International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(7s) 2026 pp. 1060-1072
https://doi.org/10.70917/ijcisim-2026-3181
Article

Evolution and Future Directions of Deep Learning in Automated Diabetic Retinopathy Detection: A Comprehensive Review

Mohammad Maaz Khan[1], Mohd Farhan Haroon[2], Girish Chandra[1], Deepa Verma[1]

[1]Department of Computer Science and Engineering, Institute of Engineering and Technology, Lucknow, Uttar Pradesh, India
[2]Department of Electrical Engineering, Indian Institute of Technology, Kanpur, India, Email: mmkmaazkhan781@gmail.com

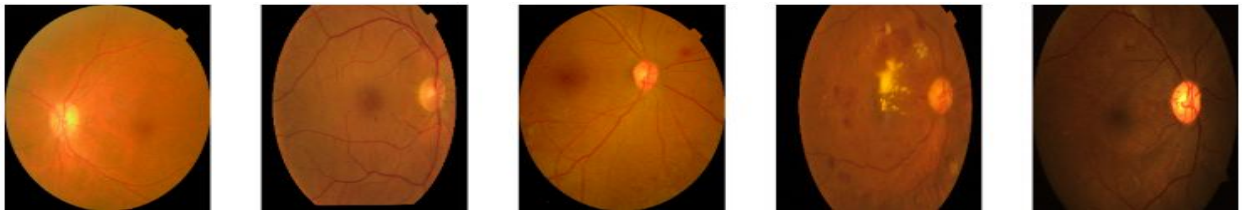
Abstract. Diabetic retinopathy (DR) remains one of the leading causes of preventable blindness globally, thus early and accurate detection is crucial. In the last decade, significant advances in computational imaging, deep learning and artificial intelligence have resulted in creation of several automated techniques to detect DR from retinal images. In this review, we summarize and compare thirty-one (31) representative studies between 2015 to 2025 from both classification-based and detection-based methods. The study proposes five main research questions to explore the transition from Convolutional Neural Networks (CNNs) towards hybrid transformer architectures, the impact of dataset diversity and the barriers towards clinical interpretability. The paper also studies the evolution of various model architectures, including convolutional neural networks (CNNs), transformer models and YOLO-based methodologies, by analyzing their performance across popular datasets like EyePACS, Messidor and IDRiD. Salient challenges, e.g., imbalanced datasets, varying image qualities, and the desired interpretability of the outputs are addressed in this review together with possible research paths moving forward. Comprehensively, this precise paper is aimed at offering a coherent, organised and extended literature resource to those researchers that are heading to automatic detection of diabetic retinopathy with better accuracies.

Keywords: Diabetic Retinopathy, Deep Learning, Fundus Images, Convolutional Neural Networks (CNN), YOLO, Lesion-Based Detection, Hybrid Models

1. INTRODUCTION

Diabetic retinopathy is a major cause of avoidable blindness and is basically one of the most common consequences of diabetes in the world. The fact that about one out of every three individuals with diabetes will ultimately experience this complication justifies the necessity of a dependable screening procedure becomes less of a technical objective and more of a communal health need. The field has, traditionally, been dependent on ophthalmologists to carefully scan the retinal fundus images to identify the smallest clues of the disease. But such a manual process is subjective and slow by nature. It also is closely connected to the existence of trained experts, which is lacking in many regions. It is this bottleneck which has given rise to the impetus behind the need to have computer-aided systems which can provide a faster, possibly more consistent, second opinion.

The pathophysiology of the disease is typified by the harmful impacts of chronic hyperglycemia on the retinal microvasculature. This injury is seen to appear as a particular collection of pathological indicators, such as microaneurysms, small haemorrhages, and that waxy exudates that are signs of leaking vessels. The absence of early indicators enables the disease to pass to more dangerous phases, starting with a mild non-proliferative phase and changing to the proliferative stage where its abnormal neovascularization may result in a permanent and irreversible loss of vision. The intensity of this progression, as Figure 1 shows, is, in actuality, the reason why the therapeutic intervention, at the earliest stage of the disease, is beneficial. The use of fundus photography has continued to be the preferred method of screening since it enables the visualization of these structures to a sufficient degree of clarity to make a decision, despite deficiencies in the area of lighting and image quality. Pathological characteristics such as haemorrhages and microaneurysms are the main visual indicators that both a doctor and an automated algorithm search when a diagnostics check is performed.



(a) (b) (c) (d) (e)

Figure 1. Clinical Progression of Diabetic Retinopathy (DR) Severity Levels: **(a)** No DR, **(b)** Mild NPDR, **(c)** Moderate NPDR, **(d)** Severe NPDR, And **(e)** Proliferative DR.

The increasing global awareness and research on this condition can be seen in search interest trends over the last ten years. Figure 2 indicates that the frequency of global search of "Diabetic Retinopathy" has consistently increased from 2015 to 2025, with maximum popularity at the end of this decade. The trend signifies a pressing need to be able to apply scalable diagnostic solutions, like machine learning, deep learning models that are discussed in this review.

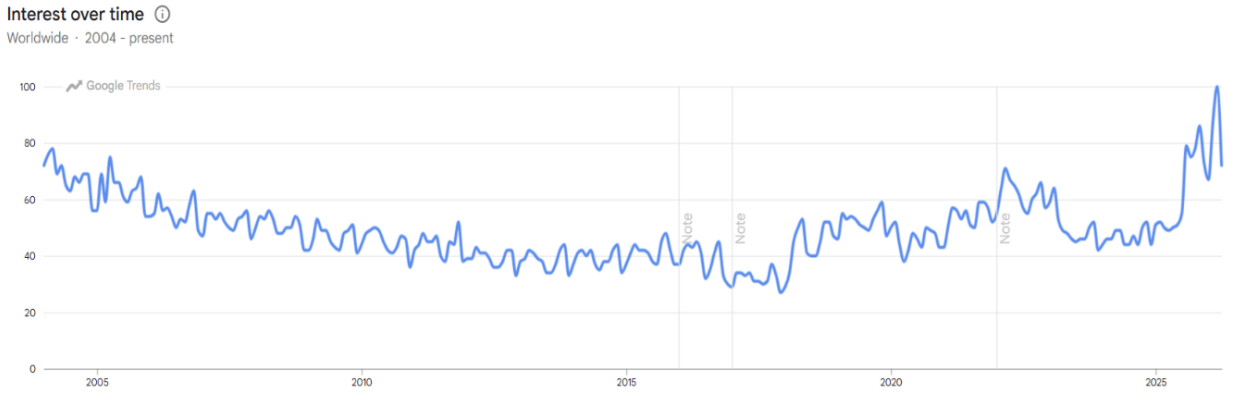


Figure 2. Google Trends Analysis (Jan 2015 – Dec 2025) Showing the Global Search Interest Trajectory for Diabetic Retinopathy.

The past decade has witnessed drastic changes in medical image analysis through deep learning [1],[2]. The industry has shifted from primary convolutional networks [3],[4] to much more ambitious transformer architectures [5] and complex hybrid pipelines [6],[7]. Even the YOLO architectures, originally developed to identify everyday objects in a short span, are being refined to detect tiny lesions in the retina with a greater precision [8],[9]. These advances have taken automated screening significantly closer to real-world implementation at community level [10],[11], which in practice seldom proceeds as smoothly as a research paper does. There are still various practical obstacles. For example, in the real-world, the quality of fundus photo can hardly be as clean as training sets due to factors such as camera hardware or uneven lighting often deteriorating the image sharpness [12],[13]. The human factor should also be taken into consideration. Minor variations in the eye anatomy or ethnicity may alter the manifestation of the lesion, and in fact, it is rather concerning that the difference in performance in the models between various demographics reaches 5-10 percent just because the training data was not diverse enough [11]. The data itself may be considered as the largest bottleneck perhaps. Manual labelling of all microaneurysms with the help of a skilled ophthalmologist is a tiresome and costly job [14],[15]. With the technical sophistication of the code, clinical implementations require regulatory approvals and stringent multi-center testing before they can know whether these systems can indeed withstand the unpredictability of a busy hospital.

In order to prove the validity of the evidence provided, thirty-one seminal studies between 2015 to 2025 were reviewed in this paper. These were extracted from high-impact publishers to ensure academic rigor. Most of the papers reviewed as shown in Figure 3 are of IEEE (35%), different Medical Journals (13%), Springer (13%), Elsevier (10%) and MDPI (10%) with 19% representing other sources. By comparing these architectures and performance results, the review outlines the situation of automated DR detection in terms of the current state and provides advice on future studies that should be conducted to have correct, interpretable, and clinically reliable solutions.

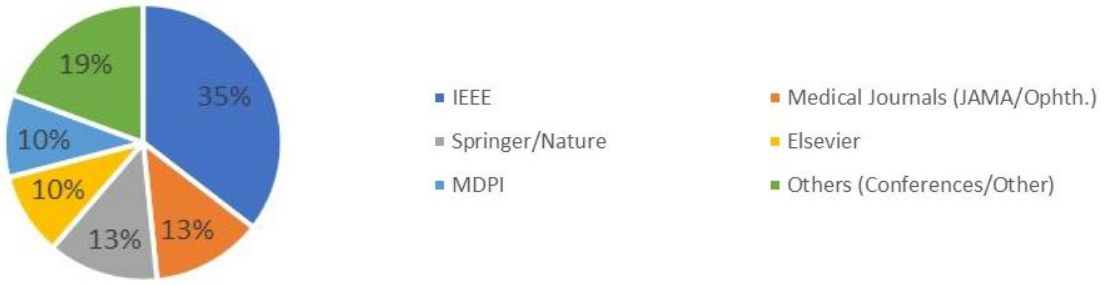


Figure 3. Distribution of the 31 Reviewed Papers Across Major Academic Publishers and Clinical Journals.

2. RESEARCH QUESTIONS AND SEARCH CRITERIA

A systematic method was adopted to identify and filter the existing literature to ensure high level of technical depth. The primary goal was to follow the rapid change in diagnostic tools since 2015 to the beginning of 2026. This section outlines the research questions that guided the analysis and the search strategy that was followed to select the thirty-one studies presented in this review.

2.1 Research Questions

A brief review of current literature is not enough to determine the current situation in terms of automated diabetic retinopathy detection. Thus, a series of specific research questions was formulated to cover the technological changes, data requirements and ongoing issues causing slow acceptance of such systems by the clinical procedures.

RQ1. In what ways have deep learning architectures transitioned from standard convolutional neural networks toward the more complex transformer-based and hybrid systems?

RQ2. When considering diagnostic accuracy and the need for clear medical explanations, how do global image classification methods compare to those that focus on specific retinal lesions?

RQ3. To what extent do variations in dataset scale, the precision of annotations, and camera acquisition settings impact the ability of a model to generalize to new patient populations?

RQ4. What are the primary technical and practical barriers, specifically regarding class imbalance and the "black box" nature of AI, that still prevent these models from being fully trusted in a clinical workflow?

RQ5. How are modern hybrid and ensemble strategies attempting to find a middle ground between high-speed computational efficiency and the demand for interpretable, explainable results?

2.2 Search Strategy and Selection Criteria

To conduct this review, the papers have been retrieved from major academic databases such as IEEE Xplore, PubMed, ScienceDirect, and SpringerLink. This was done by ensuring that salient developments were captured exhaustively, with the extraneous noise controlled. The search queries used a combination of certain keywords such as “Diabetic Retinopathy” and “Deep Learning” as well as more recent terms like “Hybrid Transformer Models,” in order to narrow down to how exactly the technology is turning.

The initial findings were filtered through on set criteria in order to get a readily available discussion. The priority was given to articles that were published in the year 2015 to 2025, as the sphere of art is changing very rapidly, and the older work becomes outdated. Further, only those studies that reported clear performance measures, like the Accuracy, Sensitivity or AUC scores, on publicly available datasets like the EyePACS and IDRiD were included. Although a model can be said to be effective, it is necessary to prove it on complex image sets that are widely used.

Certain studies could not but be put aside in the process. For example, any article without peer-reviewed validation or going off track to non-fundus imaging like MRI was rejected. This strict filtering resulted in a final list of thirty-one impactful articles. This list provides a healthy overview of the field, relating the historical milestones from 2016 with the very recent developments of the vision-transformer breakthroughs.

3. LITERATURE REVIEW

The automated diabetic retinopathy (DR) detection is a field of research that has advanced tremendously in the past decade. Initially starkly different machine-learned pipelines have been developed into deep-learning systems that are able to detect tiny retinal lesions with extremely high accuracy. This review is a summary of the key contributions of thirty-one significant articles published in 2015-2025. To further simplify the discussion, the studies are categorized into four major themes, namely, classification-based models, lesion-detection models, beyond classical CNNs approaches, and hybrid or ensemble techniques.

3.1 Classical ML and Early CNN-Based Classification Approaches

The traditional machine-learning classifiers were the first DR detection systems. Carrera et al. (2017) [16] applied the Support Vector Machines (SVMs) for automated DR detection, demonstrating that the classical pipelines were working, however, their performance was low because they used handcrafted features.

Deep learning changed the screening of DR, around 2016. One of the earliest CNN-based pipelines was presented by Doshi et al. (2016) [17], proving that automatically learned visual features are more beneficial than traditional approaches. Abràmoff et al. (2016) [2] and Gulshan et al. (2016) [1] also demonstrated that deep-learning models were able to achieve clinician-level performance when they were trained on big public datasets.

Later, deeper CNN architectures and transfer learning improved robustness. Wan et al. (2018) [3], Lam et al. (2018) [18], and Chakrabarty (2018) [19] designed more sophisticated networks that handled real-world variations. According to Khalifa et al. (2019) [4], there was no rigid data volume requirements in the systems with pre-trained models used efficiently, and Rajalakshmi et al. (2018) [10] applied the technology to smartphone users to serve rural and low-cost patient groups. Table 1 summarizes the main contributions and limitations of these classification-based studies.

Table 1: Summary of Classification-Based Approaches

Author & Year	Contribution	Limitation
Carrera et al. (2017) [16]	SVM-based DR detection using handcrafted features.	Limited generalization; weak against complex lesions.
Abràmoff et al. (2016) [2]	Integrated deep learning for automated DR detection.	Required large datasets and heavy computation.
Gulshan et al. (2016) [1]	Large-scale CNN achieving near clinician-level accuracy.	High sensitivity but lacks lesion-level interpretability.
Doshi et al. (2016) [17]	Early CNN pipeline for 5-class DR classification.	Shallow architecture struggled with subtle lesions.
Rajalakshmi et al. (2018) [10]	Smartphone-based screening using AI.	Sensitive to image quality and lighting variations.
Lam et al. (2018) [18]	Deep CNN with improved normalization/preprocessing.	Whole-image classification only; no localization.
Chakrabarty (2018) [19]	Enhanced CNN architecture for feature extraction.	Potential overfitting due to limited training data.
Wan et al. (2018) [3]	Deep CNN optimized for DR screening.	Lacks lesion-level interpretability (black-box).
Khalifa et al. (2019) [4]	Transfer learning with pre-trained ImageNet models.	May not fully capture specific retinal features.

The persistent catch with these classification systems, however, is that they often act like a black box. Although a model can give a harsh retinopathy grade, it hardly explains the microaneurysms or haemorrhages that led to the decision. Thus, these studies demonstrated that machines could be correct, although there was an enormous distance between trust. The clinicians need to see what the machine is viewing and this is why the field ultimately had to move to lesion-specific detection.

3.2 Lesion Detection and Lesion-Centric DR Screening

As the classification models levelled up, the research community realized that a simple severity rating was not sufficient to have clinical confidence. This resulted in a specific endeavor on lesion identification to enhance the

overall interpretability of such systems. For example, the combination of handmade lesion characteristics and deep learning introduced by Kar and Maity (2017) [14] is an interesting middle ground. They were able to address the divide between conventional ophthalmology and modern neural networks by clearly focusing on microaneurysms, haemorrhages, and exudates.

The Zoom-In-Net, introduced by Wang et al. (2017) [20], was a significant change. It is based on an attention-based mining approach, which basically enables the network to "zoom in" on questionable retinal areas similar to how a doctor might zoom under a magnifying glass. This was then followed by a more elaborate work that was conducted by Qiao et al. (2020) [15] to identify microaneurysm prognosis during the initial non-proliferative stage, using semantic segmentation. A YOLO-based pipeline was used by Santos et al. (2022) [8] to give real-time and accurate lesion localization. These models have had clinical transparency which cannot be equalled by the global classifiers, as indicated via the comparative analysis in Table 2.

Table 2: Summary of Lesion Detection Approaches

Author & Year	Contribution	Limitation
Kar & Maity (2017) [14]	Combined handcrafted features with deep learning.	Limited global severity grading; feature dependent.
Wang et al. (2017) [20]	Zoom-In-Net: attention-based lesion mining.	High computational cost and complex training.
Qiao et al. (2020) [15]	Semantic segmentation for microaneurysm prognosis.	Focused primarily on early-stage DR detection.
Santos et al. (2022) [8]	YOLO-based pipeline for precise lesion localization.	Detection-focused; needs integration for grading.

However, the retinal anatomy is too complicated to be captured by localized lesion observation. Although it helps to find individual spots, it is hard to determine how the spots concerned the broader health of the eye. Thus, although lesion-centric approaches appeared to provide clarity, the field had to proceed to more specialized architectures. This required a shift of the boundaries of classical CNN designs to systems capable of managing more complicated, long-range dependencies, as described in the following section.

3.3 Beyond Classical CNNs

Instead of considering standard designs, researchers started to experiment with more specially-adapted designs to the geometry of retinal pictures. Gargeya and Leng (2017) [12] developed a powerful model with improved extraction features to effectively manage the noise in fundus photography. Zeng et al. (2019) [21] introduced a binocular Siamese-like network, which was an interesting variation to single-eye research. This is just a way of emulating the methodology of a doctor where both eyes are compared to gather common anatomical features. Similarly, Gangwar and Ravi (2020) [22] in collaboration with Khan et al. (2021) [23] investigated various modified architectures such as VGG-NIN, with an aim of having a more effective method to identify subtle pathology variations.

The need for architectural creativity extended beyond the standard convolutions. Kalyani et al. (2023) [24] made use of capsule networks, which are especially fascinating as they seek to preserve the spatial correlations among lesions, which regular CNNs sometimes lose during pooling. Meanwhile, Lin and Wu (2023) [25] decided to modify the classic ResNet-50 with the sole purpose of performing retinal texture analysis by focusing on the fine-grained properties of the fundus. It is important to note as well that other studies, such as that of Li et al. (2019) [26], the concept of fundus photography is altogether discarded in favor of Optical Coherence Tomography (OCT) to show the cross-sectional structural aspects that a flat image cannot possibly show. Further performance improvement was also reported by Tymchenko et al. (2020) [27], Mushtaq and Siddiqui (2021) [28], and Dai et al. (2021) [11]. The summarization of these findings is presented in Table 3 and it offers an insight into the various means through which academics have attempted to streamline the learning process.

Table 3. Summary of Approaches Beyond Classical CNNs

Author & Year	Contribution	Limitation
Gargeya & Leng (2017) [12]	Custom residual CNN with enhanced feature extraction.	Primarily focused on classification accuracy.

Author & Year	Contribution	Limitation
Zeng et al. (2019) [21]	Binocular Siamese-like CNN for stereo-vision features.	High computational cost for binocular analysis.
Li et al. (2019) [26]	OCT-based DR detection for structural details.	Modality-specific; limited dataset availability.
Tymchenko et al. (2020) [27]	Multi-stage EfficientNet ensemble approach.	Complex multi-stage training process.
Gangwar & Ravi (2020) [22]	Specialized transfer learning hybrids.	Still lacks lesion-level clinical explanation.
Khan et al. (2021) [23]	VGG-NIN architecture for improved feature learning.	Large model size requiring significant data.
Mushtaq & Siddiqui (2021) [28]	DenseNet-169 for automated screening.	Limited lesion-level interpretability.
Dai et al. (2021) [11]	DeepDR system for full disease spectrum detection.	Computationally intensive black-box model.
Kalyani et al. (2023) [24]	Capsule networks for spatial lesion relationships.	Complex architecture with longer training times.
Lin & Wu (2023)	Revised ResNet-50 for improved texture analysis.	High computational resource requirements.

Many of these specialized designs, however, although indisputably stretching the limits of accuracy, were still associated with pure categorization or pure detection. This sustained division indicated an unresolved gap in the literature, which eventually strengthened the case of hybrid models capable of balancing between global and local specifics.

3.4 Hybrid and Ensemble Approaches

Hybrid models have been found to be more flexible as an answer to the limitations inherent in the application of a single architecture. The reasoning behind this is that one model could have a blind spot that is taken care of by another. Pires et al. (2019) [13] and Nguyen et al. (2020) [29] implemented multi-feature, data-driven pipelines that were no longer based on the conception of a "single-size-fits-all" network. Likewise, Qummar et al. (2019) [30] have also shown that a properly built CNN ensemble could surpass its individual components on the basis of integrating various visual viewpoints.

These hybrid models have been able to generate significant impetus since they lead to a more holistic view of the condition. Bilal et al. (2021) [6] developed structures that are cannot only detect but also manage severity grading in a single run. This is in a larger trend of the recent work of integrating local lesion localization with global contextual information. Indicatively, Alyoubi et al. (2021) [31] suggested a single system that can handle both the activities concurrently, in an attempt to obtain a more comprehensive diagnostic image.

Wahab Sait (2023) [9] introduced lightweight models that are directly applicable to real-time application, and Jabbar et al. (2024) [7] studied the optimization strategies such as Accelerated Particle Swarm Optimization to adjust the performance. The most advanced of such direction is the work of Waleed et al. (2025) [5], who presented YOLO-ViT. The combination of the localization accuracy of YOLO and the wide-view attention operations of a Vision Transformer helped them to create a single diagnostic tool that is much closer to the experience of a human expert as they scan a retina. Table 4 provides a summary of the details of these hybrid and ensemble models.

Table 4. Summary of Hybrid and Ensemble Approaches

Author & Year	Contribution	Limitation
Pires et al. (2019) [13]	Multi-feature, data-driven detection pipeline.	Complexity increases with feature engineering.
Qummar et al. (2019) [30]	CNN ensemble for robust DR grading.	Large ensemble size; high inference cost.
Nguyen et al. (2020) [29]	VGG-based hybrid with ensemble classifiers.	Feature fusion is difficult to optimize.

Author & Year	Contribution	Limitation
Bilal et al. (2021) [6]	Mixed model for detection and severity grading.	Larger model size and integration complexity.
Alyoubi et al. (2021) [31]	Unified lesion localization + classification system.	Requires high-quality annotated lesion data.
Wahab Sait (2023) [9]	Lightweight hybrid model for real-time use.	Potential accuracy trade-off for speed.
Jabbar et al. (2024) [7]	GoogleNet + ResNet + APSO optimization.	High architectural and optimization complexity.
Waleed et al. (2025) [5]	YOLO-ViT: Hybrid detection + Transformer attention.	Requires significant computation and data.

Finally, this transition to multi-task models indicates that the discipline is heading out of the domain of solitary models. These newer designs are much more realistic in the way that they were developed with a balance of accuracy, interpretability, and computational efficiency, in contrast to the single-architecture models that preceded them.

4. COMPARATIVE ANALYSIS OF DATASETS, PREPROCESSING, AND METHODOLOGICAL TRENDS

The reviewed papers in this study reveal the various strengths and weaknesses of the various methods used in detecting diabetic retinopathy. Although all of the approaches are designed to enhance the diagnostic accuracy, the issues are solved differently by each category.

4.1 Overview of Datasets Used in Diabetic Retinopathy Studies

The evaluation of the architecture to diabetic retinopathy (DR) detection systems is likely to result in a lack of focus, especially when trying to find sources of accuracy. However, the information which formed the basis of such studies is usually the main story. The datasets used for training and evaluation—their size, label quality, and the conditions under which the photos were captured—determine the performance of a model in the real-time. Without a thorough insight into these data properties, they would not be able to make direct comparisons of findings in the research. Table 5 summarizes the key sources of public datasets, which, most commonly, were used in the studies being reviewed.

Table 5. Major Datasets Used in Reviewed Diabetic Retinopathy Detection Studies

Dataset	Scale	Dataset Size	Annotation Type	Representative Papers
EyePACS	Large-scale	>80,000 fundus images	Image-level DR grading (5 classes)	[1], [3], [7], [9], [11], [12], [13], [17], [18], [20], [21], [23], [25], [27], [28], [30]
Messidor / Messidor-2	Medium-scale	~1,200 / 1,745 fundus images	Image-level DR grading	[1], [2], [12], [13], [14], [16], [18], [20], [22], [24], [27]
APTOS 2019	Medium-scale	~3,600 fundus images	Image-level DR grading (5 classes)	[4], [5], [9], [22], [27], [28], [31]
IDRiD	Small–medium	516 fundus images	Pixel-level lesion annotations and DR grading	7, 8, 14, 15, 27
DDR	Medium–large	~13,000 fundus images	Pixel-level lesion annotations	31, 8

Looking at the landscape, EyePACS dataset has a significant impact. Its size is large and, thus, is a preferred option of data source to use in training deep learning model with the large data volume. A number of studies have discovered that models that do well on the EyePACS don't work effectively while testing on datasets like Messidor-2. Such a shift in domain is a major barrier, indicating the fact that external validation is more reliable evidence when compared to internal system reviews.

On the other hand, some datasets as APTOS 2019 are ideal to test transfer learning and lightweight models, particularly when the researcher is operating under constrained computing limitations. There are also specialized sets like IDRiD and DDR. In spite of the fact that these sets comprise relatively small number of images, they include

lesions level annotations which are not available in EyePACS. Therefore, they cannot be overlooked by the researchers who need to detect objects or build AI systems that would effectively identify particular pathological alterations.

Though many studies are still based on smaller or more focused databases like DRIVE or STARE or even on private hospital data, they are generally used in much more specific and focused studies and not in general benchmarking. Consequently, to achieve a broad perspective in the field, then emphasis should be laid on commonly used datasets. Lastly, the literature points out that a model can be as good as the data that it was constructed. This indicates how important it is that researchers should switch to multi-dataset assessment and more uniform benchmarks.

4.2 Overview of Preprocessing Techniques

Reviewing the raw fundus images, it is clear that such images can hardly be a convenient source of knowledge about the eyes, which underlies the use of deep learning models. More often, the problems that are faced by the researchers include limited lighting, low contrast, and improper noise distributions, which differ according to the particular camera used in a clinic. By feeding raw inputs directly into a neural network, the algorithm is supposed to automatically extract and determine pathological indicators in the form of noisy representation. Therefore, preprocessing has become an underrated, yet crucial part in the literature. Despite the fact that studies may adopt a variety of methods, some generalized methodologies are seen in nearly all the thirty-one reviewed works.

The process of pipelines starts with resizing and intensity normalization. Though it is a simple method, different architectures remain compatible across various imaging hardware. Another common technique that researchers use is to crop out non-informative areas. The model is made to concentrate its computational energy on the anatomical content instead of noise by eliminating the heavy black margins that surround the retinal field.

Contrast enhancement is normally given importance to tackle the issue of poor lighting. The methods using histograms are preferable because of their ability to reveal the background microaneurysms which include the Contrast-Limited Adaptive Histogram Equalization (CLAHE). Several studies isolate the green channel of the image, which is common in ophthalmology, since the green channel gives the most enhancement to blood vessels and lesions.

Another important factor is noise reduction. Whether the researchers use a Gaussian filter to blur artifacts or a median filter to maintain edges, their objective is to offer the model a cleaner canvas. Certain methods make use of morphological operations in order to improve structural detail, but there is ever a certain intellectual aversion. It is necessary to question whether over filtering might dilute the smallest lesions that are essential in making an early diagnosis.

Furthermore, data augmentation is almost ubiquitous because of the existence of class imbalance or low spatial resolution in publicly accessible datasets. Through rotation, flipping, and scaling of the pictures, the researchers train their models to stay stable in the event that a patient blinks or the camera is slightly tilted. In the more specialized work on lesion-level segmentation, custom mask generation and region-of-interest extraction are additional steps that are regularly utilized.

Finally, although there seems to be no single gold standard pipeline, the overall agreement is obvious. Through these normalization and refinement processes, researchers can transform a noisy and inconsistent set of data into something that can actually be learnt by a model.

4.3 Overview of Models and Results

The automated diagnosis of diabetic retinopathy (DR) has experienced rapid developments in the last decade, with the advancement of innovative deep learning structures, lesion-centred strategies, and hybrid models. In order to be able to systematize this development, the studies chosen have been categorized into four groups according to their major contributions of the methods employed. This helps to conduct a proper comparison of model systems, datasets involved, and important performance indicators, focusing on past foundations as well as more recent discoveries in the realm of DR detection studies.

Classification-Based Models

Classification-based methods provided the first substantive for automated diabetic retinopathy detection. These systems are essentially designed to provide a final diagnosis that is, whether a certain image has any signs of disease and whether it has reached its advanced stages or not. Most of the earlier breakthroughs in the literature rely on the heavy lifting of convolutional neural networks. Due to the data-intensive nature, these models performed well with large and high-profile datasets where they may be trained to differentiate between the global phenotypic properties of

healthy and pathological eyes. A further comparison of the performance of these different models as well as particular architectural designs and experimental outcomes is discussed in Table 6.

Table 6. Classification-Based DR Detection Approaches

Author (Year)	Model / Architecture	Dataset(s) Used	Key Results
Carrera et al. (2017) [16]	SVM-based classifier	Messidor	85.1% Accuracy; 77% AUC
Abràmoff et al. (2016) [2]	IDx-DR (Deep Learning)	Messidor-2	96.8% Sensitivity; 0.980 AUC
Gulshan et al. (2016) [1]	Inception-v3	EyePACS, Messidor-2	0.991 AUC (EyePACS)
Doshi et al. (2016) [17]	Custom CNN	EyePACS	0.399 QWK
Rajalakshmi et al. (2018) [10]	Smartphone-based CNN	296 Smartphone images	95.8% Sens; 80.2% Spec
Lam et al. (2018) [18]	Custom CNN + GoogLeNet	EyePACS, Messidor-1	95% Binary Sensitivity
Chakrabarty (2018) [19]	Shallow CNN + Dense ANN	Local (30 images)	100% Validation Accuracy
Wan et al. (2018) [3]	VGG, ResNet, GoogLeNet	EyePACS	95.68% Acc (VGGNet-s)
Khalifa et al. (2019) [4]	AlexNet, VGG, ResNet	APTOS 2019	97.9% Acc (AlexNet)

These models usually have a high global accuracy but they are not clinical in their transparency. Even a basic severity rating, irrespective of its accuracy, does not necessarily provide the local evidence that an ophthalmologist needs to make a qualified decision. This fundamental constraint has increasingly focused the research interest to lesion-centric approach. This investigation is aimed at building the interpretability behind trust in the real-world clinical environment by putting aside the "black-box" predictions in favor of precisely localizing the specific retinal abnormalities.

Lesion Detection Methods

Lesion centric methods are designed to determine the accurate retinal defects such as microaneurysms, haemorrhages, and exudates, which determine diabetic retinopathy severity. These approaches have higher clinically interpretability as compared to pure classification models in that they identify the locations where the disease-relevant features manifest in the fundus image. The selected studies in this category cover early handcrafted–deep learning hybrids, attention-based lesion mining, and modern YOLO-based detectors. Table 7 is a summary of lesion detection representative methods that have been studied in this paper along with their datasets and results.

Table 7. Lesion Detection–Based DR Approaches

Author (Year)	Model / Architecture	Dataset(s) Used	Key Results
Kar & Maity (2017) [14]	Matched/LoG Filtering	DRIVE, STARE, MESSIDOR	97.23% MA Acc; 96.99% HE Acc
Wang et al. (2017) [20]	Zoom-in-Net	EyePACS, Messidor	0.854 QWK (Test)
Qiao et al. (2020) [15]	Semantic Segmentation	IDRiD	97.4% Sens (Dark Lesions)
Santos et al. (2022) [8]	YOLOv5	DDR, IDRiD	22.4% AP (HE); 11.1% AP (MA)

Although lesion detectors are precise, the loss of long-range spatial information is a liability and architectures like Transformers and attention mechanisms are implemented to build the contextual knowledge.

Beyond Classical CNNs

Recent studies have started to go beyond the classic CNN architecture and shift their focus to more advanced features such as attention mechanisms, transformer-based encoders, and self-supervised learning. Such strategies do not simply refer to the pursuit of higher percentages. Instead, they are a sincere attempt to address the remaining problems in a strong and context-aware manner that can be very much afflicted by simply convolutional models.

By performing long-range dependency modeling, these techniques appear to represent a far richer set of features, which contributes to them being more adaptive to various datasets or lighting conditions. It is a definite

change in the discipline, whereby it has transformed into less structural systems in the favor of more articulate and pliable systems. Table 8 provides a comprehensive discussion of these developed models strategies and reported findings.

Table 8. Deep Learning Methods Vibrant Over Classical CNNs.

Author (Year)	Model / Architecture	Dataset(s) Used	Key Results
Gargeya et al. (2017) [12]	ResNet + Gradient Boosting	EyePACS, Messidor-2	0.97 AUC (EyePACS)
Zeng et al. (2019) [21]	Binocular Siamese CNN	EyePACS	0.951 AUC; 0.829 QWK
Li et al. (2019) [26]	OCTD_Net (DenseNet-SE)	WMU (OCT Dataset)	92% Acc; 90% Sens
Tymchenko et al. (2020) [27]	EfficientNet + SE-ResNeXt	EyePACS, IDRiD	0.925 QWK
Gangwar et al. (2020) [22]	Inception-ResNet-v2 Hybrid	Messidor-1, APTOS	82.18% Acc (APTOS)
Khan et al. (2021) [23]	VGG-NiN (SPP + NiN)	EyePACS	81.8% Avg AUC
Mushtaq et al. (2021) [28]	DenseNet-169	EyePACS, APTOS	90% Accuracy
Dai et al. (2021) [11]	DeepDR (ResNet + Mask R-CNN)	CNDCS, EyePACS	0.943–0.962 AUC
Kalyani et al. (2023) [24]	Reformed Capsule Network	Messidor	97.98% Acc; 96.11% Recall
Lin & Wu (2023) [25]	Revised ResNet-50	EyePACS, JSIEC	94.85% Average Acc

However, the main difficulty is the ability to balance local detail and global thinking. It is precisely the reason of the overturning into hybrid and ensemble structures. Through combining the capabilities of the different architectures, researchers are trying to develop a highly dependable solution that can endure the uncertainty of the real-world screening context.

Hybrid and Ensemble Models

Recently, there has been a strong trend towards hybrid and ensemble designs that are establishing a very eminent niche in the realm of diabetic retinopathy studies. The main principle held is the non-reliance based on a single architecture but rather the merge of complementary technologies. A typical example of this can be in combining the local, feature-shredding properties of a CNN with the global view of a transformer to accept a single framework to simultaneously perceive both the large retinal view and the small and fine-structure lesion features.

Ensemble learning has remained of relevance, although individual models acquire more complexity, researchers discover that they can be considerably more stable when layers are stacked on one another, or that when they are designed together in hybrid form. This makes them a very attractive option for the uncertain nature of real-world clinical implementation. The fact that this transition allows models to deal with the enormous multiplicity of fundus images that project any real screening campaign is perhaps the most interesting part. Table 9 represents a summary of these hybrid and ensemble models as well as their design and performance metrics.

Table 9. Hybrid and Ensemble DR Detection Approaches

Author (Year)	Model / Architecture	Dataset(s) Used	Key Results
Pires et al. (2019) [13]	16-layer CNN + NN/RF	EyePACS, Messidor-2	98.2% AUC (Messidor-2)
Qummar et al. (2019) [30]	Ensemble (ResNet, Inception)	EyePACS	80.8% Acc; 0.97 AUC (PDR)
Nguyen et al. (2020) [29]	VGG + Custom + SVM/RF	EyePACS	80–83% Accuracy
Bilal et al. (2021) [6]	Mixed (SVM+KNN+DT)	IDRiD	92.3% Acc; 100% Spec
Alyoubi et al. (2021) [31]	CNN512 + YOLOv3	DDR, APTOS 2019	89% Acc; 89% Sens
Wahab Sait (2023) [9]	YOLOv7 + MobileNetV3	EyePACS, APTOS	98.4% Acc (EyePACS)
Jabbar et al. (2024) [7]	GoogleNet + ResNet + APSO	EyePACS	94% Acc; 0.97 Precision

Author (Year)	Model / Architecture	Dataset(s) Used	Key Results
Waleed et al. (2025) [5]	YOLO-ViT (YOLOv8 + ViT)	Not Disclosed	96.89% Acc; 0.97 F1-score

5. CHALLENGES

Although automated diabetic retinopathy (DR) detection has evolved over the years, it still has certain issues that make the technology not ideal in the real world and its large-scale adoption in clinics.

- **The Problem of Data Imbalance:** Almost in any particular set of data, it is easy to find cases of advanced disease or immaculate healthy eyes, but the small, early-stage signs like tiny microaneurysms are much rarer. Therefore, a model may be a good one in identifying a severe case where the pathology is obvious but fail to detect the mild stages. This is a tragedy in itself because those mild symptoms are the very point at which treatment may actually save sight.
- **Sensitivity to Image Quality:** A deep learning model can be quite effective when dealing with sharp, brightly lit images of the dataset, but it can entirely fail to deliver functionality as soon as faced with the blur or low light of a handheld camera in an overcrowded screening line. These differences in hardware and lighting are not just minor irritations, rather they reduce reliability which make them less useful in the real world.
- **The Lack of Standardized Benchmarking:** The use of a few publicly available datasets like EyePACS or APTOS causes geographic blind spot. In addition, several researchers continue to make use of the data which is personal and specific to the hospital and cannot be verified by any other individual. This renders it almost impossible to know if a model is actually robust or merely reciting the peculiarities of a particular camera model or type of patients.
- **Interpretability:** Convolutional neural networks classify the level of severity but not necessarily provide reasons as to why they have considered a particular eye as diseased. Although attention processes and localisation of lesions are coming into focus, many clinicians are reluctant to rely on a machine that cannot identify a particular haemorrhage or exudation to make its decision.
- **Resource Constraints in Clinical Settings:** State-of-the-art models, in particular, recent hybrid transformers, require a significant computational power. Therefore, it is easy to execute a model on a stack of high-end GPUs in a university lab, but very hard to make it efficient in a low-resource healthcare environment.

6. DISCUSSIONS AND KEY INSIGHTS

In a retrospective analysis of these thirty-one studies, it is quite obvious that the technologies of diabetic retinopathy (DR) detection have developed rather rapidly. Firstly, researchers paid much attention to classical CNNs and transfer learning. The initial models were able to provide a baseline; however, they generally showed that they did not really have an idea of retinal imaging. In the past ten years, the attention has moved towards architectures concerned with explainability and the fine-grained detail of lesions.

It is necessary to mention that no single model that has become extremely clear-cut a “winner” in all datasets. Performance seems to be dynamic and greatly influenced by the visual perception and the distribution of classes. For example, those models trained on the large library of EyePACS exhibit excellent generalization, and those trained on pixel-level annotated datasets, such as, IDRiD, exhibit excellent diagnostic depth. This affirms the fact that the quality of medical imaging resources is equally critical to the level of sophistication of the algorithms themselves.

The same approach of lesion-centered is also gaining ground as a consensus. With the aid of such tools as YOLO or attention mechanisms, researchers are eventually offering clinicians results that can be relied upon. Instead of providing a clear “yes” or “no,” such models indicate individual microaneurysms. This is especially important in early-stage screening in which the symptoms are so subtle that they would be almost tough to detect.

Furthermore, it is also clear that hybrid and ensemble systems tend to be better than single-model pipelines. The integration of CNN spatial power and the global context of a Vision Transformer is far more likely to generate more stable predictions. The downside, however, is that this benefit goes at the cost of the models being computationally intensive, and therefore not very practical in real-time clinical settings.

Another trend is related to lightweight and mobile-friendly architectures. It is not merely a convenience to have

a small model of fundus camera with the spread of smartphone-based cameras. Thus, it is necessary to bridge the gap between high-technology research laboratories and community-based healthcare in underserved rural communities.

Lastly, literature indicates that high precision is not the sole aspect of the picture. Scholars are acknowledging the fact that in order to have a valuable model, it should be transparent, fair with a variety of populations, and tested in a real situation. These technologies might require more established criteria and more validation in practice before they can be implemented with full confidence.

7. CONCLUSION AND FUTURE SCOPE

The fight against diabetic retinopathy has been fruitful in the last decade. There has been an evolution from manual and non-consistency grading to an era where a retinal scan can be quickly comprehended through the use of deep learning. Examining these thirty-one important studies between 2015-2025, a clear technological direction has been achieved. Starting from simple classifiers to the reign of powerful CNNs, the transition has finally arrived at the more complex hybrid and transformer designs that make up the present state-of-the art.

Nevertheless, despite these steady improvements in accuracy and localization of lesions, there still are several un-resolvable gaps. Even the majority of the models continue to stumble at the incoherent reality of a dismal fundus image or the fine, nearly invisible, indicator of an early-stage disease. The issue of trust still persists. Clinicians need to be given an understanding of the reasoning behind a prediction, and even an image-level rating is no longer what is expected by clinicians.

The most important conclusion that has come out of this evaluation is the fact that none of the designs have been able to address the entire challenge yet. There are great models of categorization that make excellent predictions on the global scale but they are not transparent. On the other hand, there are lesion detectors which have excellent information but fail to look at the bigger picture of the severity of the disease. The possibility of closing this gap has the most potential of innovation in the future. Cohesive and hybrid model, which is capable of processing both granular detail of lesion and the wide-range retinal information, is necessary.

Future studies of automated diabetic retinopathy detection ought to concentrate on development of robust models that are generalizable to real-life fundus photos where different equipment is used on different patient demographics. Establishing standardized, massive and clinically validated data will play a vital role in improving model reliability and equity. Another area of possible research is hybrid architectures that bring together the ability to identify lesions with high accuracy and the potential to represent global features such as combining YOLO-based detectors with CNNs or Vision Transformers. These models have the potential to enhance severity grading accuracy as well as clinical interpretability without incurring a significant computational cost to perform real-time screening on a mobile or low-cost device. Lastly, there is a need to conduct major multi-centric clinical validation of explainable AI approaches to guarantee a stable performance and reliability of the method in practical screening settings by clinicians.

References

1. V. Gulshan, L. Peng, M. Coram, M. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. Nelson, J. Mega, and D. Webster, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *JAMA*, vol. 316, pp. 2402–2410, 12 2016.
2. M. D. Abra`moff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, “Improved automated detection of diabetic retinopa-thy on a publicly available dataset through integration of deep learning,” *Investigative Ophthalmology & Visual Science*, vol. 57, pp. 5200–5206, 10 2016.
3. S. Wan, Y. Liang, and Y. Zhang, “Deep convolutional neural networks for diabetic retinopathy detection by image classification,” *Computers & Electrical Engineering*, vol. 72, pp. 274–282, 2018.
4. N. E. Khalifa, M. Loey, M. Taha, and H. Mohamed, “Deep transfer learning models for medical diabetic retinopathy detection,” *Acta Infor-matica Medica*, vol. 27, pp. 327–332, 12 2019.
5. M. Waleed, M. Yasser, M. Magdy, M. Elkady, T. Hesham, A. Adel, Y. Sherif, M. Alazzaly, and M. Marey, “YOLO-ViT: A hybrid deep learning model for eye disease classification,” in *Proceedings of the International Conference on Advanced Intelligent Systems*, pp. 3–12, Springer, 10 2025.
6. A. Bilal, G. Sun, Y. Li, S. Mazhar, and A. Khan, “Diabetic retinopathy detection and classification using mixed models for a disease grading database,” *IEEE Access*, vol. 10, pp. 18990–19010, 02 2022.
7. A. Jabbar, H. Liaqat, A. Akram, M. U. Sana, I. Azp´ıroz, I. De la Torre D´ıez, and I. Ashraf, “A lesion-based diabetic retinopathy detection through hybrid deep learning model,” *IEEE Access*, vol. 12, pp. 39024–39036, 03 2024.

8. C. Santos, M. Aguiar, D. Welfer, and B. Belloni, “A new approach for detecting fundus lesions using image processing and deep neural network architecture based on yolo model,” *Sensors*, vol. 22, no. 17, 2022.
9. A. R. Wahab Sait, “A lightweight diabetic retinopathy detection model using a deep-learning technique,” *Diagnostics*, vol. 13, no. 19, 2023.
10. R. Rajalakshmi, R. Subashini, R. Anjana, and V. Mohan, “Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence,” *Eye*, vol. 32, pp. 1138–1144, 03 2018.
11. L. Dai, L. Wu, H. Li, C. Cai, Q. Wu, H. Kong, R. Liu, X. Wang, X. Hou, Y. Liu, X. Long, Y. Wen, L. Lu, Y. Shen, Y. Chen, X. Yang, H. Zou, B. Sheng, and W. Jia, “A deep learning system for detecting dia-betic retinopathy across the disease spectrum,” *Nature Communications*, vol. 12, p. 3242, 05 2021.
12. R. Gargeya and T. Leng, “Automated identification of diabetic retinopa-thy using deep learning,” *Ophthalmology*, vol. 124, pp. 962–969, 07 2017.
13. R. Pires, S. Avila, J. Wainer, E. Valle, M. D. Abramoff, and A. Rocha, “A data-driven approach to referable diabetic retinopathy detection,” *Artificial Intelligence in Medicine*, vol. 96, pp. 93–106, 2019.
14. S. Sil Kar and S. Maity, “Automatic detection of retinal lesions for screening of diabetic retinopathy,” *IEEE Transactions on Biomedical Engineering*, vol. 65, pp. 608–618, 03 2018.
15. L. Qiao, Y. Zhu, and H. Zhou, “Diabetic retinopathy detection us-ing prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms,” *IEEE Access*, vol. 8, pp. 104292–104302, 2020.
16. E. V. Carrera, A. Gonza´lez, and R. Carrera, “Automated detection of diabetic retinopathy using SVM,” in *2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing* (IN-TERCON), (Cusco, Peru), pp. 1–4, IEEE, 08 2017.
17. D. Doshi, A. Shenoy, D. Sidhpura, and P. Gharpure, “Diabetic retinopa-thy detection using deep convolutional neural networks,” in 2016 *In-ternational Conference on Computing, Analytics and Security Trends (CAST)*, (Pune, India), pp. 261–266, IEEE, 12 2016.
18. C. Lam, D. Yi, M. Guo, and T. Lindsey, “Automated detection of diabetic retinopathy using deep learning,” *AMIA Joint Summits on Translational Science Proceedings*, vol. 2017, pp. 147–155, 2018.
19. N. Chakrabarty, “A deep learning method for the detection of diabetic retinopathy,” in *2018 5th IEEE Uttar Pradesh Section International Con-ference on Electrical, Electronics and Computer Engineering* (UPCON), (Gorakhpur, India), pp. 1–5, IEEE, 11 2018.
20. Z. Wang, Y. Yin, J. Shi, W. Fang, H. Li, and X. Wang, “Zoom-in-net: Deep mining lesions for diabetic retinopathy detection,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI* 2017, vol. 10435 of Lecture Notes in Computer Science, (Cham), pp. 267–275, Springer, 09 2017.
21. X. Zeng, H. Chen, Y. Luo, and W. Ye, “Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network,” *IEEE Access*, vol. 7, pp. 30744–30753, 03 2019.
22. A. K. Gangwar and V. Ravi, “Diabetic retinopathy detection using transfer learning and deep learning,” in *Evolution in Computational Intelligence* (V. Bhateja, S.-L. Peng, S. C. Satapathy, and Y.-D. Zhang, eds.), vol. 1176 of Advances in Intelligent Systems and Computing, pp. 679–689, Singapore: Springer, 2021.
23. Z. Khan, F. G. Khan, A. Khan, Z. U. Rehman, S. Shah, S. Qummar, F. Ali, and S. Pack, “Diabetic retinopathy detection using VGG-NIN a deep learning architecture,” *IEEE Access*, vol. 9, pp. 61408–61416, 04 2021.
24. G. Kalyani, B. Janakiramaiah, A. Karuna, and L. V. N. Prasad, “Dia-betic retinopathy detection and classification using capsule networks,” *Complex & Intelligent Systems*, vol. 9, no. 3, pp. 2651–2664, 2023.
25. C.-L. Lin and K.-C. Wu, “Development of revised ResNet-50 for diabetic retinopathy detection,” *BMC Bioinformatics*, vol. 24, p. 175, 04 2023.
26. X. Li, L. Shen, M. Shen, F. Tan, and C. S. Qiu, “Deep learning based early stage diabetic retinopathy detection using optical coherence tomography,” *Neurocomputing*, vol. 369, pp. 134–144, 2019.
27. B. Tymchenko, P. Marchenko, and D. Spodarets, “Deep learning ap-proach to diabetic retinopathy detection.” arXiv preprint, 03 2020. arXiv:2003.02261 [cs.CV].
28. G. Mushtaq and F. Siddiqui, “Detection of diabetic retinopathy using deep learning methodology,” *IOP Conference Series: Materials Science and Engineering*, vol. 1070, p. 012049, 02 2021.
29. Q. H. Nguyen, R. Muthuraman, L. Singh, G. Sen, A. C. Tran, B. P. Nguyen, and M. Chua, “Diabetic retinopathy detection using deep learn-ing,” in *Proceedings of the 4th International Conference on Machine Learning and Soft Computing,* ICMLSC ’20, (New York, NY, USA), pp. 103–107, Association for Computing Machinery, 2020.
30. S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Band, Z. U. Rehman, I. Khan, and W. Jadoon, “A deep learning ensemble approach for diabetic retinopathy detection,” *IEEE Access*, vol. 7, pp. 150530–150539, 10 2019.
31. W. L. Alyoubi, M. F. Abulkhair, and W. M. Shalash, “Diabetic retinopa-thy fundus image classification and lesions localization system using deep learning,” *Sensors*, vol. 21, no. 11, 2021.