Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 12, No. 03, September 2026.

ADVANCEMENTS IN AI-DRIVEN DIABETIC RETINOPATHY DETECTION: A SYSTEMATIC SURVEY OF METHODS, CHALLENGES AND FUTURE DIRECTIONS

Yasser A. Ali[1], Ahmed Z. Emam[2], Muhammed H. Zayyan[1], Samir Elmougy[1] and Ahmed S. Tolba[1]

(Received: 25-Mar.-2026, Revised: 31-May-2026 and 22-Jun.-2026, Accepted: 30-Jun.-2026)

ABSTRACT

Diabetic retinopathy (DR) is a leading cause of preventable vision impairment, requiring scalable and reliable screening solutions. Artificial intelligence (AI), particularly machine learning and deep learning, has significantly advanced the automation of DR classification from retinal fundus images. This paper presents a systematic review of AI-based DR classification approaches, focusing on binary classification, multi-class severity grading and lesion-based methods and introduces a taxonomy based on their clinical objectives. The review covers state-of-the-art models, including convolutional neural networks, vision transformers, attention mechanisms and ensemble techniques, as well as their pre-processing strategies and evaluation protocols. Common benchmark datasets are analyzed in terms of class distribution and their impact on performance. While current models achieve high accuracy on standard datasets, challenges remain in generalization, class imbalance and reliable detection of early-stage DR. Lesion-based approaches improve interpretability, but are often not fully integrated into grading frameworks. Overall, this review highlights key limitations and outlines future directions toward developing robust, interpretable and clinically deployable DR classification systems for real-world screening.

KEYWORDS

Diabetic Retinopathy, Automated Disease Screening, Deep Learning, Artificial Intelligence, Fundus Imaging, Explainable AI, Vision Transformers, Domain Adaptation, Class Imbalance.

1. INTRODUCTION

Diabetic Retinopathy (DR) is a significant public health issue impacting millions globally, resulting in vision impairment and blindness if not addressed, primarily affecting those with chronic diabetes [1],[2],[3]. The illness involves damage to the retinal blood vessels caused by persistent hyperglycemia, which triggers vascular leakage and capillary occlusion [4]-[5]. This process leads to specific abnormalities, including microaneurysms, hemorrhages, exudates and neovascularization [1], [6]. The quick and accurate diagnosis of DR is essential to avert vision impairment, as timely intervention can markedly enhance patient prognoses and prevent the disease from escalating to irreversible stages [2], [4], [7]-[8].

Traditional DR screening relies on manual examination of retinal fundus images by trained ophthalmologists [4], [7], [9]-[10]. Although effective, this approach is time-consuming, costly and subject to inter-observer variability, meaning that the level of agreement among specialists can vary significantly [11]-[12]. Furthermore, many regions—particularly rural and low-resource areas—lack sufficient ophthalmology specialists, leading to delayed diagnosis and treatment [6], [9]. Recent advances in artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), have transformed medical image analysis and offered promising solutions for automated DR detection [6], [9], [13]. AI-based systems can analyze retinal images efficiently, detect pathological features, classify disease severity and assist clinicians in large-scale screening programs [3], [14]. The availability of large datasets (such as EyePACS and APTOS 2019), improvements in computational power and the success of convolutional neural networks (CNNs) have accelerated research in this domain [15]-[16].

This paper presents a comprehensive review of AI techniques for the automatic detection and classification of DR. It highlights key factors that affect the performance of DR classification systems,

1. Y. A. Ali, M. H. Zayyan, Samir Elmougy and Ahmed S. Tolba are with Department of Computer Science, Faculty of Computers and Information, Mansoura University, Mansoura, Egypt. Emails: yasaliali@std.mans.edu.eg, {mhaggag, mougy, astolba}@mans.edu.eg
2. A. Z. Emam is with Department of Math & Computer Science, Faculty of Science, Menoufiya University, Menoufiya, Egypt. Email: Ahmed.emam@science.menofia.edu.eg

including image quality, feature representation, learning strategies and dataset characteristics. The reviewed studies emphasize the importance of robust AI models for early diagnosis, large-scale screening and clinical decision support. In addition, this paper summarizes current limitations, open challenges and future research directions in DR classification.

Unlike previous surveys that primarily summarize CNN-based DR classification models, this review offers a broader, more critical analysis of recent AI-driven DR detection approaches. The proposed survey introduces a unified taxonomy encompassing binary classification, multi-class grading, hybrid frameworks and lesion-based methods. Furthermore, this work critically examines emerging trends, including vision transformers (ViTs), graph neural networks, multi-modal learning and quantum-inspired approaches, while emphasizing challenges related to dataset bias, cross-dataset generalization, explainability, reproducibility and real-world clinical deployment. The review also highlights the gap between high benchmark performance and practical clinical applicability, thereby providing a more comprehensive perspective on future-research directions.

The remainder of this paper is organized as follows. Section 2 presents the systematic review methodology. Section 3 describes the general framework for automated DR classification. Section 4 presents a taxonomy of classification approaches, including binary and multi-class severity grading. Section 5 reviews existing classification methods. Section 6 summarizes the benchmark datasets commonly used for DR detection. Section 7 presents a critical discussion of the reviewed literature, highlighting methodological challenges, dataset bias, reproducibility concerns, clinical deployment barriers and emerging research directions. Finally, Section 8 concludes the paper by outlining future opportunities to advance robust, interpretable and clinically deployable DR detection systems.

2. SYSTEMATIC REVIEW METHODOLOGY

This review was conducted following a systematic literature review methodology to ensure transparency, reproducibility and comprehensive coverage of recent advances in AI-driven DR detection. Relevant studies published between 2018 and 2025 were retrieved from major scientific databases, including IEEE Xplore, Scopus, Web of Science, PubMed, ScienceDirect and SpringerLink. Search queries included combinations of keywords, such as “DR”, “DL”, “AI”, “retinal fundus images”, “CNN”, “ViT” and “DR grading”.

The inclusion criteria considered peer-reviewed studies that focused on AI-based DR detection or grading using retinal fundus images and reported quantitative evaluation metrics. Studies unrelated to DR classification, non-English publications, duplicate papers and articles lacking sufficient experimental details were excluded. The screening process involved title and abstract filtering, followed by full-text assessment. Duplicate studies were removed and the selected articles were analyzed for methodology, datasets, evaluation metrics, strengths and limitations. A PRISMA-style workflow illustrating the study selection procedure is shown in Figure 1.

2.1 Quality Assessment of Included Studies

To improve the scientific rigor of this systematic review, the included studies were qualitatively assessed rather than ranked using a numerical scoring system. The assessment considered several methodological dimensions, including dataset transparency and characteristics, annotation quality and grading procedures, the suitability of reported evaluation metrics and the handling of class imbalance. Attention was given to the validation strategy, especially the presence of external validation, the level of methodological detail provided to support reproducibility and the use of interpretability techniques, such as explainable AI methods or lesion-based analysis. In addition, the studies were evaluated with respect to their discussion of clinical relevance, practical deployment considerations and acknowledgment of limitations. This qualitative assessment supported the critical synthesis presented throughout the manuscript and facilitated a more balanced interpretation of reported performance beyond accuracy values alone.

3. THEORETICAL FRAMEWORK FOR AUTOMATED DR DETECTION

The general framework for detecting DR comprises sequential stages that transform raw retinal images into clinically meaningful diagnostic outcomes [2], [7], [17]. The process begins with data acquisition,

in which retinal fundus images are captured using specialized imaging devices, such as digital fundus cameras, either in clinical settings or from publicly available datasets, such as EyePACS, Messidor or IDRiD [2], [4], [18]. These images often vary in resolution, illumination and quality, which necessitates the next stage, pre-processing [4], [19]. Pre-processing aims to enhance image quality and standardize the input data by applying techniques, such as resizing, normalization, contrast enhancement (e.g., CLAHE), noise reduction and illumination correction [2], [19]. This stage improves the visibility of critical retinal structures and pathological signs associated with DR, such as blood vessels and lesions [1], [19].

After pre-processing, feature extraction is performed to get discriminative information from the retinal images [2], [4], [20]-[21]. Features may be handcrafted—such as texture, color and morphological descriptors—or automatically learned using DL models, which can capture complex spatial patterns related to DR lesions [1], [4], [22]. Finally, the extracted features are utilized in the classification or segmentation stage, where ML or DL models assess the presence and severity of DR or to localize specific lesions, such as microaneurysms, hemorrhages and exudates [2], [4], [7], [23]. These stages together form an integrated framework that enables accurate, efficient and automated detection of DR, supporting early diagnosis and large-scale screening programs [2], [4], [24], as shown in Figure 2.

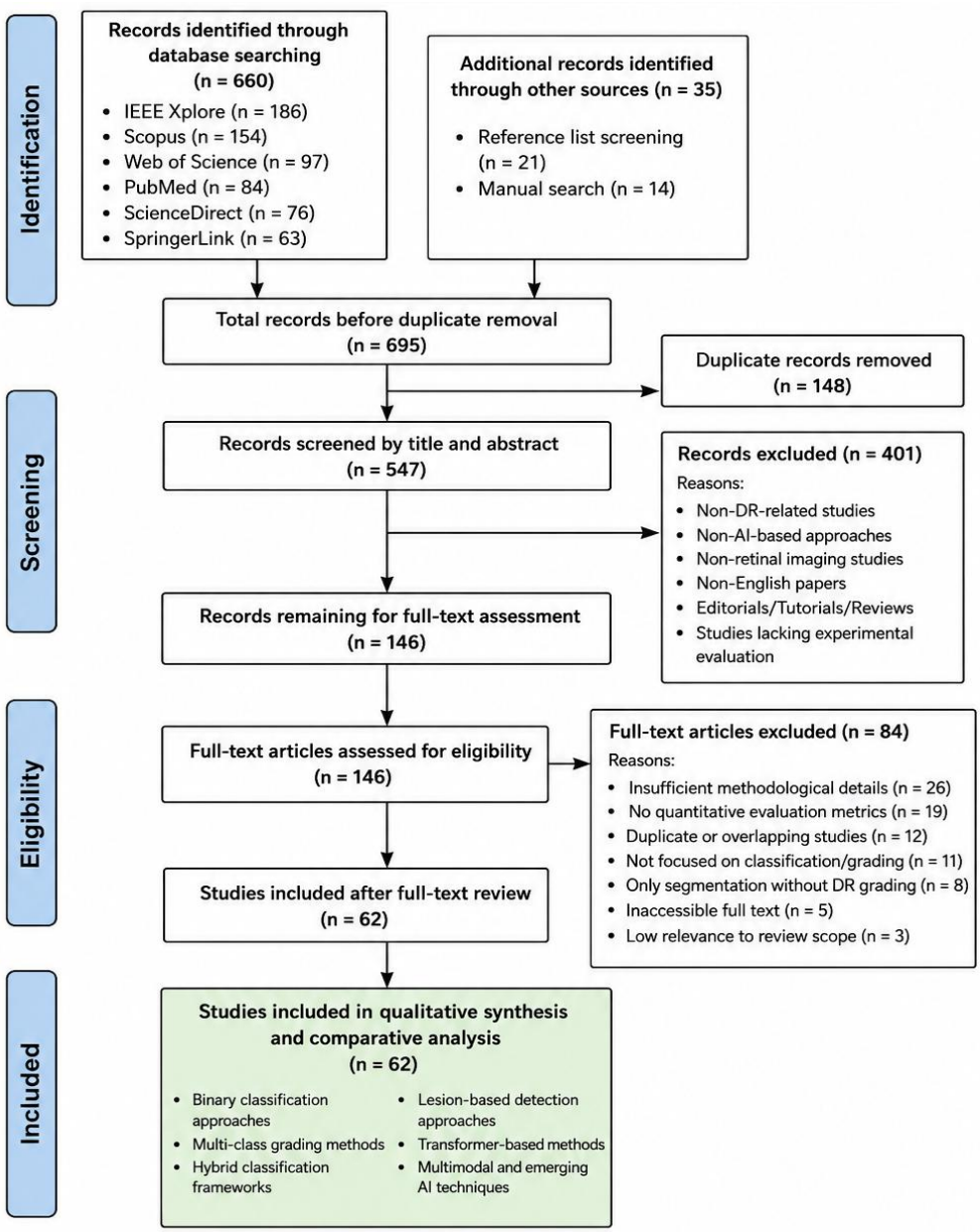


Figure 1. PRISMA-based workflow for identification, screening, eligibility assessment and inclusion of studies in the systematic review.

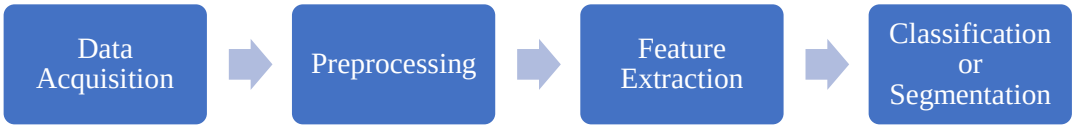


Figure 2. General framework for automated DR detection.

Although lesion segmentation plays an important role in retinal-image analysis by localizing pathological structures, such as microaneurysms, hemorrhages and exudates, the primary focus of this survey is classification-based DR detection and grading approaches. Accordingly, segmentation methods are discussed only in relation to their contribution to feature representation, interpretability and lesion-aware classification frameworks rather than as standalone segmentation systems. A comprehensive analysis of retinal lesion segmentation techniques remains an important direction for future review studies.

4. Taxonomy of DR Detection Approaches

The taxonomy of DR detection approaches, illustrated in Figure 3, is developed through a systematic analysis of the reviewed literature and reflects the major methodological and clinical directions in AI-driven DR diagnosis. The taxonomy categorizes DR detection approaches into four main groups: binary classification, multi-class severity grading, hybrid binary–multi-class frameworks and lesion-based classification approaches. Binary-classification approaches primarily focus on efficient screening tasks, such as distinguishing DR from non-DR or referable from non-referable cases, making them suitable for large-scale screening systems. In contrast, multi-class grading approaches aim to provide more clinically meaningful assessments of severity by categorizing retinal images into multiple DR stages. However, increasing grading granularity introduces greater challenges related to class imbalance, annotation variability and inter-class similarity.

Lesion-based approaches emphasize interpretability by explicitly identifying pathological retinal features, such as microaneurysms, hemorrhages and exudates. These methods provide clinically explainable predictions, but often require detailed lesion annotations and higher computational complexity. More recent hybrid and multi-modal frameworks attempt to combine the strengths of different approaches by integrating lesion localization, global contextual learning and attention-based feature extraction to improve robustness and generalization. The taxonomy also highlights the evolution of DR detection research from performance-oriented classification systems toward more explainable, generalized and clinically deployable AI frameworks. This progression reflects the growing emphasis on interpretability, fairness, cross-dataset generalization and real-world clinical applicability in modern DR detection systems.

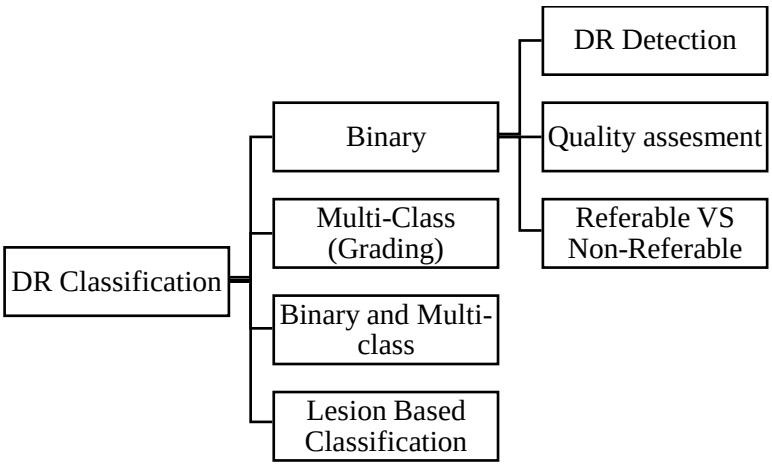


Figure 3. Taxonomy of DR detection.

Based on this taxonomy, the following sections systematically review the literature within these categories. For each category, studies are analyzed in terms of their objectives, methodologies, datasets and evaluation metrics. A critical comparison is also provided, highlighting strengths, limitations and

challenges in terms of accuracy, generalization and clinical applicability. Accordingly, Table 2 summarizes binary classification studies, Table 3 reviews multi-class severity grading, Table 4 presents combined approaches and Table 5 covers lesion-informed classification methods.

5. DR CLASSIFICATION-BASED APPROACHES

The classification of DR is crucial for early diagnosis and treatment, as it ranges from mild non-proliferative stages to severe proliferative forms that may lead to vision loss. Accurate classification helps determine the appropriate management strategies and monitor disease progression, making it a vital aspect of diabetic care [25]. DR is categorized into two main phases: non-proliferative DR (NPDR) and proliferative DR (PDR) [4]. NPDR is further divided into mild, moderate and severe stages based on the presence of specific retinal lesions. The standardized DR severity scale is summarized in Table 1, which outlines 5 distinct stages determined by lesion characteristics observed in retinal fundus images. They can also be categorized as non-referable DR (R0 and R1) and referable DR (R2, R3 and R4).

Table 1. DR grade based on the international clinical DR disease severity scale (ICDRDSS) [26].

Grade	**Description**
R0 - No DR	No visible sign of abnormalities
R1 - Mild NPDR	Only the presence of microaneurysms
R2 - Moderate NPDR	Below severe NPDR and above just microaneurysms
R3 - Severe NPDR	No signs of PDR and any of the following: 1. Each of the four quadrants has above 20 intraretinal hemorrhages 2. Two or more quadrants have definite venous beading 3. One or more quadrants with prominent IRMA
R4 - PDR	Either one or all of the following: 1. Existence of neovascularization 2. Vitreous/pre-retinal hemorrhages

5.1 Binary DR Classification

This subsection reviews studies that address binary DR classification, where retinal fundus images are categorized into two classes, such as DR *versus* non-DR, healthy *versus* unhealthy, referable *versus* non-referable DR, NPDR *versus* PDR or gradable *versus* non-gradable images. Binary classification remains widely adopted due to its effectiveness in large-scale screening and referral decision support. Table 2 summarizes representative binary DR classification studies, highlighting their pre-processing techniques, datasets, methodologies, strengths and limitations.

Hu et al. [27] proposed graph adversarial transfer learning (GATL) for DR classification to address limited annotations and domain variability across retinal imaging devices. GATL integrates self-supervised learning, graph neural network (GNN)-based feature extraction and adversarial domain adaptation for intra- and inter-domain alignment. Evaluated on the APTOS 2019 and EyePACS datasets, it achieves accuracies of 94.3% and 92.7%, respectively, with high sensitivity and specificity. Its strengths include effective domain adaptation and the utilization of unlabelled data to improve classification performance. However, the method is limited by GNN assumptions, instability in adversarial training and the complexity of retinal-image characteristics. The hierarchical DL framework (HDeep) [28] is proposed for DR classification using a combination of DenseNet-121 and EfficientNet-B7 models trained on the Kaggle APTOS dataset (5,590 retinal images). The methodology reformulates the multi-class problem into four binary classification tasks and employs image enhancement and data augmentation to address class imbalance. The model achieves a Kappa score of 0.911 and an accuracy of 85.3%, outperforming comparable approaches. Its strengths include a scalable hierarchical architecture and an effective augmentation strategy, while limitations stem from severe class imbalance and the difficulty of distinguishing between similar DR severity levels.

A hybrid CNN-based DR classification model is proposed in [4] to address the challenges of manually detecting DR in fundus images. The method combines InceptionV3 and ResNet50 for feature extraction, then concatenates the features and performs classification with a CNN. The study uses 44,119 retinal images from a public dataset across five DR severity classes, applying pre-processing and data augmentation to enhance image quality and model robustness. The proposed approach achieves 96.85% accuracy and outperforms existing methods in sensitivity, specificity, precision and F1-score. Its main

strength lies in the effective integration of multiple DL architectures for improved classification performance. However, the model is evaluated on a single dataset, which may limit its generalizability to diverse clinical populations. A faster RCNN-based DR detection framework is proposed in [2] for early diagnosis by integrating fused features extracted from fundus images. The method combines histograms of oriented gradient (HOG), Shearlet transform and RCNN features into a unified feature vector after image pre-processing and noise reduction using an adaptive filter. Experiments conducted on a dataset of 21,171 fundus images achieved 98.58% accuracy, 97.12% specificity and 95.72% sensitivity. The primary strength of the approach lies in its feature-fusion strategy, which enhances detection performance compared with existing methods. However, the model still misclassifies some DR cases and requires further training and fine-tuning to improve accuracy and reduce computational cost.

A novel retinal image-quality assessment (RIQA) standard [29] is proposed that classifies retinal images into four quality grades by integrating image quality, retinal structure and DR lesion features. The method employs an attention-dilated dense network (ADD-Net), a lightweight DenseNet-based architecture that incorporates stacked dilated convolutions and attention to enhance feature representation while maintaining low computational complexity. Using the re-labeled Kaggle retinal image quality dataset, ADD-Net achieves 97.03% accuracy, 97.06% precision and 98.31% sensitivity, outperforming existing approaches. The main strengths of the approach include the establishment of a comprehensive RIQA framework and the suitability of ADD-Net for deployment on portable devices. However, the dataset exhibits class imbalance, which may introduce bias and affect the model's generalizability. Pavithra et al. [30] proposed an ensemble DL framework for DR detection using VGG19, Inception V3 and ResNet50 for feature extraction, followed by CNN-based classification. The model is evaluated on a fundus-image dataset comprising 2,076 training, 531 validation and 231 test images, with data augmentation and image enhancement applied during pre-processing. The ensemble approach achieves an accuracy of 98.47%, outperforming the individual models. The study's main strength lies in its effective integration of multiple DL architectures for binary DR classification. However, its performance depends on the availability of high-quality fundus images, which may limit generalizability in real-world clinical settings.

Table 2. Comparison of AI-based methods for binary DR classification.

Ref.	Pre-processing	Datasets	Methods	Strengths	Limitations
[27]	Resizing (512×512)	EyePACS, APTOS	GATL	Domain adaptation and graph learning improve robustness	GNN assumptions, adversarial instability, complex retinal images
[28]	CLAHE, vessel enhancement	APTOS	HDeep (DenseNet121 + EfficientNet-B7)	Hierarchical classification strategy	Sensitive to class imbalance
[4]	Histogram equalization	OCT/Fundus	Hybrid CNN (ResNet50 + InceptionV3)	Effective feature fusion	Single-dataset evaluation
[2]	Filtering, cropping	DIARETDB1, DDR, Kaggle	Faster-RCNN + HOG + Shearlet	Excellent detection accuracy	Misclassification remains for some DR cases
[29]	Cropping, resizing	DR1, DRIMDB, Kaggle	ADD-Net	Lightweight and deployable	Dataset-quality imbalance
[30]	Normalization, cropping	Kaggle	Ensemble CNN	Ensemble improves robustness	Requires high-quality images

5.2 Multi-class DR Classification

Multi-class classification in DR involves categorizing retinal images into more than two classes, such as different disease stages or severity levels. Also, Grading in DR refers to assigning a severity level or stage to the disease based on specific criteria, such as lesions, hemorrhages, exudates or microaneurysms [20]. Table 3 summarizes the key studies reviewed in this sub-section, highlighting their pre-processing techniques, datasets, methodologies, strengths and limitations in multi-class DR classification and grading.

A novel self-supervised pre-training method, MIE-DR [31], is proposed to leverage unlabelled multi-modal ophthalmic data to learn both shared and modality-specific retinal representations for DR grading. The method is evaluated on the IDRiD and Messidor datasets, with cross-dataset validation on the EyePACS-Kaggle and Messidor-2 datasets. Experimental results demonstrated superior transfer

learning (TL) performance compared with ImageNet pre-training and other multi-modal approaches, resulting in improved DR classification and grading accuracy. The main strength of MIE-DR lies in its ability to effectively exploit multi-modal information during pre-training, thereby enhancing feature representation and generalization. However, its performance depends on the availability of multi-modal data and fine-grained DR grading remains challenging. A hybrid retinal image-enhancement framework for DR detection is proposed in [19] to address low contrast, noise and uneven illumination in fundus images. The method employs image cropping and Gaussian blurring as preprocessing steps, followed by a DL-based classification model using a modified ResNet50 architecture. Evaluation on the EyePACS and MESSIDOR datasets demonstrates substantial performance improvements, with classification accuracy increasing from 73% to 92% and from 79.2% to 93.6%, respectively. The main strength of the approach lies in its effective image-enhancement strategy, which improves feature extraction and classification performance. However, the study is limited by the need for further investigation into enhancement techniques to improve contrast and image sharpness, thereby supporting more reliable DR diagnosis. In [21], the authors propose a multi-stage patch-based deep CNN (MPDCNN) that combines the InceptionV3 and Xception models for automated DR grading. The method analyzes overlapping image patches extracted from 2,290 high-resolution fundus images, sourced from hospital and public datasets, to enhance feature representation and classification accuracy. Using five-fold cross-validation, MPDCNN achieves an accuracy of 96.2%, outperforming several existing approaches. The primary strength of the model lies in its effective integration of shallow and dense CNN features, which improves grading performance. However, its ability to further differentiate high-risk PDR cases remains limited, indicating the need for more advanced DL techniques for fine-grained severity classification.

Bogacsovics et al. [32] proposed a DR detection framework that combines hand-crafted features with neural network–extracted features to improve automated fundus-image classification. Using a retinal-image dataset, the method applies pre-processing techniques, including contrast-limited adaptive histogram equalization and illumination equalization, to enhance lesion-specific feature extraction. The results demonstrate superior predictive performance compared with conventional approaches that rely solely on neural-network features. The main strength of the study lies in the effective integration of hand-crafted and deep features, which improves classification accuracy. However, the approach is limited by its reliance on a single hidden-layer architecture and requires further investigation of more advanced architectures to optimize feature fusion and overall performance. A ViT-based DL framework is proposed in [16] for automated DR severity detection by capturing long-range image dependencies through TL. The model is trained on the fine-grained annotated DR (FGADR) dataset using image normalization, data splitting, balancing and augmentation techniques, including flipping and transposition. Experimental results demonstrate superior performance compared with state-of-the-art CNN models, achieving an F1-score of 0.825. The study's strengths include the introduction of a novel ViT architecture and the investigation of data-balancing strategies. However, its effectiveness largely depends on the availability of large, annotated datasets, which may limit its applicability in data-constrained settings. She et al. [18] proposed a generative adversarial network (GAN) model to translate color fundus (CF) images into venous and late-phase fundus fluorescein angiography (FFA) images for enhanced DR screening. Pre-processing includes pixel-level image matching and filtering out poorly registered pairs, while augmentation involves random horizontal flips and rotations during training. The model is trained on 51,370 CF–venous FFA pairs and 14,644 CF–late FFA pairs, collected from 4,438 patients and is externally evaluated on the EyePACS and MESSIDOR2 datasets. Results demonstrate improved classification performance, with area under the receiver operating characteristic curve (AUC) increasing from 0.912 to 0.939 on EyePACS and from 0.952 to 0.972 on MESSIDOR2, while diabetic macular edema (DME) detection AUC improves from 0.927 to 0.974. The framework enhances DR screening by providing FFA-like information when conventional FFA is unavailable. However, the generated images may be less accurate in detecting subtle lesions and further prospective clinical validation is required.

Mahesh Patil et al. [15] proposed a DL-based DR detection framework using a ResNet-50 architecture, trained on the APTOS dataset and evaluated on the EyePACS dataset. The method incorporates image pre-processing, data augmentation and test-time augmentation (TTA) to improve classification performance and address data imbalance. Experimental results achieve a multi-class accuracy of 97.87% and a quadratic weighted kappa (QWK) score of 0.985, demonstrating strong generalization across datasets. The main strengths of the approach include its high classification accuracy, effective use of

augmentation strategies and cross-dataset evaluation capability. However, its reliance on specific benchmark datasets may limit generalizability to other DR image distributions and real-world clinical settings. A blended DL framework for DR grading is proposed in [11], which combines predictions from multiple DL models and employs 5-fold cross-validation to improve robustness and reduce overfitting. The model is trained on 33,310 fundus images from the EyePACS, APTOS, Messidor-2 and IDRiD datasets, with contrast-limited adaptive histogram equalization (CLAHE) and data augmentation applied to enhance contrast and balance the class distribution. The blended model achieves superior performance compared with individual DL architectures, demonstrating improved predictive accuracy and reliability. Its main strength lies in reducing classification bias through ensemble learning. However, its effectiveness remains dependent on dataset quality and further refinement is required to address ungradable images and misclassified samples. Athira and Nair [33] proposed a DL-based framework for automated DR detection and classification into No DR, NPDR and PDR stages using a sub-set of 3,386 fundus images from the APTOS dataset. The methodology incorporates image resizing, grayscale conversion, Gaussian blurring, circular cropping and optimized parameter tuning, followed by classification using a ResNet50-based network. The model achieves 94.7% classification accuracy and 99.8% detection accuracy, outperforming several comparative approaches. The main strengths of the framework include effective pre-processing, optimized training and enhanced detection of subtle retinal features. However, its performance may be affected by variations in image quality across the dataset and by challenges in distinguishing visually overlapping DR classes.

A DL multi-label feature extraction and classification (ML-FEC) model [34] employs pre-trained CNN architectures, including ResNet50, ResNet152 and SqueezeNet1, for multi-label lesion detection and classification in color fundus photographs (CFPs) to support early DR diagnosis. The methodology incorporates principal component analysis (PCA)-based feature extraction and data augmentation through image rotation to improve generalization. Using a fundus-image dataset, the model achieves accuracies of 94.40%, 93.67% and 91.94% for ResNet152, ResNet50 and SqueezeNet1, respectively, with low hamming loss values, demonstrating its effectiveness for clinical-screening applications. A key strength of the proposed framework is its fully automated multi-label DR detection capability, which supports scalable screening applications. However, its effectiveness is limited by the scarcity of large publicly available datasets with comprehensive lesion annotations, emphasizing the need for validation on larger, more diverse datasets to improve generalizability and clinical reliability. A graph CNN (GCNN) integrated with a variational autoencoder (VAE) framework is proposed in [20] for DR severity classification by combining local appearance features with global retinal vessel topology. The method utilizes contrast enhancement and data augmentation to improve image quality and address class imbalance and is evaluated on the Kaggle and EyePACS datasets. Experimental results achieve 90.34% accuracy on EyePACS, outperforming existing approaches by 2.85% while maintaining high sensitivity and specificity. The main strength of the approach lies in the effective integration of GCNN and VAE for enhanced feature representation. However, its adaptability to real-world clinical settings remains limited and further optimization is required to improve deployment feasibility. Mahmood et al. [35] presented a hybrid DR classification framework that integrates morphological image processing with the Inception v3 DL model to classify fundus images into five DR stages: normal, mild NPDR, moderate NPDR, severe NPDR and PDR. The approach is evaluated on the Messidor and Kaggle datasets and incorporates pre-processing steps, including blood-vessel segmentation, optic disc and macula localization and microaneurysm (MA) and hemorrhage (HEM) detection. The model achieves an overall accuracy of 96.83% with high sensitivity and specificity, demonstrating the effectiveness of combining image processing and DL techniques for automated DR diagnosis. A key strength of the framework is its ability to perform DR classification without expert supervision, while its main limitation is its dependence on image quality, which may affect diagnostic performance.

Raiaan et al. [7] proposed RetNet-10, a lightweight DL model based on a shallow CNN architecture for five-class DR classification. The model is trained on a combined dataset of 5,819 fundus images from APTOS, Messidor2 and IDRiD, with pre-processing techniques for noise and artifact removal and data augmentation using geometric, photometric and elastic deformation methods. RetNet-10 achieves a testing accuracy of 98.65%, outperforming several state-of-the-art models while maintaining low computational complexity. The main strength of the model lies in its ability to balance high classification performance with efficient processing. However, its performance may be affected by dataset imbalance and variations in image quality across the merged datasets. Mishra et al. [36] proposed a cascaded voting

ensemble deep neural network (DNN) for DR-severity classification. The method utilizes multi-level Otsu thresholding for fundus-image segmentation and combines the gray-level co-occurrence matrix (GLCM) feature extraction with dynamic flamingo optimization. Evaluated on fundus-image datasets, the model achieves high sensitivity, specificity, accuracy, precision and F1-score, demonstrating effective DR grading. Its main strength is the integration of multiple learning strategies within a robust classification framework aligned with standard grading systems, while its limitation is the increased computational complexity of the ensemble-based architecture. The fundus gradability-recognition network (FGR-Net) is proposed in [37] for automated assessment of retinal-image quality in computer-aided design (CAD) systems. The framework combines a deep autoencoder for self-supervised feature extraction and image reconstruction with a classifier to distinguish gradable from ungradable fundus images. Evaluated on retinal-image datasets, FGR-Net achieves over 89% accuracy and 87% F1-score while providing interpretable visual feedback for ophthalmologists. Its main strength lies in enhancing the interpretability of fundus-image quality assessment; however, its performance in real-time clinical settings and across diverse datasets requires further investigation.

Nazir et al. [38] proposed a weighted-average ensemble model that combines Inception-v3, VGG16 and a custom CNN for early DR detection, using the APTOS 2019 dataset (3662 training and 1992 test images). The approach incorporates image resizing and data-augmentation techniques, including rotation, zooming and flipping, to improve data diversity and reduce overfitting. The model achieves an accuracy of 95.06%, a precision of 87.88%, a recall of 83.78% and an F1-score of 85.69%. Its primary strength lies in integrating multiple DL models to develop an effective automated DR-screening system that can support clinical decision-making. However, its evaluation is limited by the use of a competition dataset without test labels, thereby restricting comprehensive performance validation. Xu et al. [25] proposed RT2Net, a multi-view joint learning framework for DR diagnosis that integrates global fundus-image features with local vascular features to improve classification performance. The methodology employs CLAHE and vascular segmentation, followed by feature extraction through dual-branch networks. Evaluated on the EyePACS and APTOS 2019 datasets, RT2Net achieves accuracies of 88.2% and 85.4%, respectively, with corresponding AUC values of 0.98 and 0.96. The model's main strength lies in its ability to leverage complementary global and local information to enhance diagnostic accuracy. However, its reliance on both feature representations may still result in misclassification in complex cases and increase model complexity. Liu [39] proposed a cascaded hybrid attention network (CHAN), which employs residual hybrid attention modules (RHAM) to extract features by combining channel attention and multi-head self-attention. The model is tested on the MESSIDOR dataset, which is partitioned into training, validation and test sets, with data augmentation techniques applied to expand the dataset to approximately 12,000 images. The training regimen includes 100 epochs with an Adam optimizer, achieving an accuracy rate of 96.34%, placing the method second among existing DR classification techniques. The main contributions include introducing the CHAN framework and effectively integrating attention mechanisms to enhance feature extraction. Limitations include the model's reliance on dataset quality and the potential challenges of optimizing attention mechanisms.

An attention-based DL framework for automated DR grading is proposed in [40], employing separate attention mechanisms for dark and bright retinal lesions using image-level labels only. The model is evaluated on the Kaggle DR dataset (35,126 training and 53,576 testing images) and achieves 83.7% accuracy and a QWK of 0.78, outperforming several existing methods. Its primary strength lies in generating interpretable attention maps that enhance clinical explainability. However, the image-decomposition process is computationally intensive and performance remains limited for certain severity classes, particularly classes 1 and 4. The DR Grading CNN (DRGCNN) [41] addresses DR grading challenges associated with class imbalance and the neglect of binocular relationships in fundus images. The model employs a CAM-EfficientNetV2-M encoder for feature extraction and a binocular feature-fusion network (BFFN) to integrate information from both eyes. Evaluation on the EyePACS (88,702 images) and Messidor-2 (1,748 images) datasets yields kappa scores of 86.62% and 86.16%, respectively, while achieving 89.06% when trained on EyePACS and tested on Messidor-2. The model enhances grading performance through balanced channel allocation and binocular feature fusion and improves interpretability using Grad-CAM. However, its performance remains affected by class imbalance and the quality of input fundus images. Naz et al. [42] proposed a robust fuzzy local information K-means clustering algorithm for DR detection that enhances conventional K-means by incorporating localized information into the clustering process. The method is evaluated on the IDRiD, Kaggle and Shiva Netralaya fundus-image datasets, with pre-processing and data-augmentation

techniques applied. Experimental results achieve 94.4% accuracy with an average execution time of 17.11 seconds and 97% accuracy under noisy conditions. The approach improves unsupervised DR detection and robustness to noise; however, its scalability to large datasets and dependence on random centroid initialization may affect efficiency and convergence.

Table 3. Comparative analysis of multi-class DR classification methods.

Ref.	Pre-processing	Dataset	Methods	Strengths	Limitations
[31]	Resizing, normalization	EyePACS, Messidor, IDRiD	Self-supervised MIE	Multi-modal representation learning	Requires multi-modal data
[19]	Cropping, Gaussian blur	EyePACS, MESSIDOR	Modified ResNet50	Image enhancement improves grading	Sensitive to image quality
[21]	Patch generation	LFH, DIARETDB, STARE, Kaggle	MPDCNN	Multi-stage feature extraction	High computational cost
[32]	CLAHE	Kaggle, Messidor, IDRiD	CNN + handcrafted features	Combines clinical and DL features	Limited architecture exploration
[16]	Cropping, normalization	FGADR	ViT	Captures global dependencies	Requires large training data
[18]	Registration and filtering	Private, EyePACS, MESSIDOR2	GAN + CNN	Novel multi-modal image translation	Clinical validation needed
[15]	Gaussian filtering	APTOS, EyePACS	ResNet50 + TTA	Excellent cross-dataset performance	Dataset dependency
[11]	CLAHE	EyePACS, APTOS, IDRiD	Blended DL	Reduces classification bias	Moderate grading performance
[33]	Resizing, grayscale, blur	APTOS	ResNet50 Hybrid	Optimized pre-processing pipeline	Image-quality variability
[34]	Rescaling, CLAHE	Retinal benchmark dataset	PCA + CNN	Multi-label lesion analysis	Limited dataset size
[20]	Contrast enhancement	EyePACS	GCNN + VAE	Exploits vascular topology	Clinical adaptation needed
[35]	CLAHE, vessel extraction	DRIVE, Messidor, Kaggle	Morphology + InceptionV3	Lesion-aware classification	Dependent on image quality
[7]	Noise removal	APTOS, Messidor2, IDRiD	RetNet-10	Lightweight architecture	Dataset imbalance
[36]	CLAHE, Gaussian thresholding	Kaggle, IDRiD, DIARETDB	Ensemble DNN	Highest grading performance	Computational complexity
[37]	None reported	EyePACS, EyeQ	FGR-Net	Image-quality assessment	Limited clinical validation
[38]	Standardization, cropping	APTOS	Ensemble CNN	Effective ensemble learning	Competition dataset limitations
[25]	CLAHE, vessel extraction	EyePACS, APTOS	RT2Net	Multi-view learning	Complexity of dual branches
[39]	Cropping, resizing	MESSIDOR	CHAN	Hybrid attention learning	Attention-optimization challenges
[40]	Cropping, normalization	EyePACS	Attention-based Xception	Improved explainability	Lower grading accuracy
[41]	Not reported	EyePACS, Messidor-2	DRGCNN	Binocular feature fusion	Class imbalance
[42]	Noise reduction	IDRiD, Kaggle	Fuzzy K-means	Robust to noisy images	Scalability concerns
[43]	Resizing	EyePACS, APTOS	Inception-ResNet Hybrid	Dual-path feature extraction	Large model complexity
[44]	Resizing, normalization	APTOS, MESSIDOR-2	DenseNet121 + GNN	Incorporates clinical metadata	Requires patient information
[45]	ZCA whitening	Kaggle	QCNET	Quantum-inspired learning	Experimental stage

Henge et al. [43] developed a multi-decision Inception-ResNet-based hybrid model for five-class DR classification using sequential and non-sequential fundus images. The architecture employs a dual-path TL framework with 172 weighted layers, processing both color and grayscale images to enhance feature representation. Evaluated on the EyePACS and APTOS datasets, the model achieves 98.1% accuracy and a 98.12% F1-score. The integration of dual-image feature mapping and chi-square refinement improves classification performance and robustness; however, the complex architecture may increase computational cost and limit deployment in resource-constrained environments. Zedadra et al. [44] proposed DRdiag, a multi-modal framework that combines a DenseNet121-based CNN with a two-layer GNN to integrate retinal-image features and patient-specific clinical information, particularly diabetes duration. Evaluated on the MESSIDOR-2 and APTOS2019 datasets, the model achieves classification accuracies of 97.6% and 98.0%, respectively. The framework demonstrates the effectiveness of graph-aware multi-modal learning for improving DR diagnosis by leveraging both visual and clinical data. However, its reliance on GNN-based modeling and patient metadata may increase system complexity and limit applicability when clinical information is unavailable. Ikram and Imran [45] proposed a hybrid quantum-classical neural network (QCNET) that combines DenseNet201 feature extraction with a 4-qubit variational quantum circuit for DR severity grading. Evaluated on the Kaggle DR dataset using

random oversampling and ZCA whitening, the model achieves 91.0% accuracy. The integration of gradient-weighted class-activation mapping (Grad-CAM) enhances interpretability by highlighting lesion-relevant retinal regions. While QCNET demonstrates the potential of quantum-enhanced learning and explainable AI, its practical applicability remains constrained by the early-stage development and scalability challenges of quantum computing technologies.

5.3 Binary and Multi-class DR Classification

This subsection discusses the literature on binary and multi-classification problems. Most research uses binary classification to distinguish between normal and DR cases, then classifies DR cases into 4 or 5 grades (Normal, Mild, Moderate, Severe and PDR). The studies reviewed in this sub-section are systematically summarized in Table 4, which provides a comparative overview of pre-processing techniques, datasets, methodologies, strengths and limitations reported in the literature.

Jaskari et al. [46] investigated uncertainty-aware DR classification using approximate Bayesian neural networks (BNNs). The study evaluates nine BNN models on EyePACS, KSSHP, Messidor-2 and APTOS datasets under a 5-class grading scheme and introduces a novel uncertainty measure linked to classifier risk. The results demonstrate that BNNs effectively estimate prediction uncertainty, while the proposed QWK-Risk metric improves performance on benchmark datasets. The study's main strength is the incorporation of uncertainty quantification into DR classification, whereas its primary limitation is reduced generalizability from benchmark datasets to real-world clinical data. Alahmadi [47] proposed a DR classification framework that integrates style and content recalibration with texture attention and spatial-normalization modules to enhance feature representation. The model is evaluated on the APTOS Kaggle dataset (3,662 retinal images) and achieves an accuracy of 0.85, demonstrating improved discrimination between healthy and DR samples. The main strength of the approach is the incorporation of attention mechanisms to enhance the learning of texture and semantic features. However, the model does not explicitly capture local contextual dependencies, which may limit its ability to model complex retinal structures. Islam et al. [48] proposed a supervised contrastive learning (SCL)-based two-stage framework that combines supervised contrastive loss for feature representation and cross-entropy loss for classification. Using the APTOS 2019 and Messidor-2 datasets with CLAHE-based pre-processing and data augmentation, the model achieves 98.36% accuracy and 98.50% AUC for binary classification and 84.36% accuracy for five-stage DR grading. The approach improves robustness and outperforms conventional CNN-based methods; however, its performance depends on effective augmentation and representation learning strategies.

A novel few-shot learning (FSL) framework called DRNet is introduced [22] using a prototypical network with episodic learning for DR detection and grading. Developed on the APTOS2019 dataset, the model achieves 99.73% accuracy, 99.82% sensitivity and 99.63% specificity for DR detection and 98.18% accuracy, 97.41% sensitivity and 99.55% specificity for DR grading. The framework demonstrates superior performance with limited training data; however, its effectiveness decreases as the number of training episodes is reduced. Tian et al. [49] proposed the fine-grained attention and knowledge-based collaborative network (FA+KC-Net) for DR grading. The method combines FA-Net to capture subtle lesion features and KC-Net to incorporate lesion-related medical knowledge. Evaluated on DDR, Messidor, APTOS and EyePACS datasets, FA+KC-Net achieves superior performance on DDR, Messidor and APTOS, demonstrating high accuracy and stability. Its main strength lies in integrating fine-grained visual features with domain knowledge, while a limitation is KC-Net's relatively lower performance, particularly in identifying DR 4 cases. Wong et al. [8] proposed a TL-based DR grading framework that combines pre-trained ShuffleNet and ResNet-18 models for feature extraction with an error correction output code (ECOC) ensemble classifier. Using the APTOS dataset, along with a combination of the EyePACS and Messidor-2 datasets and CLAHE pre-processing, the method achieves accuracies of 82% for 5-class grading and 96% for binary classification on the APTOS dataset. The approach improves performance through adaptive differential evolution (ADE)-based optimization of feature weights and network parameters. However, its effectiveness depends on image quality and may be susceptible to overfitting.

Farag et al. [3] proposed a DL-based DR severity-grading framework that uses a modified DenseNet169 with a convolutional block attention module (CBAM) and is trained on the APTOS dataset (~13,000 fundus images). Following Gaussian-filter pre-processing, the model achieves 97% accuracy in binary classification and strong multi-class grading performance, particularly for minority classes. The

approach offers efficient DR severity assessment with reduced computational complexity; however, its performance is still affected by dataset imbalance, motivating future investigation into alternative CBAM configurations and imbalanced learning strategies. Mohammadi and Nguyen [50] evaluated ChatGPT and Vertex AI for automated DR diagnosis using automated ML (AutoML) on the Messidor-2 dataset (1,740 fundus images). Image pre-processing is performed using CLAHE assisted by ChatGPT. The proposed models achieve an area under the precision-recall curve (AUPRC) of 0.81 for DR severity detection and, for DR presence detection, precision, recall and AUPRC of 84.48%, 84.48% and 0.90, respectively. The study demonstrates the potential of AI-assisted tools to support image pre-processing and analysis without coding expertise. However, its limitations include reliance on a single dataset and uncertain generalizability across diverse populations. Phridviraj et al. [51] proposed a bi-directional long short-term memory (Bi-LSTM) model combined with the multi-scale retinex with chromaticity preservation (MSRCP) pre-processing to improve DR detection from low-quality retinal-fundus images. The Bi-LSTM-MSRCP method is evaluated on the IDRiD, MESSIDOR and Kaggle datasets, achieving an accuracy of 92.35% and an F-score of 90.56%. Its primary strength is the integration of image enhancement and DL-based classification to improve detection performance, while its main limitation is its reliance on optimal tuning of MSRCP parameters.

Ashraf and Alghamdi [17] proposed a hierarchical feature-fusion network (HFF-Net), a hybrid CNN-based framework for DR screening and grading that employs hierarchical feature fusion to extract multi-scale retinal features. The model utilizes Swish activation, sparse categorical cross-entropy loss, CLAHE-based image enhancement, region-of-interest (ROI) extraction and data augmentation. Evaluated on fundus-image datasets, HFF-Net achieves 98.70% screening accuracy and 73.77% grading accuracy, outperforming competing methods. Its main strength lies in its effective multi-scale feature representation for screening and grading; however, the relatively low grading accuracy indicates challenges in distinguishing among DR severity levels. Yang et al. [52] proposed a transformer-based multiple-instance learning (TMIL) framework for DR classification, using ViT and a global instance computing block (GICB) to process high-resolution retinal images. The model is evaluated on the APTOS dataset (3,662 labeled images) for five-class grading and on the Messidor-1 dataset for binary classification. TMIL improves classification accuracy and reduces inference time by 62% compared with direct processing of high-resolution ViT. Its main strengths include efficient handling of high-resolution images and enhanced sensitivity for certain DR grades. However, the model still exhibits a false-negative rate exceeding 30%, indicating limited sensitivity in detecting early-stage DR.

Table 4. Comparative analysis of binary and multi-class DR detection methods.

Ref.	Pre-processing	Dataset	Methods	Strengths	Limitations
[46]	Resizing, variability reduction	EyePACS, KSSHP, Messidor-2, APTOS	BNNs, VGG16	Uncertainty estimation	Clinical-transferability issues
[47]	Cropping, normalization	APTOS	Texture Attention Network	Attention-guided learning	Limited contextual modeling
[48]	CLAHE	APTOS, Messidor-2	SCL + Xception	Contrastive representation learning	More complex training
[22]	Resizing	APTOS2019	DRNet (Few-Shot Learning)	Excellent performance with limited data	Sensitive to training episodes
[49]	Contrast enhancement	DDR, Messidor, APTOS, EyePACS	FA+KC-Net	Integrates medical knowledge	Reduced performance in severe grades
[8]	CLAHE	APTOS, EyePACS, Messidor-2	ECOC Ensemble	Ensemble optimization	Potential overfitting
[3]	Cropping, Gaussian filtering	APTOS2019	DenseNet169 + CBAM	Attention-enhanced classification	Dataset imbalance
[50]	CLAHE	Messidor-2	ChatGPT + Vertex AI AutoML	Low-code clinical workflow	Single-dataset evaluation
[51]	MSRCP enhancement	IDRiD, MESSIDOR, Kaggle	Bi-LSTM	Handles low-quality images	Requires parameter tuning
[17]	CLAHE, ROI extraction	IDRiD, APTOS2019	HFF-Net	Effective screening and grading	Lower grading accuracy
[52]	Cropping, patch creation	APTOS, Messidor-1	TMIL + ViT	Efficient high-resolution analysis	High false-negative rate

5.4 Lesion-based DR Classification

Some studies focus on classifying lesions in fundus images. Lesion-based classification in DR involves

categorizing retinal images by specific lesion features, such as microaneurysms, hemorrhagic spots and exudates. Table 5 presents a comprehensive summary of representative lesion-based classification studies, highlighting their pre-processing techniques, datasets, methodologies, strengths and limitations.

Lam et al. [53] proposed a DL-based retinal lesion-detection framework using CNNs trained on 1,324 ophthalmologist-labeled image patches from Kaggle retinal images. Using a sliding-window approach, the model generates lesion-probability maps and achieves AUCs of 0.94 for microaneurysms and 0.95 for exudates. The method demonstrates effective lesion detection with limited training data and supports automated retinal screening. However, it suffers from a relatively high false-positive rate due to training-data bias and requires considerable computational time for image processing. Khojasteh et al. [54] proposed a CNN-based framework for simultaneous detection of exudates, hemorrhages and microaneurysms using probability maps generated from the softmax layer. The method is evaluated on the DIARETDB1 and e-Ophtha datasets (284 fundus images) with pre-processing and data-augmentation techniques. The model achieves sensitivities of 0.96, 0.84 and 0.85 for exudates, hemorrhages and microaneurysms, respectively and accuracies of 97.3% and 86.6% on DIARETDB1 and e-Ophtha. The approach demonstrates robust DR lesion detection performance; however, it struggles to distinguish overlapping hemorrhages and microaneurysms and is affected by dataset imbalance.

Tang et al. [1] proposed a TL-based framework for neovascularization detection in PDR using AlexNet, GoogLeNet, ResNet18, ResNet50 and a hybrid ResNet18–GoogLeNet model. The method is evaluated on 1,174 retinal-image patches that have been enhanced with CLAHE. The proposed network achieves 91.57% accuracy, 85.69% sensitivity, 97.44% specificity and 97.10% precision, outperforming individual CNN models. Its main strength lies in the effective use of TL for detecting early neovascularization; however, it does not support DR severity grading and depends on the quality of ground-truth annotations. A novel DL-based method, termed HemNet [55], is proposed for automated hemorrhage detection using DIARETDB1 (89 images) and DIARETDB0 (130 images). The method employs gamma correction and a Gaussian matched filter for pre-processing and hemorrhage localization. HemNet achieves 94.21% sensitivity and 94.76% specificity, demonstrating competitive performance while reducing training time compared with deeper networks, such as AlexNet and VGG-16. Its main strength lies in its efficiency and high detection accuracy; however, its reliance on seed-point extraction may lead to missed hemorrhages when seed points are not accurately identified. A DL-based TL framework for DR detection is proposed in [56] using a modified VGG16 model for feature extraction and a logistic regression classifier. The approach is evaluated on an Indian DR dataset containing four lesion classes: microaneurysms, soft exudates, hard exudates and hemorrhages. Image pre-processing and data-augmentation techniques are employed to improve performance, achieving an accuracy of 90.4%. The main strength of the method is its reliable screening capability through CNN-based feature learning, while its limitations include insufficient noise reduction, lack of metadata integration and limited feature-selection strategies.

Table 5. Comparison of lesion-based approaches for automated DR detection.

Ref.	Pre-processing	Dataset	Methods	Strengths	Limitations
[53]	Normalization	Kaggle, E-Ophtha	CNN	Excellent lesion detection	Computationally intensive
[54]	Contrast enhancement	DIARETDB1, E-Ophtha	CNN	Multi-lesion detection	Difficulty with overlapping lesions
[1]	CLAHE	Messidor, DIARETDB0	TL	Detects neovascularization	No grading capability
[55]	Gamma correction	DIARETDB0/1	HemNet	Lightweight model	Seed-point dependency
[56]	Scaling, color correction	IDRiD	VGG16 / VGG19 TL	Strong lesion representation	Needs better feature selection

6. BENCHMARK DATASETS FOR DR DETECTION

This section summarizes key datasets widely used in DR research, particularly for classification tasks, such as severity grading and lesion identification. These datasets consist of retinal-fundus images captured under varying conditions, with differences in resolution, field of view (FOV) and annotation detail and include labels for pathologies, such as microaneurysms, hemorrhages, exudates and DME. Table 6 provides a detailed comparison of various DR datasets commonly used in research. Commonly used datasets include Messidor [57] (1,200 images with DR and DME grading), DIARETDB0/1 (130 and 89 images with expert lesion annotations) and Kaggle DR (EyePACS) [58], a large-scale dataset of 35,126 images with five-class DR grading, though affected by variations in image quality. Messidor-2

[57][59] extends Messidor with 1,748 images from 874 examinations, but lacks standardized DR annotations.

Other datasets provide more specialized or detailed annotations. E-ophtha [60] focuses on lesion-level labeling for microaneurysms and exudates, while DRiDB [61] offers detailed annotations of lesions and anatomical structures from multiple experts. IDRiD [62] includes 516 high-resolution images with both DR grading and pixel-level lesion annotations, supporting multi-task learning. APTOS 2019 [63] contains 3,662 images labeled according to ICDRSS severity levels under diverse real-world conditions. Larger and more recent datasets include DDR [64] (13,673 images from multiple hospitals, including poor-quality cases) and FGADR [65] (2,842 images with fine-grained lesion annotations and high-quality grading labels from multiple ophthalmologists), both of which enable more robust evaluation of DR classification models.

Table 6. Publicly available dataset for DR detection.

Dataset Name	**Images**	**No DR**	**Mild DR**	**Moderate DR**	**Severe DR**	**PDR**	**Fundus Camera**	**Annotation Level**	**Usage**
Messidor [57]	1200	546	153	247	254	-	Topcon TRC NW6 non-mydriatic, 45° FOV	Image-level	DR grading
DIARETDB0/ DIARETDB1 [66]	130/89	20 / 5	110/84				Zeiss FF450 + 50° FOV	Pixel-level and image-level	DR grading, Lesion Detection
Kaggle DR (EyePACS) [58]	88702 (35126 train)	73726 25810	7351 2443	7025 5292	708 873	1892 708	Various cameras (Centervue DRS, Optovue iCam, Canon CR1 / DGi / CR2, Topcon NW)	Image-level	DR grading
Messidor-2 [57]	1748	874	174	524	87	89	Topcon TRC NW6 non-mydriatic, 45° FOV	Image-level	DR grading
E-Ophtha [60]	463	-	-	-	-	-	Topcon TRC-NW6S, Canon CR-DGi	Pixel-level	DR diagnosis, Lesion Detection
DRiDB [61]	50	14	36				Zeiss VISUCAM 200, 45° FOV	Pixel-level and image-level	DR diagnosis, Lesion Detection
IDRiD [62]	516	134	20	193	65	104	Kowa VX-10 Alpha, 50° FOV	Pixel-level and image-level	DR grading, Lesion detection
APTOS 2019 [63]	5593	1805	370	999	193	295	Various fundus cameras	Image-level	DR grading
DDR [64]	13673	6266	630	4477	236	913	Various fundus cameras	Image-level	DR grading, Lesion Detection
FGADR [65]	2842	143	125	566	105	61	Not specified	Pixel-level and image-level	DR grading, Lesion Detection

Although these benchmark datasets have significantly advanced AI-based DR research, several important limitations remain regarding dataset diversity, annotation consistency and real-world generalization. Despite the availability of large benchmark datasets, significant dataset bias remains a major challenge in AI-driven DR detection. Public datasets often differ substantially in their demographic distributions, ethnicities, image acquisition devices, illumination conditions and annotation protocols. Such variability can undermine cross-dataset generalization and degrade performance when models are deployed in real-world clinical environments. Furthermore, annotation inconsistency among ophthalmologists introduces label uncertainty, particularly for intermediate DR stages. These limitations highlight the need for standardized evaluation protocols, diverse multi-center datasets and fairness-aware learning strategies.

7. DISCUSSION AND RESEARCH GAPS

The literature reviewed in this paper clearly demonstrates that AI–driven solutions for DR detection have undergone rapid methodological evolution, moving from coarse binary screening systems to sophisticated multi-task frameworks that integrate classification, lesion detection, segmentation and clinical interpretability. This progression reflects a broader shift in medical-image analysis from purely performance-driven models toward systems that aim to approximate clinical reasoning, scalability and trustworthiness.

Binary DR classification remains the most widely adopted paradigm, largely due to its direct applicability in population-level screening programs. Studies focusing on DR *versus* non-DR or referable *versus* non-referable DR consistently report high accuracy, sensitivity and specificity, as

demonstrated by GATL [27], HDeep [28] and ensemble CNN approaches [30]. These models effectively reduce ophthalmologists' workload by filtering out normal cases, thereby enabling targeted referral. However, a closer examination reveals that these impressive metrics are often achieved under controlled experimental settings with curated datasets, such as EyePACS and APTOS. While domain-adaptation strategies, such as GATL [27], attempt to mitigate inter-dataset variability, most binary classifiers still struggle when exposed to heterogeneous clinical environments characterized by diverse imaging devices, inconsistent illumination and variable patient demographics. Moreover, binary classification inherently oversimplifies DR pathology, failing to capture disease-progression dynamics that are essential for personalized treatment planning.

Multi-class DR classification and grading represent a more clinically meaningful task, as they directly correspond to treatment decisions and follow-up strategies. The reviewed studies highlight a clear trend toward deeper architectures, attention mechanisms and transformer-based models to address the subtle inter-class differences among DR stages. Self-supervised and multi-modal pre-training strategies, such as MIE-DR [31], demonstrate that leveraging unlabeled ophthalmic data significantly enhances feature representation, particularly when labeled datasets are limited. Similarly, ViT-based approaches [16] [52] successfully model long-range spatial dependencies, which are critical for capturing global retinal context. Attention-driven CNNs [40] [39] further improve discriminative power by focusing on lesion-relevant regions. Despite these advances, grading performance remains inconsistent across DR stages. Intermediate classes (mild and moderate NPDR) are frequently misclassified due to overlapping lesion characteristics and gradual disease transitions. This limitation is repeatedly observed even in high-performing systems, such as DRGCNN [41]and RT2Net [25]. The problem is exacerbated by severe class imbalance, where minority classes receive limited representation during training, leading to biased decision boundaries.

Lesion-based classification approaches address a critical limitation of pure image-level classifiers by explicitly modeling pathological features, such as microaneurysms, hemorrhages and exudates. Early CNN-based lesion detectors [53] [54] established the feasibility of automated lesion recognition with relatively small datasets, highlighting the potential for explainable DR screening systems. However, lesion-based methods often operate in isolation, focusing on individual lesion types without fully integrating their spatial distribution, co-occurrence patterns or clinical significance, while newer methods incorporate TL and hybrid architectures [56][55]. Lesion overlap and visual similarity—particularly between microaneurysms and hemorrhages—remain persistent challenges. Consequently, lesion-based classification alone is insufficient for reliable severity grading, reinforcing the need for integrated frameworks.

Another important consideration when interpreting the literature is the heterogeneity of the diagnostic tasks across studies. Although many works are collectively categorized under DR detection, they often target fundamentally different objectives, including binary DR screening, referable *versus* non-referable DR classification, multi-class severity grading, lesion detection, image-quality assessment and segmentation-assisted diagnosis. Consequently, performance metrics reported across these studies are not always directly comparable, because task complexity, annotation requirements, class distributions and evaluation protocols differ substantially. Binary screening tasks generally achieve higher accuracy due to simpler decision boundaries, whereas multi-class grading and lesion-level analysis present greater challenges related to inter-class similarity, annotation variability and clinical ambiguity. Therefore, reported performance should be interpreted in the context of the specific diagnostic objective and experimental setting, rather than as a universally comparable indicator of methodological superiority.

Despite these advances, several critical challenges persist. Most studies rely on a limited set of public datasets (EyePACS, APTOS, IDRiD, MESSIDOR), leading to weak cross-dataset generalization and performance degradation on unseen clinical data [15], [46]. Although many studies report high classification accuracy and AUC values, direct comparison among models remains difficult, because the evaluated methods use different datasets, pre-processing techniques, train-test splits, grading protocols and augmentation strategies. Consequently, reported performance metrics may not provide a fully fair or standardized comparison of state-of-the-art methods. Future research should focus on unified benchmarking frameworks and standardized evaluation protocols to enable more reliable comparative analysis. Reproducibility remains a major challenge in AI-based DR research. Many studies do not provide publicly available source codes, detailed pre-processing pipelines or standardized train-validation-test splits, making independent validation difficult. Furthermore, inconsistent

evaluation metrics and dataset protocols complicate fair comparison among models. Establishing open benchmarking frameworks and reproducible evaluation standards is essential for improving transparency, facilitating independent verification and accelerating clinical translation. Class imbalance continues to hinder reliable detection of early and intermediate DR stages, increasing the risk of false negatives. Reported high-performance metrics should be interpreted cautiously, as several studies rely on highly imbalanced datasets, limited external validation or extensive augmentation strategies that may artificially inflate performance. Furthermore, differences in experimental protocols and evaluation settings reduce reproducibility and complicate fair comparison across studies. These issues highlight the importance of robust validation using diverse multi-center datasets and standardized testing frameworks. Although explainability techniques, such as attention maps and Grad-CAM, improve interpretability, many models still lack clinically meaningful explanations aligned with ophthalmological reasoning. Furthermore, high computational complexity limits deployment in resource-constrained environments, despite the emergence of lightweight models, such as RetNet-10 and ADD-Net. Finally, most approaches rely solely on fundus images, with limited use of multi-modal and longitudinal data. Addressing these challenges is essential to bridge the gap between experimental performance and real-world clinical deployment of AI-based DR classification systems.

Although many AI-based DR detection systems achieve high performance on benchmark datasets, their translation into routine clinical practice remains limited. Most studies rely on retrospective evaluations using public datasets and lack prospective clinical validation across diverse patient populations and healthcare settings. In addition, substantial variation exists in validation protocols across studies, including differences in train-validation-test splits, cross-validation strategies, external validation practices and performance-reporting standards, making reliable comparison and clinical assessment difficult. Furthermore, regulatory approval of AI-based medical devices requires robust external validation, reliability assessment, explainability and evidence of clinical benefit. Practical deployment also faces challenges related to integration with existing healthcare workflows, interoperability with electronic health record systems, data privacy and security requirements and clinician trust in automated decision-support systems. Consequently, future research should prioritize clinically oriented validation studies, standardized regulatory-compliant evaluation protocols and human-centered AI frameworks to facilitate the safe and effective adoption of AI-driven DR screening systems in real-world clinical environments.

Recent research has also explored emerging paradigms beyond conventional CNN- and transformer-based architectures. Emerging approaches, such as multi-modal learning and quantum-inspired neural networks, demonstrate promising capability for improving DR classification performance. Multi-modal systems can integrate complementary information from fundus images, OCT scans and clinical metadata, thereby enhancing robustness and diagnostic precision. However, these methods typically require larger datasets, increased computational resources and more complex training pipelines. Similarly, quantum-inspired models remain in early experimental stages and currently face scalability and hardware limitations that restrict practical clinical deployment. Despite these challenges, such approaches represent promising directions for future research toward more comprehensive and clinically reliable DR detection systems.

8. Conclusion and Future Directions

This review examines AI-based approaches for DR classification, including binary detection and multi-class severity grading. Recent advances in deep learning, particularly CNNs, attention mechanisms and transformer-based models, have achieved high classification performance, in some cases approaching expert-level accuracy. However, key challenges remain. Many models suffer from limited generalization, as they rely heavily on a few benchmark datasets, leading to performance drops in real-world settings. Class imbalance further affects the reliable detection of early and intermediate DR stages. Additionally, most classification models rely on global image features and integrate only limited lesion-level information, while current interpretability methods often fall short of clinical expectations. From a practical perspective, computational complexity limits deployment, especially for transformer-based and ensemble models and achieving efficiency–accuracy balance remains challenging. Moreover, most approaches rely on single-modality fundus images, restricting their ability to capture disease variability. In conclusion, despite strong progress, bridging the gap between high experimental accuracy and real-

world clinical deployment remains a key challenge for DR classification systems.

Future research should focus on developing generalized, clinically reliable DR detection systems that use domain adaptation and federated learning strategies to preserve data privacy while improving cross-center generalization. Self-supervised and foundation-model approaches may reduce dependence on large, labeled datasets, while multi-modal learning can integrate fundus imaging with OCT and clinical metadata to improve diagnostic robustness. In addition, lightweight edge AI architectures are needed for deployment in resource-constrained screening environments. Standardized benchmarking frameworks and fairness-aware evaluation protocols should also be established to improve reproducibility and clinical reliability.

REFERENCES

[1] M. C. S. Tang, S. S. Teoh, H. Ibrahim and Z. Embong, "A Deep Learning Approach for the Detection of Neovascularization in Fundus Images Using Transfer Learning," IEEE Access, vol. 10, pp. 20247–20258, DOI: 10.1109/ACCESS.2022.3151644, 2022.

[2] Md. Nur-A-Alam, Md. M. K. Nasir, M. Ahsan, Md. A. Based, J. Haider and S. Palani, "A Faster RCNN-based Diabetic Retinopathy Detection Method Using Fused Features From Retina Images," IEEE Access, vol. 11, pp. 124331–124349, DOI: 10.1109/ACCESS.2023.3330104, 2023.

[3] 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, DOI: 10.1109/ACCESS.2022.3165193, 2022.

[4] G. Ali, A. Dastgir, M. W. Iqbal, M. Anwar and M. Faheem, "A Hybrid Convolutional Neural Network Model for Automatic Diabetic Retinopathy Classification From Fundus Images," IEEE J. of Translational Engineering in Health and Medicine, vol. 11, pp. 341–350, DOI: 10.1109/JTEHM.2023.3282104, 2023.

[5] M. Shafiq, Q. Fan, F. H. Alghamedy and W. J. Obidallah, "DualEye-FeatureNet: A Dual-stream Feature Transfer Framework for Multi-modal Ophthalmic Image Classification," IEEE Access, vol. 12, pp. 143985–144008, DOI: 10.1109/ACCESS.2024.3469244, 2024.

[6] M. N. Hasan, M. E. R. Pial, S. Das, N. Siddique and H. Wang, "DIA-VXNET: A Framework for Automated Diabetic Eye Disease Detection Using Transfer Learning with Feature Fusion Network," Biomedical Signal Processing and Control, vol. 100, p. 106907, Feb. 2025.

[7] M. A. K. Raiaan et al., "A Lightweight Robust Deep Learning Model Gained High Accuracy in Classifying a Wide Range of Diabetic Retinopathy Images," IEEE Access, vol. 11, pp. 42361–42388, DOI: 10.1109/ACCESS.2023.3272228, 2023.

[8] W. K. Wong, F. H. Juwono and C. Apriono, "Diabetic Retinopathy Detection and Grading: A Transfer Learning Approach Using Simultaneous Parameter Optimization and Feature-weighted ECOC Ensemble," IEEE Access, vol. 11, pp. 83004–83016, DOI: 10.1109/ACCESS.2023.3301618, 2023.

[9] T. Palaniswamy and M. Vellingiri, "Internet of Things and Deep Learning Enabled Diabetic Retinopathy Diagnosis Using Retinal Fundus Images," IEEE Access, vol. 11, pp. 27590–27601, 2023.

[10] M. Nawal Sad-Houari Hicham Reguieg, Chaimaa Bachiri, "Automated Diabetes Disease Prediction System Based on Risk Factors Assessment: Taking Charge of Your Health," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 09, no. 04, pp. 377–394, DOI: 10.5455/jjcit.71-1694107771, 2023.

[11] F. C. Monteiro, "Diabetic Retinopathy Grading Using Blended Deep Learning," Procedia Computer Science, vol. 219, pp. 1097–1104, DOI: 10.1016/j.procs.2023.01.389, 2022.

[12] B. N. Jagadesh, M. G. Karthik, D. Siri, S. K. K. Shareef, S. V. Mantena and R. Vatambeti, "Segmentation Using the IC2T Model and Classification of Diabetic Retinopathy Using the Rock Hyrax Swarm-based Coordination Attention Mechanism," IEEE Access, vol. 11, pp. 124441–124458, DOI: 10.1109/ACCESS.2023.3330436, 2023.

[13] S. Samira Ait Kaci Azzou, D. Boukredera and S. Baouz, "A Hybrid CNN-Transformer Approach for Precise Three-class Diabetic Retinopathy Classification," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 11, no. 03, pp. 279–299, DOI: 10.5455/jjcit.71-1738691198, 2025.

[14] Q. Abbas, M. E. A. Ibrahim and A. R. Baig, "Transfer Learning-based Computer-aided Diagnosis System for Predicting Grades of Diabetic Retinopathy," Computers, Materials & Continua, vol. 71, no. 3, pp. 4573–4590, doi: 10.32604/cmc.2022.023670, Jan. 2022.

[15] M. S. Patil, S. Chickerur, C. Abhimalya, A. Naik, N. Kumari and S. Maurya, "Effective Deep Learning Data Augmentation Techniques for Diabetic Retinopathy Classification," Proc. of the Int. Conf. on Machine Learning and Data Engineering, vol. 218, pp. 1156–1165, Jan. 2023.

[16] W. Nazih, A. O. Aseeri, O. Y. Atallah and S. El-Sappagh, "Vision Transformer Model for Predicting the Severity of Diabetic Retinopathy in Fundus Photography-based Retina Images," IEEE Access, vol. 11, pp. 117546–117561, DOI: 10.1109/ACCESS.2023.3326528, 2023.

[17] M. H. Ashraf and H. Alghamdi, "HFF-Net: A Hybrid Convolutional Neural Network for Diabetic

Retinopathy Screening and Grading," Biomedical Technology, vol. 8, pp. 50–64, Dec. 2024.

[18] D. Shi et al., "Translation of Color Fundus Photography into Fluorescein Angiography Using Deep Learning for Enhanced Diabetic Retinopathy Screening," Ophthalmology Science, vol. 3, no. 4, p. 100401, Dec. 2023.

[19] S. H. Abbood, H. N. A. Hamed, M. S. M. Rahim, A. Rehman, T. Saba and S. A. Bahaj, "Hybrid Retinal Image Enhancement Algorithm for Diabetic Retinopathy Diagnostic Using Deep Learning Model," IEEE Access, vol. 10, pp. 73079–73086, DOI: 10.1109/ACCESS.2022.3189374, 2022.

[20] S. Sundar and S. Sumathy, "Classification of Diabetic Retinopathy Disease Levels by Extracting Topological Features Using Graph Neural Networks," IEEE Access, vol. 11, pp. 51435–51444, DOI: 10.1109/ACCESS.2023.3279393, 2023.

[21] V. Deepa, C. S. Kumar and T. Cherian, "Ensemble of Multi-stage Deep Convolutional Neural Networks for Automated Grading of Diabetic Retinopathy Using Image Patches," J. of King Saud Uni. - Computer and Information Sciences, vol. 34, no. 8, Part B, pp. 6255–6265, Sep. 2022.

[22] M. Murugappan, N. B. Prakash, R. Jeya, A. Mohanarathinam, G. R. Hemalakshmi and M. Mahmud, "A Novel Few-shot Classification Framework for Diabetic Retinopathy Detection and Grading," Measurement, vol. 200, p. 111485, DOI: 10.1016/j.measurement.2022.111485, Aug. 2022.

[23] J. Amin, I. Shazadi, M. Sharif, M. Yasmin, N. A. Almujally and Y. Nam, "Localization and Grading of NPDR Lesions Using ResNet-18-YOLOv8 Model and Informative Features Selection for DR Classification Based on Transfer Learning," Heliyon, vol. 10, no. 10, p. e30954, May 2024.

[24] K. Aurangzeb, R. S. Alharthi, S. I. Haider and M. Alhussein, "Systematic Development of AI-enabled Diagnostic Systems for Glaucoma and Diabetic Retinopathy," IEEE Access, vol. 11, pp. 105069–105081, DOI: 10.1109/ACCESS.2023.3317348, 2023.

[25] X. Xu, D. Liu, G. Huang, M. Wang, M. Lei and Y. Jia, "Computer Aided Diagnosis of Diabetic Retinopathy Based on Multi-view Joint Learning," Computers in Biology and Medicine, vol. 174, p. 108428, May 2024.

[26] 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, Sep. 2003.

[27] J. Hu, H. Wang, L. Wang and Y. Lu, "Graph Adversarial Transfer Learning for Diabetic Retinopathy Classification," IEEE Access, vol. 10, pp. 119071–119083, DOI: 10.1109/ACCESS.2022.3220776, 2022.

[28] T. F. de Sousa and C. G. Camilo, "HDeep: Hierarchical Deep Learning Combination for Detection of Diabetic Retinopathy," Proc. of the Int. Neural Network Society Workshop on Deep Learning Innovations and Applications (INNS DLIA 2023), vol. 222, pp. 425–434, Jan. 2023.

[29] P. Zhang et al., "Image Quality Assessment of Diabetic Retinopathy Based on ADD-Net," IEEE Access, vol. 11, pp. 105130–105139, DOI: 10.1109/ACCESS.2023.3318876, 2023.

[30] S. Pavithra, D. Jaladi and K. Tamilarasi, "Optical Imaging for Diabetic Retinopathy Diagnosis and Detection Using Ensemble Models," Photodiagnosis and Photodynamic Therapy, vol. 48, p. 104259, Aug. 2024.

[31] Á. S. Hervella, J. Rouco, J. Novo and M. Ortega, "Multimodal Image Encoding Pre-training for Diabetic Retinopathy Grading," Computers in Biology and Medicine, vol. 143, p. 105302, Apr. 2022.

[32] G. Bogacsovics, J. Toth, A. Hajdu and B. Harangi, "Enhancing CNNs through the Use of Hand-crafted Features in Automated Fundus Image Classification," Biomedical Signal Processing and Control, vol. 76, p. 103685, DOI: 10.1016/j.bspc.2022.103685, Jul. 2022.

[33] T. R. Athira and J. J. Nair, "Diabetic Retinopathy Grading From Color Fundus Images: An Autotuned Deep Learning Approach," Proc. of the Int. Conf. on Machine Learning and Data Engineering, vol. 218, pp. 1055–1066, DOI: 10.1016/j.procs.2023.01.085, Jan. 2023.

[34] T. M. Usman, Y. K. Saheed, D. Ignace and A. Nsang, "Diabetic Retinopathy Detection Using Principal Component Analysis Multi-label Feature Extraction and Classification," Int. J. of Cognitive Computing in Eng., vol. 4, pp. 78–88, DOI: 10.1016/j.ijcce.2023.02.002, Jun. 2023.

[35] M. A. I. Mahmood, N. Aktar and Md. F. Kader, "A Hybrid Approach for Diagnosing Diabetic Retinopathy from Fundus Image Exploiting Deep Features," Heliyon, vol. 9, no. 9, p. e19625, Sep. 2023.

[36] A. Mishra, M. Pandey and L. Singh, "Segmentation and Classification of Diabetic Retinopathy Using Ensemble Deep Neural Network," Proc. of the 5th Int. Conf. on Innovative Data Communication Technologies and Application (ICIDCA 2024), vol. 233, pp. 226–234, DOI: 10.1016/j.procs.2024.03.212, Jan. 2024.

[37] S. Khalid, H. A. Rashwan, S. Abdulwahab, M. Abdel-Nasser, F. M. Quiroga and D. Puig, "FGR-Net: Interpretable Fundus Image Gradeability Classification Based on Deep Reconstruction Learning," Expert Systems with Applications, vol. 238, p. 121644, DOI: 10.1016/j.eswa.2023.121644, Mar. 2024.

[38] K. Nazir, J. Kim and Y.-C. Byun, "Enhancing Early-stage Diabetic Retinopathy Detection Using a Weighted Ensemble of Deep Neural Networks," IEEE Access, vol. 12, pp. 113565–113579, 2024.

[39] Y. Liu, "Automatic Grading Diabetic Retinopathy in Color Fundus Image: Cascaded Hybrid Attention Network," Journal of Radiation Research and Applied Sciences, vol. 17, no. 4, p. 101160, Dec. 2024.

[40] R. Romero-Oraá, M. Herrero-Tudela, M. I. López, R. Hornero and M. García, "Attention-based Deep Learning Framework for Automatic Fundus Image Processing to Aid in Diabetic Retinopathy Grading," Computer Methods and Programs in Biomedicine, vol. 249, p. 108160, Jun. 2024.

[41] Z. Hai et al., "A Novel Approach for Intelligent Diagnosis and Grading of Diabetic Retinopathy," Computers in Biology and Medicine, vol. 172, p. 108246, Apr. 2024.

[42] H. Naz, T. Saba, F. S. Alamri, A. S. Almasoud and A. Rehman, "An Improved Robust Fuzzy Local Information K-means Clustering Algorithm for Diabetic Retinopathy Detection," IEEE Access, vol. 12, pp. 78611–78623, DOI: 10.1109/ACCESS.2024.3392032, 2024.

[43] S. K. Henge, N. R. Viraati, M. Alhussein, A. S. Kushwaha, K. Aurangzeb and R. Singh, "Detection of Diabetic Retinopathy Using a Multi-decision Inception-ResNet-blended Hybrid Model," IEEE Access, vol. 13, pp. 8988–9005, DOI: 10.1109/ACCESS.2024.3525154, 2025.

[44] A. Zedadra, O. Zedadra, M. Yassine Salah-Salah and A. Guerrieri, "Graph-aware Multimodal Deep Learning for Classification of Diabetic Retinopathy Images," IEEE Access, vol. 13, pp. 74799–74810, 2025.

[45] A. Ikram and A. Imran, "Fundus Imaging Based Quantum Computing: Grading of Severity of Vision Threatening Diabetic Retinopathy," Informatics in Medicine Unlocked, vol. 61, p. 101744, DOI: 10.1016/j.imu.2026.101744, Mar. 2026.

[46] J. Jaskari et al., "Uncertainty-aware Deep Learning Methods for Robust Diabetic Retinopathy Classification," IEEE Access, vol. 10, pp. 76669–76681, DOI: 10.1109/ACCESS.2022.3192024, 2022.

[47] M. D. Alahmadi, "Texture Attention Network for Diabetic Retinopathy Classification," IEEE Access, vol. 10, pp. 55522–55532, DOI: 10.1109/ACCESS.2022.3177651, 2022.

[48] M. R. Islam et al., "Applying Supervised Contrastive Learning for The Detection of Diabetic Retinopathy and Its Severity Levels from Fundus Images," Computers in Biology and Medicine, vol. 146, p. 105602, DOI: 10.1016/j.compbiomed.2022.105602, Jul. 2022.

[49] M. Tian, H. Wang, Y. Sun, S. Wu, Q. Tang and M. Zhang, "Fine-grained Attention & Knowledge-based Collaborative Network for Diabetic Retinopathy Grading," Heliyon, vol. 9, no. 7, p. e17217, Jul. 2023.

[50] S. S. Mohammadi and Q. D. Nguyen, "A User-friendly Approach for the Diagnosis of Diabetic Retinopathy Using ChatGPT and Automated Machine Learning," Ophthalmology Science, vol. 4, no. 4, p. 100495, 2024.

[51] M. S. B. Phridviraj, R. Bhukya, S. Madugula, A. Manjula, S. Vodithala and M. S. Waseem, "A Bi-directional Long Short-term Memory-based Diabetic Retinopathy Detection Model Using Retinal Fundus Images," Healthcare Analysis, vol. 3, p. 100174, DOI: 10.1016/j.health.2023.100174, Nov. 2023.

[52] Y. Yang, Z. Cai, S. Qiu and P. Xu, "A Novel Transformer Model with Multiple Instance Learning for Diabetic Retinopathy Classification," IEEE Access, vol. 12, pp. 6768–6776, 2024.

[53] C. Lam, C. Yu, L. Huang and D. Rubin, "Retinal Lesion Detection with Deep Learning Using Image Patches," Investigative Ophthalmology & Visual Science, vol. 59, no. 1, p. 590, Jan. 2018.

[54] P. Khojasteh, B. Aliahmad and D. K. Kumar, "Fundus Images Analysis Using Deep Features for Detection of Exudates, Hemorrhages and Microaneurysms," BMC Ophthalmology, vol. 18, no. 1, p. 288, Dec. 2018.

[55] T. Aziz, C. Charoenlarpnopparut and S. Mahapakulchai, "Deep Learning-based Hemorrhage Detection for Diabetic Retinopathy Screening," Scientific Reports, vol. 13, no. 1, p. 1479, Jan. 2023.

[56] D. Singh and D. C. Dobhal, "A Deep Learning-based Transfer Learning Approach Fine-tuned for Detecting Diabetic Retinopathy," Proc. of the 5th Int. Conf. on Innovative Data Communication Technologies and Application (ICIDCA 2024), vol. 233, pp. 444–453, DOI: 10.1016/j.procs.2024.03.234, Jan. 2024.

[57] E. Decencière et al., "Feedback on a Publicly Distributed Image Database: The Messidor Database," Image Analysis & Stereology, vol. 33, no. 3, Art. no. 3, DOI: 10.5566/ias.1155, Aug. 2014.

[58] Emma Dugas, Jared, Jorge and Will Cukierski, "Diabetic Retinopathy Detection," Diabetic Retinopathy Detection, [Online], Available: https://kaggle.com/diabetic-retinopathy-detection, Accessed: Dec. 14, 2024.

[59] G. P. Maffre Gervais Gauthier et al., "Messidor-2," ADCIS, [Online], Available: https://www.adcis.net/en/third-party/messidor2/, Accessed: Dec. 14, 2024.

[60] E. Decencière et al., "TeleOphta: Machine Learning and Image Processing Methods for Teleophthalmology," ANR TECSAN - Technologies for Health and Autonomy, vol. 34, no. 2, pp. 196–203, DOI: 10.1016/j.irbm.2013.01.010, Apr. 2013.

[61] P. Prentašić et al., "Diabetic Retinopathy Image Database (DRiDB): A New Database for Diabetic Retinopathy Screening Programs Research," Proc. of the 2013 8th Int. Symposium on Image and Signal Processing and Analysis (ISPA), pp. 711–716, DOI: 10.1109/ISPA.2013.6703830, 2013.

[62] S. P. Prasanna Porwal, "Indian Diabetic Retinopathy Image Dataset (IDRiD)," IEEE DataPort, DOI: 10.21227/H25W98, 2018.

[63] M. Karthik and S. Dane, "APTOS 2019 Blindness Detection," [Online], Available: https://kaggle.com/aptos2019-blindness-detection, Accessed: Dec. 14, 2024.

[64] 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, DOI: https://doi.org/10.1016/j.ins.2019.06.011, 2019.

[65] Y. Zhou, B. Wang, L. Huang, S. Cui and L. Shao, "A Benchmark for Studying Diabetic Retinopathy: Segmentation, Grading and Transferability," IEEE Transactions on Medical Imaging, vol. 40, no. 3, pp. 818–828, DOI: 10.1109/TMI.2020.3037771, Mar. 2021.

[66] T. Kauppi et al., "DIARETDB1 Diabetic Retinopathy Database and Evaluation Protocol," [Online], Available: https://webpages.tuni.fi/vision/public_data/publications/bmvc2007_diaretdb1.pdf, 2007.

ملخص البحث:

يُعدّ اعتلال الشّبكية السُّكّري سبباً رئيسياً لضعّف البصر الّذي يمكن الوقاية منه، ممّا يستدعي توفير حلول فحص موثوقة وقابلة للتّوسّع. وقد ساهم الذّكاء الاصطناعي – ولا سيما التّعلُّم الآلي والتّعلُّم العميق – بشكلٍ كبير في أتمتة تصنيف اعتلال الشّبكية السّكّري استناداً إلى صور قاع العين.

تقدّم هذه الورقة مراجعة منهجية لأساليب تصنيف اعتلال الشّبكية السّكّري القائمة على الذّكاء الاصطناعي، مع التّركيز على التّصنيف الثّنائي، وتصنيف درجات الشّدّة متعدّد الفئات، والطّرق القائمة على الآفات. كما تطرح تصنيفاً لهذه الأساليب بناءً على أهدافها السّريرية. وتغطّي المراجعة أحدث النّماذج المتطوّرة، بما في ذلك الشّبكات العصبية التّلافيفية، ومحوّلات الرّؤية، وآليات الانتباه، وتقنيات النّماذج المجمّعة، بالاضافة إلى استراتيجيات المعالجة المسبقة وبروتوكولات التّقييم الخاصّة بها. كما يتمّ تحليل مجموعات البيانات المرجعية الشّائعة من حيث توزيع الفئات وتأثيرها على الأداء.

وعلى الرّغم من أنّ النّماذج الحالية تحقّق دقّة عالية في مجموعات البيانات القياسية، فإنّ التّحدّيات لا تزال قائمةً فيما يتعلّق بالقُدرة على التّعميم، وعدم توازن الفئات، والموثوقية في الكشْف عن المرض في مراحله المبكّرة. وتجدر الإشارة إلى أنّ الأساليب القائمة على الآفات تعزّز قابلية التّفسير، إلّا أنّها غالباً لا تكون مُدمجةً بالكامل في أطر تصنيف درجات المرض.

وبشكلٍ عام، تسلّط هذه المراجعة الضّوء على القيود الرّئيسية وتحدّد التّوجُّهات المستقبلية نحو تطوير أنظمة قوية وقابلة للتّفسير وقابلة للتّطبيق السّريري لتصنيف اعتلال الشّبكية السّكّري في سياق الفحص الميداني الفعلي.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).