Computerized Diagnosis of Diabetic Retinopathy using Dual Attention Mechanisms

T. Monisha Birlin[1], C. Divya[2]
[1]*Research Scholar, Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India*
[2]*Assistant Professor, Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India*

***Abstract*—Diabetic retinopathy (DR) is a leading cause of preventable vision loss, and computerized grading of retinal fundus photographs can support large-scale screening. DR grading is difficult because the decisive lesions (microaneurysms, haemorrhages, and exudates) are small, scattered, and unevenly distributed across severity grades. Attention mechanisms address this by teaching a convolutional neural network where to look and which features to trust. This paper reviews dual attention mechanisms for DR diagnosis, that is, designs that combine two complementary attention branches, most often channel attention ("what") and spatial attention ("where"), or class-agnostic global attention and class-specific category attention. We explain the building blocks (squeeze-and-excitation, CBAM, and the dual attention network), group ten published DR models into six design families, and compare their reported results on EyePACS, APTOS 2019, Messidor, Messidor-2, IDRiD, and DDR. In the two ablation studies reviewed, combining two attention branches improved accuracy by 1.4 to 3.1 percentage points over the better single branch. However, results are reported under different datasets, splits, and metrics; few studies test on external data, and performance falls on unseen datasets. We close with recommendations for fair evaluation: quadratic weighted kappa, patient- and duplicate-aware splits, external validation and full reporting of ablations.**

***Index Terms*—Attention mechanism, Channel attention, Deep learning, Diabetic retinopathy, Dual attention, Fundus images, Spatial attention**

I. INTRODUCTION

Diabetic retinopathy (DR) is a microvascular complication of diabetes and one of the main causes of preventable blindness in working-age adults. About one in five people with diabetes has DR, and the number of people affected worldwide is projected to reach about 160 million by 2045 [1]. Because early DR has no symptoms, regular screening with colour fundus photographs is recommended, and the international clinical scale grades each eye into five levels: no DR, mild, moderate and severe non-proliferative DR, and proliferative DR [2]. In many countries, including India, there are too few eye specialists to grade every screening image by hand, which has motivated computerized diagnosis.

Deep convolutional neural networks (CNNs) have reached specialist-level performance in detecting referable DR [3]. Fine-grained grading into five levels remains harder for three reasons. First, the lesions that separate neighbouring grades, such as a few microaneurysms in mild DR, occupy a tiny fraction of a high-resolution image. Second, public datasets are strongly imbalanced, with many healthy eyes and few severe or proliferative cases. Third, images differ in camera, illumination and quality. A plain CNN that pools the whole image into one feature vector can easily overlook small lesions or be dominated by the majority class.

Attention mechanisms were introduced to help networks concentrate on the informative parts of their input [4, 5]. In DR grading, an attention module can highlight lesion regions (spatial attention), emphasise feature channels that respond to lesions (channel attention), or focus on features that discriminate each severity grade (category attention). Early work such as Zoom-in-Net used attention maps to localise suspicious regions and re-examine them at high resolution [6]. Many recent models combine two complementary attention branches, which we call dual attention mechanisms.

This paper reviews dual attention mechanisms for computerized DR diagnosis. Its contributions are:

(i) a plain explanation of the attention building blocks that DR models reuse;
(ii) a grouping of published dual attention DR models into six design families;
(iii) a side-by-side comparison of their reported results, including ablation evidence on whether the second attention branch helps; and
(iv) a discussion of evaluation problems and recommendations for future studies.

The results quoted in this review are those reported by the original authors; no new experiments were performed.

II. DATASETS AND EVALUATION METRICS

A. Public datasets

Table 1 summarises the public fundus datasets used by the reviewed studies. EyePACS, released through a Kaggle competition in 2015, is the largest [7]. APTOS 2019 is a smaller five-grade dataset from a later Kaggle competition [8]. Messidor provides 1,200 images with four retinopathy grades and is often used for referable versus non-referable classification [9]. IDRiD, collected at an eye clinic in Nanded, India, provides DR and diabetic macular edema (DME) grades and pixel-level lesion masks [10]. DDR contains 13,673 images, including an ungradable class [11].

Table 1. Public fundus datasets used for DR grading in the reviewed studies

Dataset	Images	Labels	Typical use
EyePACS (Kaggle 2015) [7]	35,126 train + 53,576 tests	5 DR grades	Large-scale training and testing
APTOS 2019 [8]	3,662 labelled	5 DR grades	Five-grade grading
Messidor [9]	1,200	4 retinopathy grades (0–3)	Referable vs non-referable DR
Messidor-2 [9]	1,748	DR grades	External testing
IDRiD [10]	516 (413 train, 103 test)	5 DR grades + DME risk	Joint DR/DME grading; lesion masks
DDR [11]	13,673	5 DR grades + ungradable	Five-grade grading

B. Evaluation metrics

Accuracy counts exact agreement with the reference grade but treats every error alike. Because DR grades are ordered, many studies also report the quadratic weighted kappa (QWK), the official metric of the EyePACS and APTOS competitions. QWK penalises a disagreement in proportion to the square of its distance, so calling a severe eye healthy cost far more than calling it moderate:

$$\kappa = 1 - (\Sigma_{i,j}\, w_{ij}\, O_{ij}) / (\Sigma_{i,j}\, w_{ij}\, E_{ij}), \quad w_{ij} = (i-j)^2 / (K-1)^2 \qquad (1)$$

Where O is the observed confusion matrix, E is the matrix expected by chance from the marginal totals and K = 5 is the number of grades. QWK equals 1 for perfect agreement and 0 for chance agreement. For binary referable-DR tasks, the area under the ROC curve (AUC), sensitivity and specificity are used. For IDRiD, the challenge metric is the joint accuracy, the fraction of images whose DR and DME grades are both correct [12].

III. ATTENTION BUILDING BLOCKS

A. Channel attention

A CNN feature map F of size H × W × C contains C channels, each responding to a different visual pattern. The squeeze-and-excitation (SE) block [5] learns how much each channel matters. It first squeezes every channel to one number by global average pooling, then passes this vector through two fully connected layers and a sigmoid to obtain channel weights s, which rescale the channels:

$$s = \sigma(W_2\, \delta(W_1\, z)), \quad z_c = (1/HW) \Sigma_i \Sigma_j F_c(i,j), \quad F'_c = s_c \cdot F_c \qquad (2)$$

where δ is the ReLU function and σ the sigmoid. In DR images, channel attention can raise channels that respond to red lesions or bright exudates and suppress channels that respond to the optic disc or background.

B. Spatial attention

Spatial attention learns where to look. In the convolutional block attention module (CBAM) [13], the feature map is average-pooled and max-pooled across channels, the two resulting maps are stacked and passed through a 7 × 7 convolution and a sigmoid to give a weight for every location:

$$M_s(F) = \sigma(\, f^{7\times 7}([\mathrm{AvgPool}(F); \mathrm{MaxPool}(F)])\,) \qquad (3)$$

Locations containing lesions should receive high weights and the black border or uniform background low weights.

C. Sequential and parallel dual attention

CBAM applies channel attention and then spatial attention in sequence (Fig. 1a): $F' = Mc(F) \otimes F$ and $F'' = Ms(F') \otimes F'$ [13]. The dual attention network (DANet) instead runs two self-attention branches in parallel and adds their outputs (Fig. 1b) [14]. Its position attention module relates every pixel to every other pixel, so that similar lesions in distant parts of the image reinforce each other, and its channel attention module relates every channel to every other channel:

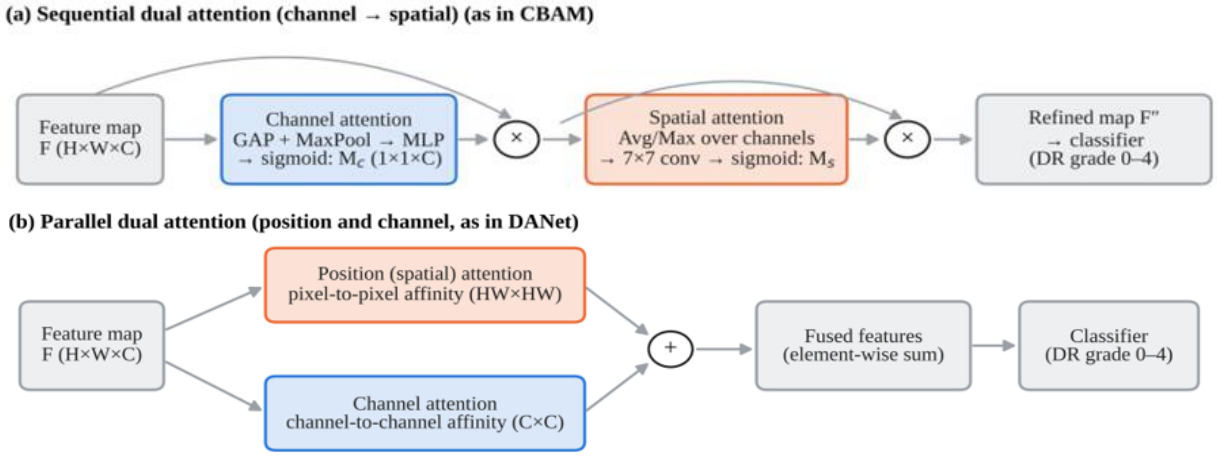


Figure 1. Two ways of combining channel and spatial attention: (a) sequential, as in CBAM [13]; (b) parallel, as in DANet [14].

IV. DUAL ATTENTION MODELS FOR DIABETIC RETINOPATHY

We grouped the dual attention DR models found in the literature into six design families according to which two kinds of attention they combine. Table 2 lists each model with its datasets and the results reported by its authors.

A. Sequential channel and spatial attention

Farag et al. [15] added a CBAM block on top of a DenseNet169 encoder and trained it on APTOS 2019. They reported 82% accuracy and a QWK of 0.888 for five-grade severity grading, and 97% accuracy with 97% sensitivity and 98.3% specificity for the binary DR/no-DR task, which shows how much harder five-grade grading is than detection. Saleem et al. [16] built a lesion-aware dual attention network on an EfficientNet-B4 backbone, in which a CBAM-style spatial module (channel pooling, 7 × 7 convolution, sigmoid) is followed by an SE channel module, and combined it with adversarial augmentation and uncertainty estimation. They reported 93.8% accuracy and a QWK of 0.912 on APTOS 2019 and 91.2% accuracy on Messidor-2 used for external validation.

$$E_j = \alpha \sum_i (s_{ji} D_i) + A_j, \quad s_{ji} = \exp(B_i \cdot C_j) / \sum_i \exp(B_i C_j) \tag{4}$$

where A is the input feature map, B, C and D are its convolutional projections and α is a learned scale; the channel branch has the same form, with channel-to-channel affinities computed from A itself [14]. Parallel self-attention captures long-range context but costs memory proportional to $(HW)^2$, which matters for high-resolution fundus images.

B. Parallel position and channel attention

Following DANet, Ramakrishnaiah et al. [17] combined a position attention mechanism and a channel attention mechanism with EfficientNet-B0 (DAM-EfficientNet-B0) and reported 99.21% accuracy on EyePACS and 99.47% on Messidor. Bhati et al. [18] proposed IDANet, which pairs a bi-directional spatial attention block, a self-attention module that reweights every pixel, with a channel-wise parallel attention block that removes redundant channels. IDANet was evaluated on four DR grading benchmarks, and IDRiD was used to show that its attention maps highlight lesions.

C. Global and category attention

CABNet [19] targets class imbalance. Its global attention block applies channel and spatial attention to obtain class-agnostic attention maps that capture small

lesions, and its category attention block learns separate region-wise features for every DR grade so that rare grades are not swamped by common ones. Both blocks can be added to standard backbones with few extra parameters. CABNet reported 78.98% accuracy and a QWK of 0.786 on DDR, and an AUC of 0.969 with 93.1% accuracy for referable DR on Messidor. Li et al. [20] extended this idea in DACNN, adding a stochastic atrous spatial pyramid pooling module and learnable layer connections to a global attention mechanism and a category attention mechanism; they reported 88.0% accuracy and a kappa of 0.886 for five-grade grading on EyePACS, and an AUC of 0.985 with 93.8% accuracy for referable DR on Messidor. Hannan et al. [21] attached the global and category attention blocks to three lightweight backbones; on APTOS 2019 their DenseNet-169 version reached 83.6% accuracy and a kappa of 0.88, and the MobileNetV3-small version achieved 82% accuracy with only 1.6 million parameters.

D. Cross-disease attention

CANet [12] grades DR and DME together. A disease-specific attention module selects features useful for each disease, and a disease-dependent attention module passes information between the two, reflecting the clinical fact that DME is a complication of DR. With a ResNet50 backbone and only image-level labels, CANet reached a joint accuracy of 65.1% on the IDRiD challenge test set, the best result in that challenge, and an AUC of 0.963 with 92.6% accuracy for referable DR on Messidor (10-fold cross-validation).

E. Frequency-domain dual attention

Yu et al. [22] argued that ordinary attention works only in the spatial domain and ignores frequency information that may describe fine lesion texture. Their FF-ResNet-DR adds a frequency channel attention module and a frequency spatial attention module to ResNet, and reported an AUC of 0.981 for referable DR on Messidor and a joint accuracy of 64.1% on IDRiD.

F. Dual attention with multi-scale fusion

Zhang et al. [23] combined a dual weighted attention mechanism, which modulates features in both the spatial and the channel dimension, with a multi-scale feature integration module in DWAM-MSFINet. On a composite dataset assembled from several public sources they reported 82.59% top-1 accuracy, compared with 81.68% for ResNet50 and 80.26% for the Swin Transformer, and 78.6% on Messidor used as a separate test set. acc., accuracy; AUC, area under the ROC curve; QWK, quadratic weighted kappa; ‖, parallel; →, sequential. Values are as reported by the original authors under their own data splits and are not directly comparable across rows.

Table 2. Dual attention models for DR diagnosis and the results reported by their authors

Model (year)	Attention design	Backbone	Dataset(s)	Reported results
CANet (2020) [12]	Disease-specific + disease-dependent attention	ResNet50	IDRiD; Messidor	IDRiD joint acc. 65.1%; Messidor referable AUC 0.963, acc. 92.6%
CABNet (2021) [19]	Global (channel + spatial) + category attention	Standard CNNs (plug-in)	DDR; Messidor	DDR acc. 78.98%, QWK 0.786; Messidor AUC 0.969, acc. 93.1%
Farag et al. (2022) [15]	CBAM (channel → spatial)	DenseNet169	APTOS 2019	5-grade acc. 82%, QWK 0.888; binary acc. 97%
DACNN (2022) [20]	Global + category attention, sASPP	Custom	EyePACS; Messidor	EyePACS acc. 88.0%, kappa 0.886; Messidor AUC 0.985, acc. 93.8%
IDANet (2024) [18]	Bi-directional spatial ‖ channel-wise parallel attention	Plug-in block	Four DR benchmarks; IDRiD for maps	Better than compared methods (values in full text)
DAM-EfficientNet-B0 (2024) [17]	Position ‖ channel attention	EfficientNet-B0	EyePACS; Messidor	Acc. 99.21% (EyePACS), 99.47% (Messidor)
FF-ResNet-DR (2025) [22]	Frequency channel + frequency spatial attention	ResNet	Messidor; IDRiD	Messidor AUC 0.981; IDRiD joint acc. 64.1%
Hannan et al. (2025) [21]	Global (channel + spatial) + category attention	DenseNet-169, MobileNetV3-	APTOS 2019; EyePACS	APTOS acc. 83.6%, kappa 0.88 (DenseNet-169);

		small, EfficientNet-b0		EyePACS acc. 80% (EfficientNet-b0)
DWAM-MSFINet (2025) [23]	Spatial + channel weighting, multi-scale fusion	Custom	Composite set; Messidor	Top-1 acc. 82.59%; 78.6% on Messidor
Saleem et al. (2026) [16]	Spatial → SE channel (lesion-aware)	EfficientNet-B4	APTOS 2019; Messidor-2	APTOS acc. 93.8%, QWK 0.912; Messidor-2 acc. 91.2%

V. COMPARATIVE ANALYSIS AND DISCUSSION

A. Does the second attention branch help?

Only some studies isolate the contribution of each attention branch. Table 3 collects the two ablation studies in which each branch was tested alone and together on the same data. In Hannan et al. [21], adding either block to DenseNet-169 raised APTOS accuracy from 79.0% to about 82%, and combining both raised it to 83.6%, 1.4 points above the better single block. In Saleem et al. [16], dual attention was 2.6 points more accurate than spatial attention alone on APTOS 2019 and 3.1 points more accurate on Messidor-2. Both studies therefore support the idea that channel and spatial information are complementary, but the extra gain from the second branch is modest, a few percentage points, and each was measured on a single split without confidence intervals.

Table 3. Ablation results: single versus dual attention (accuracy, %)

Study / dataset	No attention	First branch only	Second branch only	Both (dual)
Hannan et al. [21], DenseNet-169, APTOS 2019	79.0	82.2 (global)	82.1 (category)	83.6
Saleem et al. [16], EfficientNet-B4, APTOS 2019	–	91.2 (spatial)	90.4 (channel)	93.8
Saleem et al. [16], EfficientNet-B4, Messidor-2	–	88.1 (spatial)	87.2 (channel)	91.2

B. Comparing reported results

Table 2 should not be read as a ranking. The models were trained and tested on different datasets, with different train/test splits, image resolutions and metrics, and some report five-grade accuracy while others report binary referable-DR results, which are much easier: Farag et al. obtained 97% accuracy for detection but 82% for five-grade grading with the same model [15]. Training choices can matter as much as architecture. Huang et al. [24] showed that a plain ResNet-50 with carefully chosen preprocessing, augmentation and training settings reached a test kappa of 0.8581 on EyePACS, above several specially designed models, including a CABNet configuration with a ResNet-50 backbone (0.8456) listed in their comparison. A fair test of an attention module therefore requires the baseline to be trained with exactly the same settings.

Some reported figures also need careful interpretation. Accuracies above 99% on EyePACS and Messidor [17] are far higher than those of other attention models on the same datasets, for example 88.0% five-grade accuracy on EyePACS for DACNN [20]. Such differences may arise from the data split, the class definitions or the evaluation protocol rather than from the attention design, so the full protocol must be reported for results to be reproducible.

C. Generalisation to new data

Performance usually falls when a model meets images from another source. DWAM-MSFINet dropped from 82.59% on its development data to 78.6% on Messidor [23], and the joint DR/DME accuracy of the best models on the IDRiD test set remains around 64–65% [12, 22]. Only a minority of the reviewed studies report external validation [16, 23]. Since a screening model will be used on cameras and populations it has not seen, external testing should be standard.

D. Data leakage and duplicates

Several studies divide a single dataset at random, for example 50% for training, 30% for validation and 20% for testing [21]. Random image-level splitting can place images of the same patient, or repeated copies of the same photograph, in both training and test sets; public fundus datasets are known to contain such repeated images [25]. This kind of data leakage makes test results look better than they would be on genuinely new patients [26]. Splits should therefore be made by patient and should keep duplicate images together.

E. Interpretability and cost

A practical advantage of attention is that the attention maps can be shown to clinicians as visual evidence. Zoom-in-Net [6], CANet [12], IDANet [18] and DACNN [20] all used attention or activation maps to show that the models focus on lesions. Attention also adds computation. Channel attention and CBAM are light, whereas parallel self-attention grows with the square of the number of pixels [14]. Lightweight designs, such as attention blocks on MobileNetV3-small with 1.6 million parameters [21], are attractive for screening on low-cost hardware.

VI. CHALLENGES AND FUTURE DIRECTIONS

Based on this review, we suggest the following for future work on dual attention DR models. (1) Report QWK, together with accuracy, macro-F1 and per-grade sensitivity, because accuracy alone hides errors on the rare but dangerous severe and proliferative grades. (2) Use patient-level and duplicate-aware splits, and preferably cross-validation with confidence intervals, rather than a single random split. (3) Validate on at least one external dataset. (4) Report a full ablation in which the baseline, each single attention branch and the dual design are trained with identical settings. (5) Release code and data splits. (6) Explore attention that combines spatial, channel, frequency and category information, and compare it fairly with transformer-based models. (7) Evaluate lightweight dual attention models on images from smartphone and low-cost cameras used in community screening.

VII. CONCLUSION

Dual attention mechanisms help CNNs focus on small DR lesions and on the features that separate severity grades. The reviewed models combine channel and spatial attention in sequence or in parallel, global and category attention, cross-disease attention, or frequency-domain attention. Where ablations were reported, combining two attention branches improved accuracy by a few percentage points over a single branch. However, the published results come from different datasets, splits and metrics, few are externally validated, and performance drops on unseen data. Consistent, leakage-free and externally validated evaluation is now the main requirement for judging which dual attention design is most useful for computerized DR screening.

REFERENCES

[1]. Z. L. Teo et al., “Global prevalence of diabetic retinopathy and projection of burden through 2045: Systematic review and meta-analysis,” *Ophthalmology*, vol. 128, no. 11, pp. 1580–1591, 2021, doi: 10.1016/j.ophtha.2021.04.027.

[2]. C. P. Wilkinson et al., “Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales,” *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, 2003, doi: 10.1016/S0161-6420(03)00475-5.

[3]. V. Gulshan et al., “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016, doi: 10.1001/jama.2016.17216.

[4]. A. Vaswani et al., “Attention is all you need,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998–6008.

[5]. J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7132–7141.

[6]. Z. Wang, Y. Yin, J. Shi, W. Fang, H. Li, and X. Wang, “Zoom-in-Net: Deep mining lesions for diabetic retinopathy detection,” in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2017, pp. 267–275.

[7]. Kaggle and EyePACS, “Diabetic Retinopathy Detection,” Kaggle Competition, 2015. [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection

[8]. Asia Pacific Tele-Ophthalmology Society, “APTOS 2019 Blindness Detection,” Kaggle Competition, 2019. [Online]. Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection

[9]. E. Decencière et al., “Feedback on a publicly distributed image database: The Messidor database,” *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, 2014, doi: 10.5566/ias.1155.

[10]. P. Porwal et al., “Indian Diabetic Retinopathy Image Dataset (IDRiD): A database for diabetic retinopathy screening research,” *Data*, vol. 3, no. 3, p. 25, 2018, doi: 10.3390/data3030025.

[11]. T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Information Sciences*, vol. 501, pp. 511–522, 2019, doi: 10.1016/j.ins.2019.06.011.

[12]. X. Li, X. Hu, L. Yu, L. Zhu, C.-W. Fu, and P.-A. Heng, "CANet: Cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1483–1493, 2020, doi: 10.1109/TMI.2019.2951844.

[13]. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. European Conf. Computer Vision (ECCV)*, 2018, pp. 3–19.

[14]. J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 3146–3154.

[15]. M. M. Farag, M. Fouad, and A. T. Abdel-Hamid, "Automatic severity classification of diabetic retinopathy based on DenseNet and convolutional block attention module," *IEEE Access*, vol. 10, pp. 38299–38308, 2022, doi: 10.1109/ACCESS.2022.3165193.

[16]. S. Saleem, R. Talib, M. K. Hanif, and M. U. Sarwar, "Clinical-guided deep learning framework for diabetic retinopathy: Integrating lesion-aware attention, adversarial augmentation, and uncertainty quantification," *Scientific Reports*, vol. 16, p. 24161, 2026, doi: 10.1038/s41598-026-58690-w.

[17]. N. K. Ramakrishnaiah, N. Rangahanumaiah, V. Anantharamaiah, B. U. M. B. Nanjundaswamy, and K. M. K. Timmappa, "Classification of diabetic retinopathy stages using dual attention mechanism with EfficientNet-B0," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 6, pp. 975–985, 2024, doi: 10.22266/ijies2024.1231.73.

[18]. A. Bhati, N. Gour, P. Khanna, A. Ojha, and N. Werghi, "An interpretable dual attention network for diabetic retinopathy grading: IDANet," *Artificial Intelligence in Medicine*, vol. 149, p. 102782, 2024, doi: 10.1016/j.artmed.2024.102782.

[19]. A. He, T. Li, N. Li, K. Wang, and H. Fu, "CABNet: Category attention block for imbalanced diabetic retinopathy grading," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 143–153, 2021, doi: 10.1109/TMI.2020.3023463.

[20]. F. Li, S. Tang, Y. Chen, and H. Zou, "Deep attentive convolutional neural network for automatic grading of imbalanced diabetic retinopathy in retinal fundus images," *Biomedical Optics Express*, vol. 13, no. 11, p. 5813, 2022, doi: 10.1364/BOE.472176.

[21]. A. Hannan, Z. Mahmood, R. Qureshi, and H. Ali, "Enhancing diabetic retinopathy classification accuracy through dual-attention mechanism in deep learning," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 13, no. 1, 2025, doi: 10.1080/21681163.2025.2539079.

[22]. C. Yu, Q. Ma, J. Li, Q. Zhang, J. Yao, B. Yan, and Z. Wang, "FF-ResNet-DR model: A deep learning model for diabetic retinopathy grading by frequency domain attention," *Electronic Research Archive*, vol. 33, no. 2, pp. 725–743, 2025, doi: 10.3934/era.2025033.

[23]. M. Zhang, Q. Liu, J. Zhan, J. Gao, D. Xie, and J. Liu, "A dual attention and multi-scale fusion network for diabetic retinopathy image analysis," *Frontiers in Medicine*, vol. 12, p. 1614046, 2025, doi: 10.3389/fmed.2025.1614046.

[24]. Y. Huang, L. Lin, P. Cheng, J. Lyu, R. Tam, and X. Tang, "Identifying the key components in ResNet-50 for diabetic retinopathy grading from fundus images: A systematic investigation," *Diagnostics*, vol. 13, no. 10, p. 1664, 2023. Originally available as arXiv:2110.14160, 2021.

[25]. A. I. Bursov and D. M. Safonova, "Study of images in open datasets of the ocular fundus in diabetic retinopathy designed for training neural network algorithms," *Siberian Journal of Clinical and Experimental Medicine*, vol. 40, no. 1, pp. 218–225, 2025, doi: 10.29001/2073-8552-2025-40-1-218-225.

[26]. S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, p. 100804, 2023, doi: 10.1016/j.patter.2023.100804.