IJVAST | Vol. 1, Issue 4 | April 2026 | Open Access — Peer Reviewed Review Article

A Review Paper on Machine Learning Approaches for Early Diabetic Retinopathy Detection

Ajay Bagwat[1], Pravin Pawade[2]
Department of Computer Science and Engineering
D. Y. Patil Medical College, Pune, Maharashtra, India

Abstract—Diabetic Retinopathy (DR) is a leading cause of preventable blindness among working-age adults worldwide, arising as a microvascular complication of long-standing diabetes mellitus. Early detection through regular retinal screening is critical for preventing irreversible vision loss, yet manual grading by ophthalmologists is time-consuming, subject to inter-observer variability, and difficult to scale to population-wide screening, particularly in resource-constrained settings. This review paper examines the evolution of Machine Learning (ML) and Deep Learning (DL) approaches for early and automated DR detection, covering classical machine learning pipelines built on hand-crafted features, convolutional neural network (CNN) architectures trained end-to-end, transfer learning strategies, and ensemble/hybrid models. We review representative studies published between 2016 and 2024 evaluated on major public fundus image datasets including EyePACS, APTOS 2019, Messidor, Messidor-2, and IDRiD. Preprocessing techniques such as contrast enhancement and vessel/lesion segmentation, together with explainability methods such as Grad-CAM and attention maps that support clinical trust, are discussed. The paper compares reported sensitivity, specificity, and area-under-curve performance across representative studies, identifies persistent challenges such as class imbalance, dataset shift across cameras and populations, small-lesion detection, and limited external validation, and outlines future directions including multi-modal learning, self-supervised pretraining, federated learning, and point-of-care screening tools.

Index Terms—Diabetic Retinopathy, Deep Learning, Convolutional Neural Networks, Fundus Image Analysis, Transfer Learning, Automated Screening, Image Classification, Explainable AI, Class Imbalance.

I. Introduction

Diabetes mellitus affects hundreds of millions of people globally, and Diabetic Retinopathy (DR) develops in a substantial proportion of long-term diabetic patients as damage accumulates within the retinal microvasculature. DR progresses through recognized stages — no DR, mild, moderate, and severe non-proliferative DR, and proliferative DR — with vision-threatening complications such as diabetic macular edema (DME) emerging as severity increases. Early detection at the mild or moderate stage allows timely referral and treatment, such as laser photocoagulation or anti-VEGF injections, that can prevent progression to blindness. Systematic screening, however, is constrained by the limited availability of trained ophthalmologists, particularly in low- and middle-income regions, motivating the development of automated, ML-based DR detection systems capable of operating on fundus camera images at scale.

Early rule-based and hand-crafted-feature systems gave way, following the resurgence of deep learning, to convolutional neural networks that learn discriminative features directly from retinal images. A landmark validation study demonstrated that a deep learning algorithm could detect referable DR from fundus photographs with sensitivity and specificity comparable to expert ophthalmologists [1], catalyzing a wave of research applying CNNs, transfer learning, and ensemble methods to DR grading and lesion detection.

This review surveys the ML/DL literature on DR detection published between 2016 and 2024. The primary

III. ML and DL Approaches for DR Detection

A. Classical Machine Learning Pipelines

Early automated DR detection systems relied on hand-crafted feature extraction — including microaneurysm and hemorrhage detection via morphological operations, vessel segmentation, and exudate detection via color thresholding — followed by classifiers such as Support Vector Machines, Random Forests, and k-Nearest Neighbors [2]. These pipelines achieved moderate accuracy but were sensitive to image quality and illumination variation and required extensive domain-specific feature engineering, limiting scalability across datasets collected with different camera equipment.

B. Convolutional Neural Network Architectures

CNN-based approaches learn hierarchical features directly from raw fundus images, removing the need for manual feature engineering. Pratt et al. proposed an end-to-end CNN architecture for DR grading, achieving reasonable sensitivity on the EyePACS dataset while highlighting the difficulty of detecting early-stage lesions [3]. Subsequent work applied deeper architectures such as VGG, ResNet, and Inception, adapted for five-class DR severity grading, generally improving accuracy at the cost of requiring larger labeled training sets.

C. Transfer Learning Strategies

Given the relatively limited size of expert-labeled DR datasets compared to natural image corpora, transfer learning from ImageNet-pretrained CNNs has become a dominant strategy. Fine-tuning pretrained backbones on fundus

contributions are: (1) a structured comparison of classical ML and deep learning approaches to DR detection and grading; (2) a review of preprocessing, segmentation, and explainability techniques used to improve model performance and clinical interpretability; (3) a comparative analysis of representative studies and their reported performance metrics; (4) identification of open challenges including class imbalance and generalization across populations; and (5) a discussion of emerging research directions.

II. Background on Diabetic Retinopathy

A. Clinical Staging

DR severity is commonly graded using the International Clinical Diabetic Retinopathy (ICDR) severity scale into five classes based on the presence and extent of microaneurysms, hemorrhages, exudates, venous beading, and neovascularization. Automated grading systems are typically framed either as binary classification (referable vs. non-referable DR) or multi-class classification across the five ICDR grades.

B. Screening Workflow

Conventional DR screening involves pupil dilation, fundus photography, and grading by a trained reader or ophthalmologist, followed by referral for patients with referable disease. Automated ML-based grading is intended to serve as a first-line triage tool, flagging likely-referable cases for specialist review and reducing the grading burden on a limited ophthalmic workforce.

images, combined with aggressive data augmentation, consistently improves convergence speed and generalization relative to training from scratch, particularly under limited and imbalanced training data [4].

D. Ensemble and Hybrid Models

To improve robustness, several studies combine multiple CNN architectures or integrate deep features with traditional classifiers. Ensemble approaches that aggregate predictions from several independently trained networks have been shown to improve quadratic-weighted kappa scores on multi-class DR grading benchmarks relative to single-model baselines, at increased computational cost [5].

IV. Comparison of Approaches

Table I summarizes representative approaches to automated DR detection. Classical ML pipelines remain the most interpretable and computationally lightweight but underperform on early-stage/mild DR relative to CNN-based approaches. Transfer learning consistently improves performance over training CNNs from scratch, especially given the limited size of expert-annotated DR datasets, while ensemble and hybrid methods yield the strongest reported metrics at the cost of increased model complexity and training time.

TABLE I — Comparison of ML/DL Approaches for DR Detection

Approach	Key Technique	Interpretability	Data Requirement	Early-Stage Sensitivity
Classical ML (SVM/RF) [2]	Hand-crafted features	High	Low	Low
CNN (from scratch) [3]	End-to-end learning	Moderate	High	Moderate
Transfer Learning [1],[4]	Pretrained backbone fine-tuning	Moderate	Moderate	Moderate-High
Ensemble/Hybrid [5]	Multi-model aggregation	Low-Moderate	High	High
Attention-based CNN [11]	Lesion-focused attention maps	Moderate-High	High	High

V. DR Detection Applications

A. Referable vs. Non-Referable DR Detection

The majority of deployed automated screening systems are framed as binary classifiers distinguishing referable DR (moderate NPDR or worse, or the presence of diabetic macular edema) from non-referable disease, prioritizing high sensitivity to minimize missed referrals [1].

B. Multi-Class Severity Grading

Grading across the full five-point ICDR scale is more clinically informative but more challenging due to severe class imbalance — mild and moderate NPDR cases dominate most public datasets while proliferative DR cases are comparatively rare, leading models to underperform on minority severity classes [6].

VII. Preprocessing and Explainability Techniques

A. Image Preprocessing

Common preprocessing steps include contrast-limited adaptive histogram equalization (CLAHE), circular field-of-view cropping, illumination normalization, and green-channel extraction to emphasize vascular and lesion contrast, along with resizing and patch extraction for computational efficiency.

B. Data Augmentation

Given limited labeled data and severe class imbalance, augmentation strategies including rotation, flipping, color jitter, and synthetic oversampling of minority severity classes are widely used to improve model robustness and reduce overfitting to majority classes [4].

C. Lesion-Level Detection

Detecting individual lesions — microaneurysms, hemorrhages, hard and soft exudates — supports both severity grading and clinical interpretability, but microaneurysms in particular are small, low-contrast structures that remain difficult to detect reliably, especially in early-stage disease [7].

D. Diabetic Macular Edema Detection

DME, a vision-threatening complication that can occur at any DR severity level, is typically assessed using optical coherence tomography (OCT) in clinical practice; fundus-image-only models estimate DME risk indirectly via hard-exudate proximity to the macula, motivating multi-modal approaches that combine fundus and OCT imaging.

VI. Comparative Analysis

Table II compares representative studies by dataset, technique, and reported performance. Across studies, transfer-learning-based CNNs and ensembles consistently report high area-under-curve (AUC) values for referable DR detection, frequently approaching expert-level performance on internal validation sets [1],[8]. Performance on external or independent test sets collected from different populations or camera equipment is often markedly lower than on internal held-out test data, underscoring persistent generalization challenges [9].

C. Explainability via Attention and Saliency

To support clinical trust, several studies incorporate Grad-CAM or attention mechanisms that highlight image regions driving the model's grading decision, allowing clinicians to visually verify that predictions are driven by lesion-relevant regions rather than spurious artifacts [11].

D. Confidence Calibration

Probability calibration and uncertainty estimation techniques are increasingly incorporated so that automated systems can flag low-confidence cases for mandatory human review rather than issuing an automated grade, an important safety consideration for clinical deployment.

VIII. Datasets in DR Research

Table III summarizes major public and challenge datasets used in DR detection research, spanning the widely used EyePACS/Kaggle Diabetic Retinopathy dataset, APTOS 2019 Blindness Detection, Messidor and Messidor-2 [10], and IDRiD, which additionally provides pixel-level lesion annotations alongside severity grades [7].

TABLE II — Comparative Analysis of Representative Studies

Author & Year	Dataset	Technique	Task	Reported Performance
Gulshan et al. [1] (2016)	EyePACS, Messidor-2	Inception-v3 (transfer learning)	Referable DR detection	AUC ~0.99; high sensitivity/specificity
Pratt et al. [3] (2016)	Kaggle EyePACS	CNN (from scratch)	5-class grading	~75% accuracy; weak on mild DR
Gargeya & Leng [8] (2017)	EyePACS, Messidor-2	CNN + gradient boosting	Referable DR detection	AUC 0.94-0.97
Ting et al. [12] (2017)	Multi-ethnic Asian cohorts	Deep learning system	Referable DR + DME	High sensitivity/specificity across populations
Voets et al. [9] (2019)	EyePACS (replication)	Inception-v3 replication	Referable DR detection	Lower performance on external test set
Qummar et al. [5] (2019)	Kaggle EyePACS	5-model CNN ensemble	5-class grading	Improved kappa vs. single CNN
Porwal et al. [7] (2018)	IDRiD	Lesion + grading annotations	Lesion detection + grading	Benchmark for microaneurysm/lesion tasks

TABLE III — Key Datasets Used in DR Detection Research

Dataset	Annotation	Size	Classes	Access
EyePACS / Kaggle DR	Severity grade	~88,000 images	5-class	Public
APTOS 2019 [13]	Severity grade	~5,590 images	5-class	Public (Kaggle)
Messidor / Messidor-2 [10]	Severity grade, macular edema risk	~1,200 / ~1,748 images	4-class	Public
IDRiD [7]	Pixel-level lesion + grade	516 images	5-class + lesion masks	Public
DRIVE [14]	Vessel segmentation masks	40 images	N/A (segmentation)	Public

Dataset	Annotation	Size	Classes	Access
STARE [15]	Vessel segmentation masks	400 images	N/A (segmentation)	Public

IX. Key Challenges

A. Class Imbalance

Public DR datasets are heavily skewed toward "no DR" and mild cases, with severe and proliferative DR substantially underrepresented, causing models to underperform on rare but clinically critical severity classes unless mitigated through resampling, class-weighted losses, or synthetic augmentation [6].

B. Small-Lesion Detection

Microaneurysms, the earliest visible sign of DR, are small, low-contrast structures often only a few pixels wide in downsampled images, making their reliable detection a persistent technical bottleneck for early-stage screening [7].

C. Dataset Shift and Generalization

Models trained on one dataset, camera, or population frequently show reduced performance when evaluated on external datasets collected with different equipment or from different ethnic populations, highlighting the need for more diverse, multi-site training and validation data [9],[12].

D. Image Quality Variability

Gradable image quality is not guaranteed in real-world screening settings; poor illumination, blur, and artifacts degrade model performance and necessitate automated image-quality assessment as a pre-screening step before grading.

E. Clinical Trust and Regulatory Validation

Deployment of automated DR screening in clinical workflows requires prospective validation, explainability, and regulatory clearance; despite strong retrospective performance, prospective real-world deployment studies remain comparatively limited [12].

X. Future Research Directions

A. Multi-Modal Learning

Combining fundus photography with OCT, OCT-angiography, and electronic health record data such as HbA1c and diabetes duration may improve early detection of both DR and diabetic macular edema beyond what fundus images alone can capture.

B. Self-Supervised and Foundation Model Pretraining

Self-supervised pretraining on large unlabeled fundus image collections, followed by fine-tuning on smaller labeled DR datasets, is a promising direction to reduce dependence on scarce expert annotations, particularly for lesion-level tasks.

C. Federated and Privacy-Preserving Learning

Given that fundus image data is subject to patient privacy regulations and is often siloed across hospitals, federated learning approaches that train shared models without centralizing raw images are a natural extension for building more generalizable DR detection models across institutions.

D. Point-of-Care and Low-Resource Deployment

Lightweight, quantized model architectures deployable on smartphone-based fundus cameras represent an important direction for extending automated screening to low-resource and rural settings lacking specialist ophthalmic infrastructure.

XI. Conclusion

This review has examined the evolution of ML and DL approaches for early diabetic retinopathy detection, from classical hand-crafted-feature pipelines to CNN-based, transfer-learning, and ensemble methods. The evidence indicates that deep learning-based systems, particularly those leveraging transfer learning, can achieve referable-DR detection performance approaching expert ophthalmologist grading on internal test sets, though generalization to external populations and camera types remains an open challenge. Persistent issues including class imbalance, small-lesion detection, dataset shift, and the need for clinical explainability must be addressed before automated DR screening can be reliably deployed at scale. Future work in multi-modal learning, self-supervised pretraining, federated learning, and point-of-care deployment is expected to be central to translating ML-based DR detection into routine clinical screening practice, particularly in resource-constrained settings.

XII. References

[1] Gulshan, V., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402-2410.

[2] Acharya, U.R., et al. (2009). Automated diagnosis of diabetic retinopathy using digital fundus images. Journal of Medical Systems, 33(2), 121-127.

[3] Pratt, H., Coenen, F., Broadbent, D.M., Harding, S.P., and Zheng, Y. (2016). Convolutional neural networks for diabetic retinopathy. Procedia Computer Science, 90, 200-205.

[4] Wan, S., Liang, Y., and Zhang, Y. (2018). Deep convolutional neural networks for diabetic retinopathy detection by image classification. Computers & Electrical Engineering, 72, 274-282.

[5] Qummar, S., et al. (2019). A deep learning ensemble approach for diabetic retinopathy detection. IEEE Access, 7, 150530-150539.

[6] Islam, M.M., et al. (2020). Applications of machine learning in diagnosis of diabetic retinopathy: a review. Health Information Science and Systems, 8(1), 21.

[7] Porwal, P., et al. (2018). Indian Diabetic Retinopathy Image Dataset (IDRiD): a database for diabetic retinopathy screening research. Data, 3(3), 25.

[8] Gargeya, R., and Leng, T. (2017). Automated identification of diabetic retinopathy using deep learning. Ophthalmology, 124(7), 962-969.

[9] Voets, M., Mollersen, K., and Bongo, L.A. (2019). Reproduction study using public data of a deep learning algorithm for detection of diabetic retinopathy in fundus photographs. PLoS ONE, 14(6), e0217541.

[10] Decenciere, E., et al. (2014). Feedback on a publicly distributed image database: the Messidor database. Image Analysis & Stereology, 33(3), 231-234.

[11] Selvaraju, R.R., et al. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proc. IEEE ICCV, pp. 618-626.

[12] Ting, D.S.W., et al. (2017). Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA, 318(22), 2211-2223.

[13] APTOS 2019 Blindness Detection. (2019). Kaggle competition dataset, Asia Pacific Tele-Ophthalmology Society.

[14] Staal, J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., and van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. IEEE Trans. Medical Imaging, 23(4), 501-509.

[15] Hoover, A., Kouznetsova, V., and Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. Medical Imaging, 19(3), 203-210.

[16] Abramoff, M.D., et al. (2016). Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Investigative Ophthalmology & Visual Science, 57(13), 5200-5206.

ISSN: Applied For (2026) | DOI: 10.XXXXX/IJVAST.2026.XXXXX | Open Access — CC BY 4.0 License