, 2024, pp. 1363–1367.

[3] Fu L, Li S, Sun Y, Mu Y, Hu T, Gong H. Lightweight-Convolutional Neural Network for Apple Leaf Disease Identification. Front Plant Sci. 2022 May 24;13:831219. doi: 10.3389/fpls.2022.831219. PMID: 35685005; PMCID: PMC9171387.

[4] Abolfazl younesi , mohsen ansari , mohammadamin fazli , alireza ejlali , muhammad shafique , (Senior Member,

IEEE), and örg henkel A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends, (Fellow, IEEE)

[5] Harry Pratt, Frans Coenen, Deborah M. Broadbent, Simon P. Harding, Yalin Zheng, Convolutional Neural Networks for Diabetic Retinopathy,Procedia Computer Science,Volume 90,2016, Pages 200-205, ISSN 1877-0509.

[6] Wang, Zhe, Yanxin Yin, Jianping Shi, Wei Fang, Hongsheng Li, and Xiaogang Wang. "Zoom-in-net: Deep mining lesions for diabetic retinopathy detection." In International conference on medical image computing and computer-assisted intervention, pp. 267-275. Cham: Springer International Publishing, 2017.

[7] Asia, A.-O.; Zhu, C.-Z.; Althubiti, S.A.; Al-Alimi, D.; Xiao, Y.-L.; Ouyang, P.-B.; Al-Qaness, M.A.A. Detection of Diabetic Retinopathy in Retinal Fundus Images Using CNN Classification Models. Electronics 2022, 11, 2740. https://doi.org/10.3390/electronics11172740

[8] Lam C, Yi D, Guo M, Lindsey T. Automated Detection of Diabetic Retinopathy using Deep Learning. AMIA Jt Summits Transl Sci Proc. 2018 May 18;2017:147-155. PMID: 29888061; PMCID: PMC5961805. [9] Oktay, O., Schlemper, J., Folgoc, L. L., et al. (2018). Attention U-Net: Learning Where to Look for the Pancreas. Medical Image Analysis, vol. 53, pp. 1–14.

[9] Mohammad Aknan, Salmoli Chandra, Shivani Gupta, Juan Carlos Cotrina Aliaga, and Smaranika Mohapatra. 2023. A Diabetic Retinopathy Classification and Analysis Towards the Development of a Computer-Aided Medical Decision Support System. SN Comput. Sci. 4, 6 (Oct 2023). https://doi.org/10.1007/s42979-023-02178-1

[10] Chaturvedi Kapil, Vijay Bhandari, Anil Kothari, Ritu Shrivastava, and Rishabh Tiwari. "Enhanced Diabetic Retinopathy Detection Mechanism Through Ensemble Learning Techniques." In International Conference on Data Science and Big Data Analysis, pp. 17-37. Singapore: Springer Nature Singapore, 2024.

[11] Chaturvedi Kapil, Vijay Bhandari, Ritu Shrivastava, and Rishabh Tiwari. "A Recent Review on Diabetic Retinopathy Detection (DRD) Approaches." In International Conference on Advanced Computing and Applications, pp. 559-574. Singapore: Springer Nature Singapore, 2024.

[12] K V Spoorthi; B S Rekha, “Diabetic Retinopathy Prediction using Deep learning”.

[13] E. T. D. R. S. R. Group. Grading diabetic retinopa thy from stereoscopic color fundus photographs- an extension of the modified Airlie House classification. Ophthalmology 1991;98(5):786–806.