Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00226-4

Original Article

Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning

Amol C. Adamuthe [1] , Shilpa R. Bhagat [2] , Vijay Kalmani [3,]

1. *Department of Information Technology, Kasegaon Education Society's Rajarambapu Institute of Technology, Shivaji University, Ishwarpur, IND*
2. *Department of Computer Science and Engineering, Kasegaon Education Society's Rajarambapu Institute of Technology, Shivaji University, Ishwarpur, IND*
3. *Department of Computer Science and Engineering, Rajarambapu Institute of Technology, Shivaji University, Ishwarpur, IND*

Received: June 13, 2026 | Review began: July 23, 2026 | Review ended: July 26, 2026 | Published: August 17, 2026

© **Copyright** 2026
This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Diabetic retinopathy (DR) is a leading cause of preventable blindness worldwide, necessitating accurate and reliable automated severity grading systems for early diagnosis and timely intervention. However, existing deep learning methods often exhibit limited lesion localization, poor cross-dataset generalization, inadequate modeling of the ordinal progression of disease severity, and limited prediction reliability. To address these challenges, this study proposes a lesion-guided domain-adaptive framework that integrates an EfficientNetV2-S backbone for retinal feature extraction, lesion-guided attention, adversarial domain adaptation, Earth Mover's Distance-based ordinal learning, Monte Carlo (MC) Dropout for uncertainty estimation, and Grad-CAM for model interpretability. Experiments were conducted on a harmonized dataset created by combining the APTOS2019 and MESSIDOR datasets, with the original labels consolidated into three severity classes: No DR, Mild DR, and Severe DR. The proposed framework achieved an accuracy of 90.26%, precision of 90.10%, recall of 90.26%, F1-score of 90.14%, and a Quadratic Weighted Kappa (QWK) score of 89.28%. Uncertainty-aware evaluation using MC Dropout achieved 89.89% accuracy, 88.99% QWK, and a mean predictive entropy of 0.2571. Comparative evaluations, backbone comparisons, and ablation studies further demonstrated the effectiveness of the proposed components. These results indicate that the proposed framework provides accurate, robust, interpretable, and reliable DR severity assessment across heterogeneous retinal image datasets.

Categories: Safe and trustworthy AI in healthcare, Computer Vision, Deep Learning

Keywords: diabetic retinopathy, deep learning, domain adaptation, ordinal learning, explainable ai (xai), uncertainty estimation

Introduction

Clinical importance of diabetic retinopathy

Diabetic retinopathy (DR) is among the most frequent microvascular complications of diabetes mellitus and remains a major cause of preventable vision impairment worldwide. The continuously increasing global diabetic population has led to a sharp rise in the number of people susceptible to retinal abnormalities, thus increasing the burden on healthcare systems and society [1,2]. Recent epidemiological reports over the last three decades have reported large increases in the prevalence of blindness and vision impairment from DR, particularly in working-age adults, which calls for effective screening and management strategies [1,3].

The clinical impact of DR goes beyond ophthalmology, as progression of retinal damage may result in irreparable visual impairment and subsequent effects on the quality of life, independence, work capabilities, and general health of diabetic patients [4]. Besides its health complications, DR entails substantial financial losses associated with regular screening, proper diagnostics, prolonged treatment, and vision rehabilitation costs. With advancing DR, the necessity of additional therapeutic procedures, including laser photocoagulation, intravitreal injections, and vitreoretinal surgeries, occurs and raises further health care expenditures [5].

Early detection of disease and prompt management play an important role in the prevention of potentially avoidable blindness caused by DR. Nevertheless, the existing screening programs are largely based on retinal fundus imaging analysis performed manually by experienced ophthalmologists and hence cannot be implemented widely enough to satisfy the growing needs due to limited medical staff availability. The increasing number of patients suffering from diabetes aggravates this problem further [6]. In this regard, the development of computerized artificial intelligence (AI) methods for retinal image analysis seems like a viable way to enhance the screening process and decrease diagnostic burden [6,7]. Automated DR grading systems allow for rapid and scalable analysis of retinal images, thus supporting population-level screening efforts and improving access to eye care, especially in resource-limited settings.

Challenges and limitations of automated DR grading

Despite the significant progress in deep learning for medical image analysis, reliable and clinically deployable systems for DR grading are still challenging to construct. The main challenge is the high heterogeneity of the retinal image datasets. The domain discrepancies that hinder model generalization [8,9] are often caused by the differences in imaging devices, acquisition protocols, illumination conditions, image quality, and patient demographics. Therefore, models trained on a specific dataset may suffer a significant performance drop when they are evaluated on images acquired from other clinical settings.

Another significant problem is the highly imbalanced distribution of DR data. Publicly available databases show a predominance of examples of either healthy eyes or initial stages of DR development. The presence of such an imbalanced dataset may introduce biases in learning algorithms toward classes with more examples, resulting in an inability to correctly detect advanced cases that require urgent attention from the doctor [10,11]. Furthermore, DR develops through different stages according to its progressively growing complexity. Multi-class classifications typically consider these stages as separate classes, ignoring their inherent ordinal nature and, hence, leading to potentially unreliable results [12].

The final obstacle also relates to the medical field. Despite impressive accuracy rates, deep neural networks remain difficult to interpret because their algorithms do not provide clear reasons for their decisions. This situation makes clinicians suspicious of regarding the output provided by the machine and prevents its practical use in clinical environments [13]. Moreover, most algorithms do not incorporate uncertainty estimates into their output predictions. Such an outcome is undesirable since cases with high levels of uncertainty must be highlighted to attract special attention from a physician as opposed to deterministic predictions [14]. These limitations collectively highlight the need for accurate, robust, interpretable, and trustworthy DR grading frameworks across a range of clinical settings.

Motivation and research objectives

These aforementioned challenges point to the need for more comprehensive and clinically oriented DR grading frameworks. Most existing methods focus on improving classification accuracy and pay relatively little attention to cross-dataset robustness, lesion-aware feature learning, ordinal disease progression, prediction uncertainty, and model explainability. Consequently, there remains a large gap between the experimental performance reported on benchmark datasets and the requirements of real-world clinical deployment.

Driven by these limitations, we propose a unified framework that integrates robust feature representation, domain-invariant learning, clinically meaningful severity modeling, and trustworthy decision support into a single architecture. To this end, an EfficientNetV2-S backbone is used to extract discriminative retinal features, and a lesion-guided attention mechanism is incorporated to highlight pathological regions associated with disease progression. We propose a gradient

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

reversal-based domain adaptation strategy that promotes domain-invariant feature learning to better generalize across heterogeneous datasets. Further, an Earth Mover's Distance (EMD)-based loss function is explicitly designed to model the ordinal nature of DR severity, which helps the network to better preserve the progression characteristics of the disease.

In addition to predictive performance, the study also emphasizes model transparency and reliability. To improve interpretability from a clinical perspective, Gradient-weighted Class Activation Mapping (Grad-CAM)-based visual explanations are used to gain insight into the retinal regions that influence the network's decisions. Further, Monte Carlo dropout is employed at inference time to estimate predictive uncertainty and identify ambiguous cases when specialist assessment may be required. The proposed framework integrates lesion-guided attention, domain adaptation, ordinal optimization, explainability, and uncertainty estimation to push automated DR grading toward a more robust and trustworthy clinical deployment. The main objectives of this study are to achieve the following:

1. To develop a lesion-guided deep learning framework that can highlight clinically relevant retinal abnormalities for DR grading.

2. To enhance cross-dataset generalization by domain-adaptive feature learning on heterogeneous retinal image datasets.

3. To incorporate ordinal severity modeling that preserves the progressive aspect of DR.

4. To improve model transparency and reliability via Grad-CAM-based explainability and Monte Carlo dropout-based uncertainty estimation.

5. To perform an extensive evaluation of the proposed framework, including quantitative performance analysis, ablation studies, and comparison experiments of backbones on the combined APTOS2019 and MESSIDOR datasets.

Deep convolutional approaches for retinal severity assessment

Convolutional Neural Network (CNN) models have become a prominent tool for automatic grading of DR due to their capacity to learn hierarchical representations of retinal structure automatically from fundus images. Different CNN models have been developed and examined in recent years, such as ResNet, DenseNet, InceptionNet, NASNet, and hybrid models, to improve the severity classification [15,16]. Multi-stream ensemble models that utilize different backbones and complementary representations of features such as ResNet-50 and DenseNet-121 [17] yielded impressive results in terms of classification on EyePACS and Messidor datasets. Similarly, hybrid CNN and recurrent neural network (RNN) models with an attention mechanism have been applied to consider both spatial features and temporal trends of the disease and yielded competitive results on multiple databases such as DRIVE, Kaggle, and EyePACS [18]. The use of transfer learning approaches with different CNN models, NASNet-Large and ResNet-50, improved classification accuracy while reducing complexity with the aid of data augmentation [19]. It is concluded that more complex models and advanced fusion of features may lead to better results in automated DR grading.

Although CNN models have achieved successful results, they still face some challenges. Attention-based frameworks like AG-CNN have enhanced the accuracy of lesion localization and classification by simultaneously learning the global and local retinal representations [20]. In addition, hybrid architectures such as combining InceptionV3, ResNet-50, and other feature extraction networks have shown high sensitivity and specificity with high feature diversity [21]. Many of the existing methods, however, require heavy preprocessing, class balancing techniques, or computationally expensive ensemble learning frameworks [22,23]. Moreover, model performance often degrades when tested on external datasets due to domain shifts arising from differences in imaging devices, acquisition conditions, and patient populations. Further, some deep convolutional approaches for retinal severity assessment systems have limited clinical applicability because of their limited interpretability, poor uncertainty estimation, and poor handling of ordinal disease progression. Table *1* summarizes the main characteristics of representative CNN-based approaches for DR grading.

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

Method	Backbone	Dataset	Accuracy (%)	Key Observation
Multi-stream Ensemble [17]	ResNet-50 + DenseNet-121	EyePACS, Messidor-2	95.58	Ensemble feature learning improves robustness and classification performance.
Hybrid CNN-RNN [18]	CNN + RNN + Attention	DRIVE, Kaggle, EyePACS	94.04–97.50	Captures spatial and progression-related retinal features.
Hybrid CNN-DenseNet [24]	CNN + DenseNet	Private Dataset	96.22	Dense feature reuse improves multi-stage DR grading.
AG-CNN [20]	DenseNet-121 + Attention Branch	APTOS 2019	98.48	Lesion-focused attention enhances discriminative feature learning.
Lightweight CNN [22]	Custom CNN	EyePACS	90.60	Image preprocessing improves performance with lower computational complexity.
Transfer Learning Framework [19]	NASNet-Large, ResNet-50	APTOS 2019	>96.00	Pretrained backbones improve convergence and classification accuracy.
Feature Fusion CNN [21]	InceptionV3 + ResNet50	Public Fundus Datasets	96.85	Multi-backbone fusion improves sensitivity and specificity.
CNN-RBF Hybrid [23]	CNN + RBF Classifier	Multiple Public Datasets	96.10	Combines handcrafted and deep features for improved detection.

TABLE 1: Representative CNN-based diabetic retinopathy grading approaches and their characteristics

CNN, Convolutional Neural Network; RNN, Recurrent Neural Network; DR, Diabetic Retinopathy; RBF, Radial Basis Function

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

Attention-driven retinal representation learning

Attention mechanisms are increasingly adopted in modern DR grading systems, since they allow deep learning models to focus selectively on clinically relevant retinal structures and to mitigate redundant background information. Early attention-based approaches predominantly employed channel and spatial attention modules, e.g., the Convolutional Block Attention Module (CBAM), to enhance feature representation by concentrating on informative feature channels and lesion-specific spatial regions [9]. These mechanisms have been shown to be effective for the improvement of retinal vessel segmentation and lesion detection by guiding the network toward pathological features associated with DR progression [25,26]. Recently, self-attention and transformer-based architectures have gained popularity because they are able to model long-range dependencies and global contextual relationships that are often ignored by the traditional convolutional operations. Vision Transformer (ViT)-based frameworks and hybrid transformer models have demonstrated promising performance in modeling complex retinal patterns and inter-lesion relationships. This is achieved through global feature aggregation and multi-scale attention learning [27,28]. Moreover, cross-domain attention strategies that fuse complementary feature representations from different domains have further boosted discriminative learning and grading performance [29].

Despite these advances, there remain several limitations that affect the effectiveness of attention-driven retinal representation learning. Transformer-based approaches usually need large training datasets and high computational resources, which makes their deployment difficult in resource-constrained clinical settings [27,28]. Moreover, the quadratic complexity of self-attention mechanisms can be a bottleneck for high-resolution retinal images. To tackle these problems, lesion-focused attention frameworks have been proposed to explicitly localize pathological regions and guide the feature extraction toward clinically meaningful abnormalities such as microaneurysms, hemorrhages, and exudates [30]. These approaches can boost interpretability and classification accuracy by combining lesion information with the overall retinal environment. However, they are sensitive to the localization of lesions in the retina and can suffer negative effects from incorrect lesion localization. Therefore, despite the success of the attention mechanisms in achieving improved performance for learning features and diagnosis tasks, finding an appropriate balance between awareness of pathological changes, efficiency, interpretation of results, and generalizability of models across different databases is still an open problem for automated DR grading.

Domain adaptation approaches

Domain adaptation has become another promising direction in DR grading owing to differences in various retinal image databases. The adverse effect of domain shifts on generalization capability due to differences in hardware characteristics, data acquisition procedures, quality, and demographics has prompted numerous investigations of adversarial learning and feature alignment strategies to learn invariant features. For instance, Graph Adversarial Transfer Learning is based on adversarial training using graph neural networks for inter-domain and intra-domain alignments. Thus, the approach allows robust feature extraction and does not require a substantial amount of annotated data from the target domain [31]. Furthermore, several works focused on exploring multi-source domain adaptation techniques to address distribution shifts in multiple databases. In particular, the researchers proposed a novel DRStageNet that combines multi-source domain adaptation and self-supervised learning with vision transformers for improving prediction performance in different testing domains [32].

More recently, self-supervised learning (SSL) emerged as an alternative strategy to address domain variability and the scarcity of annotated data. SSL methods exploit the availability of large sets of unlabeled fundus images to learn transferable retinal representations that generalize well across datasets. The Saliency-guided Self-supervised Image Transformer uses saliency-aware contrastive learning and auxiliary segmentation objectives to guide representation learning toward clinically relevant retinal structures, leading to substantial improvements in downstream DR grading performance [33]. At the same time, multi-domain and multi-source training approaches have further shown the advantages of training a model on diverse imaging conditions. For example, it has been shown that models trained on combinations of EyePACS, DRIVE, Kaggle, and other public datasets achieve improved robustness compared to single-domain models [18,32]. However, problems of training complexity, computational cost, and complete domain invariance

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

are still unsolved, although promising results have been achieved by adversarial learning, self-supervised pre-training, and multi-domain optimization. Thus, building efficient domain-adaptive frameworks that can consistently perform across heterogeneous clinical environments is an active research area in automated DR grading.

Ordinal learning methods

DR severity grading is a natural ordinal classification problem where the progression of the disease has a natural continuum from no DR to proliferative DR. However, many existing deep learning models formulate DR grading as a standard multi-class classification problem, ignoring the ordinal relationships between disease stages. To overcome this limitation, researchers have explored ordinal learning, ordinal regression, and ranking-based optimization methods that explicitly incorporate the severity ordering into the learning process. These approaches attempt to penalize predictions based on the distance between predicted and true severity grades, making the predictions more clinically meaningful. In ranking-based frameworks, contrastive grading losses have also been studied to retain the relative severity relationship among retinal images, which have resulted in improved discrimination between the neighboring DR stages and improved multi-class grading performance [34]. More generally, the significance of ordinal learning paradigms is acknowledged as an important direction to improve the clinical relevance and consistency of automated DR grading systems [14].

Besides ordinal regression and ranking-based learning, severity-aware optimization methods have gained increasing attention for medical image classification. Metrics like Quadratic Weighted Kappa (QWK) and weighted loss functions commonly emphasize clinically relevant grading errors and improve agreement with expert annotations. For example, DenseNet-CBAM-based frameworks optimized by QWK achieved better grading performance by considering the ordinal structure of the disease severity [35]. Likewise, ensemble learning methods with severity-aware optimization have demonstrated less label discrepancy and better robustness for DR classification tasks [36]. The EMD loss has been proposed as an effective way to model ordinal relationships by aligning severity across distributions, but its use in DR grading has been limited by the additional computational complexity and implementation challenges. Therefore, while ordinal learning is gaining recognition as a clinically meaningful alternative to conventional classification, standardized ordinal frameworks, large-scale validation studies, and computationally efficient optimization strategies remain active research challenges in automated DR grading.

Explainable and uncertainty-aware AI

The rising application of AI in medicine has raised awareness of the need for explainability and uncertainty quantification, which enhance the reliability of automated systems for identifying DR. Explainable AI seeks to provide an explanation of the rationale behind decisions made by neural networks in terms of the identification of lesions or retinal layers that play the most significant role in the corresponding prediction. In this regard, Grad-CAM is one of the most popular approaches used to obtain explanations for a particular decision in the form of heat maps, which show areas of lesions and retinal layers that impact a predicted score for the specific DR severity grade [37,38]. Additionally, SHapley Additive exPlanations represents another method aimed at interpreting model decisions in terms of feature importance using game theoretic approaches [39,40]. While saliency-based methods may help with visualizing relevant regions within a lesion, they have shown different localization accuracy due to varying lesion features, network structure, and imaging modality [41].

In addition to explainability, confidence awareness in the predictions made by the models has emerged as an essential requirement for implementing AI systems in critical healthcare applications. Bayesian deep learning algorithms like Monte Carlo (MC) dropout and ensembles may produce probability values instead of point estimates for the purpose of measuring uncertainty of prediction [42,43]. This will help identify those ambiguous instances that require expert supervision and avoid the risks of misclassification by an AI system with too much certainty about the predicted class label. The confidence level of the classifier can be determined by calculating measures of entropy [42,43]. While improvements made in explainability and uncertainty assessment contributed to better performance of DR screening systems, some of their limitations include visual inconsistency of the explanation results, high costs associated with

Bayesian inference, and lack of a standard evaluation protocol. However, including confidence-based decision-making alongside interpretable models is another important step toward developing clinically viable AI solutions for DR diagnosis.

Limitations of current retinal grading frameworks

Numerous studies have focused on automated DR grading using CNNs, attention models, domain adaptation techniques, ordinal regression approaches, and trustworthy AI models, but many aspects still prevent the clinical implementation of these solutions. Nonetheless, even high-performing CNN-based approaches suffer from data bias, inability to generalize across different imaging environments, and limited interpretability of the results [14,44]. Attention-based methods provide better lesion localization and feature visualization, but they involve additional computational complexity and require large amounts of labeled images for efficient training [29,44]. Adversarial learning, feature alignment, and multisource domain adaptation techniques significantly increased generalization, but residual differences between domains, training instabilities, and the necessity of precise and well-balanced training datasets remain to be tackled [31-33]. Besides, even though the idea of ordinal regression and severity awareness in learning proved to be efficient, there is no standardized framework allowing the incorporation of ordinal relationships into DR grading models [14,36].

Apart from the aforementioned challenges related to model performance, current approaches do not ensure necessary transparency and reliability. The problem of explainability can be addressed by methods such as Grad-CAM or saliency-based visualization, but they cannot always localize retinal lesions correctly and yield visually inconsistent explanations due to differences between the architectures [38,41]. In the same way, the Bayesian approach and MC dropout methods are useful to evaluate uncertainties, yet they are rarely used due to their high computational complexity and lack of interpretability for uncertainty scores [42]. Table *2* presents an analysis of all identified limitations for each direction of studies. Based on these findings, future DR classification systems should incorporate properties such as cross-domain transferability, lesion-based feature extraction, the ordinal nature of disease progression, explainability, and uncertainty estimation all at once.

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

Research Direction	Major Limitations
CNN-Based Methods	Limited cross-dataset generalization, susceptibility to dataset bias and class imbalance, inadequate focus on subtle lesions, and limited interpretability despite strong classification performance [14,21,44].
Attention-Based Methods	High computational complexity, dependency on large annotated datasets, and difficulty in consistently capturing subtle lesion characteristics and long-range pathological relationships [29,45].
Domain Adaptation Methods	Residual domain shifts, instability of adversarial training, increased optimization complexity, and incomplete utilization of unlabeled retinal data [31-33].
Severity-Aware and Ordinal Modeling Techniques	Limited adoption in DR grading, lack of standardized ordinal learning frameworks, and additional optimization overhead associated with ranking-based and severity-aware loss functions [14,36].
Explainability Methods	Inconsistent visual explanations, limited lesion localization accuracy, and lack of standardized quantitative evaluation protocols for explanation quality [38,41].
Uncertainty-Aware Methods	Increased inference time, limited integration into clinical decision support systems, and challenges in interpreting uncertainty measures for routine clinical use [42].

TABLE 2: Summary of limitations identified in existing diabetic retinopathy grading approaches

CNN, Convolutional Neural Network; DR, Diabetic Retinopathy

Materials And Methods

This section describes the datasets, preprocessing steps, and data preparation techniques used to develop the proposed DR grading framework. Two publicly available benchmark datasets, APTOS2019 and MESSIDOR [46], were integrated to create a diverse and representative dataset. Subsequently, label harmonization, stratified data partitioning, and image augmentation techniques were applied to improve class balance, enhance model robustness, and reduce overfitting. The experiments were conducted using a preprocessed retinal fundus image dataset obtained from the Kaggle repository "Diabetic Retinopathy 224 × 224 (2019 Data)," which combines images from the APTOS 2019 Blindness Detection Challenge and the MESSIDOR database.

Datasets and preprocessing

Label Harmonization Approach

Although both datasets provide DR severity annotations, differences in dataset characteristics and severity label representation necessitate a unified labeling strategy prior to dataset integration. Therefore, the original five DR severity grades were consolidated into three clinically meaningful classes to standardize the severity scale across both datasets and facilitate consistent cross-domain learning, as summarized in Table 3.

Original Grade	Mapped Class
0	No DR
1,2	Mild DR
3,4	Severe DR

TABLE 3: Label harmonization approach used for DR classification

DR, Diabetic Retinopathy

Dataset	Number of Images
APTOS2019	**3662**
MESSIDOR	1748
Total	**5410**

TABLE 4: Composition of the integrated retinal fundus image dataset

The integrated dataset consisted of 5,410 retinal fundus images collected from the APTOS2019 and MESSIDOR datasets. Table 4 summarizes the composition of the integrated dataset prior to model development.

Dataset Integration and Split Methodology

The retinal images from the two datasets were integrated to form a unified dataset, defined in Equation (1).

$$\mathcal{D} = \{(x_i, y_i, d_i)\}_{i=1}^{N} \tag{1}$$

where x_i denotes the i^{th} retinal image, y_i represents the corresponding DR severity label, d_i denotes the domain label indicating the source dataset, and N is the total number of retinal images in the integrated dataset.

For domain adaptation, images from the APTOS2019 dataset were assigned to Domain 0, whereas images from the MESSIDOR dataset were assigned to Domain 1. The integrated dataset was subsequently divided into training, validation, and testing subsets, as illustrated in Equation (2).

$$\mathcal{D} = \mathcal{D}_{train} \cup \mathcal{D}_{val} \cup \mathcal{D}_{test} \tag{2}$$

where $\mathcal{D}_{train}$, $\mathcal{D}_{val}$, and $\mathcal{D}_{test}$ represent the training, validation, and testing sets, respectively.

A stratified sampling strategy was adopted to preserve the class distribution across all subsets. The dataset partitioning is expressed in Equation (3).

$$\mathcal{D} = 0.70\mathcal{D}_{train} + 0.15\mathcal{D}_{val} + 0.15\mathcal{D}_{test} \quad (3)$$

According to Equation (3), the integrated dataset was partitioned into training, validation, and testing subsets using a stratified sampling strategy to preserve the class distribution across all partitions. The resulting dataset partition is summarized in Table 5. The training subset was used for model optimization, the validation subset for hyperparameter tuning and model selection, and the testing subset was reserved for final performance evaluation.

Dataset Split	Number of Images	Percentage
Training	3787	70%
Validation	812	15%
Testing	811	15%
Total	5410	100%

TABLE 5: Stratified partitioning of the integrated retinal fundus image dataset

Data Augmentation and Image Enhancement

To improve model generalization and reduce overfitting, geometric and photometric augmentation techniques, together with contrast enhancement, were applied exclusively during model training. All preprocessing and augmentation operations were implemented using the Albumentations library. Input retinal fundus images were resized to 384 × 384 pixels prior to augmentation. Validation and testing datasets followed a deterministic preprocessing pipeline without stochastic augmentation. The augmentation and enhancement hyperparameters used in this study are summarized in Table 6.

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

Stage	Transform	Parameters	Probability
Train	Resize	384 × 384	Always
Train	Horizontal Flip	—	0.5
Train	Vertical Flip	—	0.2
Train	Rotation	±20°	0.5
Train	Brightness & Contrast	Brightness_limit = 0.2, contrast_limit = 0.2	0.5
Train	CLAHE	clip limit = 4.0, tile grid size = 8 × 8	0.5
Train	Gaussian Noise	Albumentations default variance range	0.2
Train/Validation/Test	Normalize	ImageNet μ = (0.485,0.456,0.406), σ = (0.229,0.224,0.225)	Always
Validation/Test	Resize	384 × 384	Always
Validation/Test	Random Augmentation	Disabled	—

TABLE 6: Data augmentation and image enhancement hyperparameters used during model training

The augmentation procedure is mathematically defined by Equation (4).

$$I' = T(I) \tag{4}$$

where I is the original retinal image, I' is the transformed image, and $T(\cdot)$ denotes a random transformation operator comprising the applied augmentation techniques.

For image enhancement, CLAHE was employed using the parameters listed in Table *6* to improve local contrast and enhance the visibility of retinal lesions. Its operation is defined by Equation (5).

$$I_{\text{CLAHE}} = f_{\text{CLAHE}}(I) \tag{5}$$

where $f_{\text{CLAHE}}(\cdot)$ denotes the adaptive histogram equalization operation. CLAHE enhances low-contrast retinal regions, thereby improving the visibility of clinically relevant features such as microaneurysms, hemorrhages, and exudates.

The combined processes of label harmonization, stratified data partitioning, image augmentation, and image enhancement constitute the preprocessing pipeline used for developing the proposed DR grading framework.

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

Proposed lesion-guided domain-adaptive framework

The proposed lesion-guided domain-adaptive framework for DR grading is illustrated in Figure *1*. The framework integrates retinal image preprocessing, deep feature extraction, lesion-guided attention, adversarial domain adaptation, ordinal severity learning, uncertainty estimation, and explainable AI into a unified architecture.

Retinal fundus images from the APTOS2019 and MESSIDOR datasets are first preprocessed through label harmonization, CLAHE-based image enhancement, and data augmentation. The processed images are then provided as input to the EfficientNetV2-S backbone network to extract hierarchical retinal feature representations.

The extracted feature maps are refined using a lesion-guided attention module comprising channel and spatial attention mechanisms. The resulting lesion-aware feature representations are subsequently forwarded to the domain-adaptive ordinal learning module.

The domain-adaptive ordinal learning module performs DR severity classification while simultaneously learning domain-invariant feature representations. An ordinal classification branch predicts the disease severity, whereas a Gradient Reversal Layer (GRL) and a domain discriminator minimize domain discrepancies between the APTOS2019 and MESSIDOR datasets through adversarial learning.

During inference, the optimized feature representation is used to predict the DR severity grade. Predictive uncertainty is estimated using MC Dropout with multiple stochastic forward passes, while Grad-CAM generates visual explanations by highlighting retinal regions contributing to the model predictions.

The framework produces three outputs: the predicted DR severity grade, a confidence score derived from uncertainty estimation, and a Grad-CAM heatmap highlighting the retinal regions contributing to the prediction.

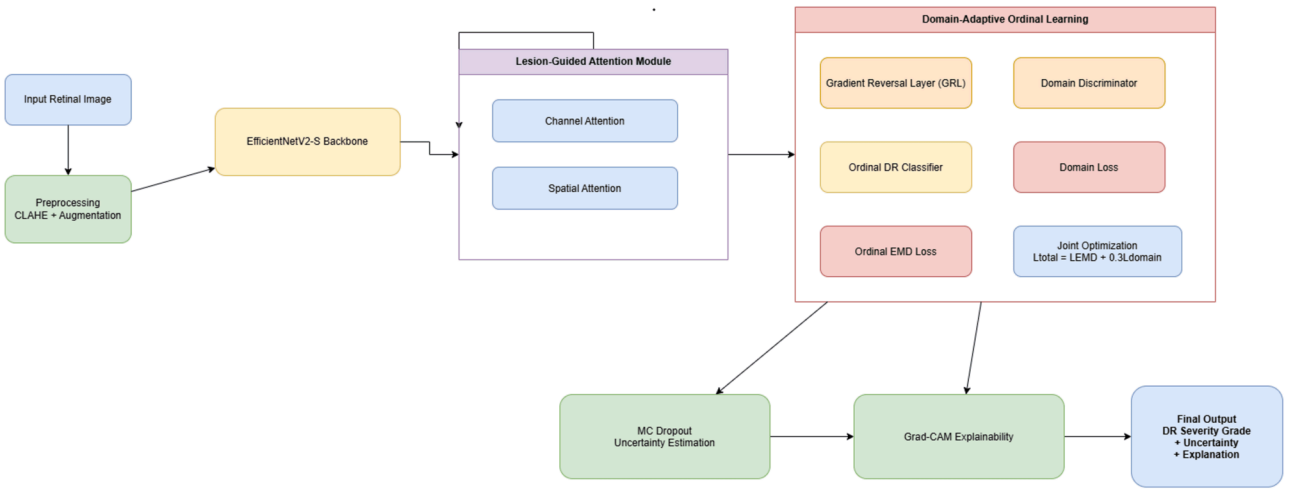


FIGURE 1: Overview of the proposed retinal severity assessment pipeline integrating lesion-focused attention, adversarial feature alignment, ordinal optimization, uncertainty estimation, and explainability modules

DR, Diabetic Retinopathy; Grad-CAM, Gradient-weighted Class Activation Mapping; EMD, Earth Mover's Distance

The overall workflow of the proposed framework consists of the following five stages:

1. Deep feature extraction using the EfficientNetV2-S backbone network.

2. Lesion-guided feature refinement through channel and spatial attention mechanisms.

3. Adversarial domain adaptation using a gradient reversal layer and domain discriminator.

4. Ordinal DR severity classification.

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

5. Uncertainty estimation and explainability using MC Dropout and Grad-CAM.

The complete training procedure of the proposed framework is summarized in Algorithm (1).

Algorithm 1: Training Procedure of the Proposed Lesion-Guided Domain-Adaptive Framework

Input: Integrated retinal dataset $(\mathcal{D})$, severity labels (y), domain labels (d)

Output: Optimized DR grading model

Initialize EfficientNetV2-S backbone network.

Initialize lesion-guided attention module.

Initialize GRL.

Initialize domain discriminator.

Initialize ordinal classification head.

For each training epoch:

a. For each mini-batch $((x, y, d))$:

i. Extract retinal features using the EfficientNetV2-S backbone.

ii. Refine lesion-aware feature representations through the lesion-guided attention module.

iii. Apply adversarial domain adaptation using the GRL.

iv. Predict domain labels using the domain discriminator.

v. Predict DR severity grade using the ordinal classification head.

vi. Compute ordinal classification loss: (L_{EMD})

vii. Compute domain adaptation loss: (L_{domain})

viii. Compute total optimization loss: $(L_{total} = L_{EMD} + \lambda L_{domain})$

ix. Update network parameters through backpropagation.

End mini-batch loop.

End training epoch loop.

Return the optimized DR grading model.

Algorithm (1) summarizes the end-to-end optimization process of the proposed framework. The model jointly learns discriminative retinal representations for DR severity grading and domain-invariant features for cross-dataset generalization. During inference, the optimized network generates severity predictions together with confidence estimates and visual explanations, thereby providing a reliable and interpretable decision-support system for automated DR screening.

Lesion-guided feature learning module

The proposed framework incorporates a lesion-guided feature learning module that combines EfficientNetV2-S with channel and spatial attention mechanisms to generate lesion-aware retinal feature representations.

EfficientNetV2-S Feature Extraction

EfficientNetV2-S is employed as the backbone network for extracting hierarchical retinal representations. Given an input retinal image x, the backbone network generates a deep feature representation according to Equation (6).

$$F = \Phi(x) \tag{6}$$

where $\Phi(\cdot)$ denotes the EfficientNetV2-S backbone network and $F \in \mathbb{R}^{C \times H \times W}$ represents the extracted feature tensor, where C, H, and W denote the channel dimension, height, and width, respectively.

The extracted feature maps are subsequently refined using channel and spatial attention mechanisms.

Channel Attention Mechanism

The channel attention mechanism recalibrates feature channels by assigning higher weights to informative retinal representations. Given the feature tensor F, global average pooling and global max pooling are performed independently across the spatial dimensions. The resulting descriptors are processed through a shared multilayer perceptron (MLP) to generate the channel attention map, as defined in Equation (7).

$$M_c(F) = \sigma\Big(MLP(AvgPool(F)) + MLP(MaxPool(F))\Big) \tag{7}$$

where $M_c(F)$ denotes the channel attention map, $AvgPool(\cdot)$ and $MaxPool(\cdot)$ denote global average and max pooling operations, respectively, $MLP(\cdot)$ denotes a shared multilayer perceptron, and $\sigma(\cdot)$ represents the sigmoid activation function.

The computed attention map is used to recalibrate the feature channels according to Equation (8).

$$F_c = M_c(F) \otimes F \tag{8}$$

where $\otimes$ denotes element-wise multiplication and F_c represents the channel-refined feature representation.

Spatial Attention Mechanism

The spatial attention mechanism further refines the extracted feature maps by emphasizing lesion-relevant spatial regions. The spatial attention map is generated by applying average pooling and max pooling across the channel dimension, followed by channel-wise concatenation and convolution, as expressed in Equation (9).

$$M_s(F) = \sigma\Big(f^{7\times7}\big[Avg(F); Max(F)\big]\Big) \tag{9}$$

where $M_s(F)$ denotes the spatial attention map, $f^{7\times7}$ represents a convolution operation with a kernel size of 7×7, and $[\cdot;\cdot]$ denotes channel-wise concatenation.

The spatial attention map is subsequently used to refine the feature representation according to Equation (10).

$$F_s = M_s(F) \otimes F \tag{10}$$

where F_s denotes the spatially refined feature representation.

Lesion-Guided Feature Representation

The channel-refined and spatially refined feature maps are sequentially combined to generate the final lesion-guided representation, expressed as Equation (11).

$$F_{LG} = M_s(F_c) \otimes F_c \tag{11}$$

where F_{LG} denotes the final lesion-guided feature representation that is subsequently forwarded to the domain adaptation and ordinal classification modules.

Domain-adaptive ordinal learning

The proposed framework incorporates adversarial domain adaptation and ordinal severity learning to obtain domain-invariant retinal representations while preserving the ordinal relationship among DR severity grades.

Adversarial Domain Adaptation

The objective of domain adaptation is to learn retinal representations that are discriminative for DR grading while remaining invariant to dataset-specific characteristics. Let F_{LG} denote the lesion-guided feature representation obtained from Equation (11). A GRL is introduced before the domain discriminator to encourage domain-invariant feature learning.

The GRL performs an identity transformation during forward propagation and reverses the gradient direction during backpropagation. The forward operation is defined by Equation (12).

$$R(F_{LG}) = F_{LG} \tag{12}$$

where $R(\cdot)$ denotes GRL.

During backpropagation, the gradient transformation is given by Equation (13).

$$\frac{\partial R}{\partial F_{LG}} = -\alpha I \tag{13}$$

where α denotes the gradient reversal coefficient and I represents the identity matrix.

The transformed feature representations are subsequently forwarded to the domain discriminator, which predicts the source domain associated with each retinal image. The domain prediction process is expressed by Equation (14).

$$\hat{d} = D\big(R(F_{LG})\big) \tag{14}$$

where $D(\cdot)$ denotes the domain discriminator network and $\hat{d}$ represents the predicted domain label.

The domain adaptation objective is optimized using the cross-entropy loss defined in Equation (15).

$$L_{domain} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} d_{ij}\log(\hat{d}_{ij}) \tag{15}$$

where N denotes the number of training samples, M denotes the number of domains, d_{ij} represents the ground-truth domain label, and $\hat{d}_{ij}$ denotes the predicted domain probability.

Ordinal Severity Learning Using Earth Mover's Distance

An EMD-based ordinal loss is employed to preserve the ordered relationship among DR severity grades.

Let the predicted class probability distribution be represented by

$$P = [p_1, p_2, \ldots, p_K] \tag{16}$$

where K denotes the number of severity classes and p_k represents the predicted probability of class k.

The cumulative distribution function of the predicted probabilities is computed according to Equation (17).

$$CDF_P(k) = \sum_{i=1}^{k} p_i \tag{17}$$

where $CDF_P(k)$ denotes the cumulative predicted probability up to class k.

Similarly, the cumulative distribution corresponding to the ground-truth label is computed as

$$CDF_T(k) = \sum_{i=1}^{k} t_i] \tag{18}$$

where t_i denotes the one-hot encoded ground-truth distribution.

The ordinal EMD loss is calculated according to Equation (19).

$$L_{EMD} = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}\left(CDF_P^{(n)}(k) - CDF_T^{(n)}(k)\right)^2 \quad (19)$$

where N denotes the batch size.

Joint Optimization Objective

The proposed framework jointly optimizes the ordinal classification objective and the domain adaptation objective. The overall optimization function is defined by Equation (20).

$$L_{total} = L_{EMD} + 0.3\,L_{domain} \quad (20)$$

where the weighting coefficient 0.3 corresponds to the domain adaptation factor adopted during training.

Algorithm 2 summarizes the proposed domain-adaptive ordinal learning procedure. The lesion-guided feature representations are simultaneously optimized using the EMD-based ordinal classification loss and the adversarial domain adaptation loss. The overall network parameters are updated using the joint optimization objective defined in Equation (20).

Algorithm 2: Domain-Adaptive Ordinal Learning

Input: Lesion-guided feature representation (F_{LG}) , severity labels (y), domain labels (d)

Output: Domain-invariant ordinal DR grading model

For each training mini-batch:

a. Extract lesion-guided feature representations (F_{LG}).

b. Predict diabetic retinopathy severity grades using the ordinal classification head $[\hat{y} = C(F_{LG})]$.

c. Compute the ordinal EMD loss $[L_{EMD}]$.

d. Apply the GRL to obtain domain-adaptive feature representations $[F_{GRL} = R(F_{LG})]$.

e. Predict the source domain using the domain discriminator $[\hat{d} = D(F_{GRL})]$.

f. Compute the domain classification loss $[L_{domain}]$.

g. Compute the total optimization loss $[L_{total} = L_{EMD} + 0.3L_{domain}]$.

h. Update the parameters of the feature extractor, lesion-guided attention module, ordinal classifier, and domain discriminator using backpropagation.

End mini-batch loop.

Return the optimized domain-adaptive ordinal DR grading model.

Explainability, uncertainty estimation, and complexity analysis

To improve clinical trustworthiness and reliability, the proposed framework incorporates mechanisms for explainability and uncertainty estimation. In addition, computational complexity analysis is performed to evaluate the practicality of deploying the proposed model in real-world DR screening systems.

Grad-CAM-Based Explainability

Grad-CAM is employed to identify retinal regions contributing to DR grading decisions.

Let A^k denote the k^{th} feature map extracted from the lesion-guided attention module. The importance weight associated with feature map A^k is computed according to Equation (21).

$$\alpha_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{21}$$

where y^c denotes the prediction score of class c and Z represents the total number of spatial locations.

The Grad-CAM localization map is generated according to Equation (22).

$$L_{GradCAM}^c = ReLU\left(\sum_k \alpha_k A^k \right) \tag{22}$$

where $L_{GradCAM}^c$ highlights the retinal regions contributing to the predicted DR severity grade.

MC Dropout-Based Uncertainty Estimation

MC dropout is employed during inference to estimate predictive uncertainty by performing multiple stochastic forward passes while maintaining dropout activation.

Let T denote the total number of stochastic forward passes. The predictive probability distribution is computed according to Equation (23).

$$\bar{P} = \frac{1}{T} \sum_{t=1}^{T} P_t \tag{23}$$

where P_t denotes the prediction probability obtained during the t^{th} stochastic forward pass.

The predictive uncertainty is quantified using entropy, as defined in Equation (24).

$$H(\bar{P}) = -\sum_{k=1}^{K} \bar{P}_k \log(\bar{P}_k) \tag{24}$$

where $\bar{P}_k$ denotes the mean probability of class k.

In the present study, $T = 20$ stochastic forward passes were used during MC dropout inference to estimate predictive uncertainty.

Computational Complexity Analysis

The computational complexity of the proposed framework is primarily determined by the EfficientNetV2-S backbone, lesion-guided attention module, domain discriminator, and ordinal classification head.

The overall computational complexity can be approximated by Equation (25).

$$\mathcal{O}_{total} = \mathcal{O}_{backbone} + \mathcal{O}_{attention} + \mathcal{O}_{domain} + \mathcal{O}_{classifier} \tag{25}$$

where $\mathcal{O}_{backbone}$, $\mathcal{O}_{attention}$, $\mathcal{O}_{domain}$, and $\mathcal{O}_{classifier}$ represent the computational costs of the corresponding modules.

The total number of trainable parameters is computed as

$$P_{total} = \sum_{i=1}^{L} P_i \tag{26}$$

where P_i denotes the trainable parameters of the $i\text{-th}$ layer and L represents the total number of learnable layers.

Experimental evaluation demonstrates that the proposed framework achieves an effective balance between classification performance and computational complexity. The incorporation of lesion-guided attention, adversarial domain adaptation, ordinal learning, Grad-CAM explainability, and MC Dropout uncertainty estimation introduces only a modest computational overhead while providing substantial improvements in robustness, interpretability, and clinical reliability.

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

Results And Discussion

Experimental setup

This section describes the experimental environment, training configuration, and evaluation metrics employed to validate the effectiveness of the proposed lesion-guided domain-adaptive framework for DR grading. All experiments were conducted under identical training conditions to ensure a fair and reproducible evaluation of the proposed methodology and comparative models. All reported experimental results were obtained using the independent testing subset comprising 811 retinal fundus images, corresponding to the 15% testing partition of the integrated dataset summarized in Table 5.

Hardware Configuration and Software Environment

The proposed framework was implemented and evaluated using the Kaggle cloud-based computational platform equipped with two NVIDIA Tesla T4 GPUs, each providing 16 GB of dedicated GDDR6 memory, to accelerate deep learning computations. The software stack included Python 3.x, PyTorch, OpenCV, Albumentations, timm, and Scikit-learn. GPU acceleration was utilized throughout feature extraction, lesion-guided attention learning, domain adaptation, uncertainty estimation, and model optimization, while the Kaggle runtime provided multi-core Intel Xeon processors and approximately 30 GB of system RAM to support efficient data preprocessing and model execution.

Training Configuration

To ensure stable convergence and a fair comparison across all experiments, a unified set of training hyperparameters was adopted for the proposed framework and all comparative models. Network optimization was performed using the AdamW optimizer with an initial learning rate of $\mathbf{3 \times 10^{-4}}$weight decay of $\mathbf{1 \times 10^{-4}}$, momentum coefficients $\boldsymbol{\beta_1 = 0.9}$ and $\boldsymbol{\beta_2 = 0.999}$, and a numerical stability parameter of $\boldsymbol{\varepsilon = 1 \times 10^{-8}}$. A cosine annealing learning-rate scheduler was employed to gradually reduce the learning rate during optimization with $\boldsymbol{T_{max} = 15}$ epochs and a minimum learning rate of $\mathbf{1 \times 10^{-6}}$ Mini-batch training was performed using a batch size of 8 for 15 training epochs. The loss weight for domain adaptation was empirically fixed at $\boldsymbol{\lambda = 0.3}$, and all experiments were conducted using a fixed random seed of 42 to ensure experimental reproducibility.

The full set of hyperparameters used for training is given in Table 7.

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

Parameter	Value
Input image size	384 × 384
Batch size	8
Number of epochs	15
Optimizer	AdamW
Initial learning rate	3×10^{-4}
Learning rate scheduler	Cosine Annealing
Weight decay	1×10^{-4}
Dropout rate	0.35
Domain-adaptation weight (λ)	0.3
Number of DR classes	3
Number of domains	2
Random seed	42

TABLE 7: Training configuration used for the proposed lesion-guided domain-adaptive DR grading framework

DR, Diabetic Retinopathy

As shown in Table *7*, the selected hyperparameters provide a balance between computational efficiency and model generalization. The relatively small batch size improves gradient diversity, whereas the cosine annealing scheduler facilitates smoother convergence during later training stages. Furthermore, the inclusion of dropout regularization and weight decay reduces the risk of overfitting when training on heterogeneous retinal image datasets.

Evaluation Metrics

The proposed framework was evaluated using multiple performance metrics to comprehensively assess classification performance and ordinal grading consistency. Since DR grading represents an ordinal multi-class classification problem, both conventional classification metrics and clinically relevant agreement measures were considered. The mathematical definitions of the primary classification metrics are summarized in Table *8*.

Metric	Formula
Accuracy	(TP + TN)/(TP + TN + FP + FN)
Precision	TP/(TP + FP)
Recall	TP/(TP + FN)
F1-score	2 × (Precision × Recall)/(Precision + Recall)

TABLE 8: Evaluation metrics used for diabetic retinopathy grading

TP, True Positive; TN, True Negative; FP, False Positive; FN, False Negative

Table *8* summarizes the primary classification metrics used in this study. Accuracy measures the overall proportion of correctly classified retinal images, whereas precision, recall, and F1-score evaluate the reliability, sensitivity, and balanced classification performance of the proposed framework, respectively. In addition to these conventional metrics, QWK was employed to assess the agreement between predicted and true DR severity grades while accounting for the ordinal relationship among disease stages. Unlike accuracy, QWK penalizes larger grading errors more heavily than adjacent-stage misclassifications and is therefore widely regarded as one of the most clinically relevant evaluation metrics for DR grading. Collectively, these metrics provide a comprehensive assessment of classification performance and ordinal grading consistency, thereby enabling a rigorous evaluation of the proposed lesion-guided domain-adaptive DR grading framework.

Overall DR grading performance

The proposed lesion-guided domain-adaptive framework achieved an accuracy of 90.26%, a precision of 90.10%, a recall of 90.26%, an F1-score of 90.14%, and a QWK of 89.28% on the independent test dataset. These results indicate that the framework correctly grades the severity of DR while still agreeing strongly with the ground-truth labels. Figure *2* presents the confusion matrix summarizing the class-wise classification performance of the proposed framework on the test dataset.

The confusion matrix generated by the proposed framework on the test dataset is shown in Figure *2*, illustrating the class-specific classification performance of the model.

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

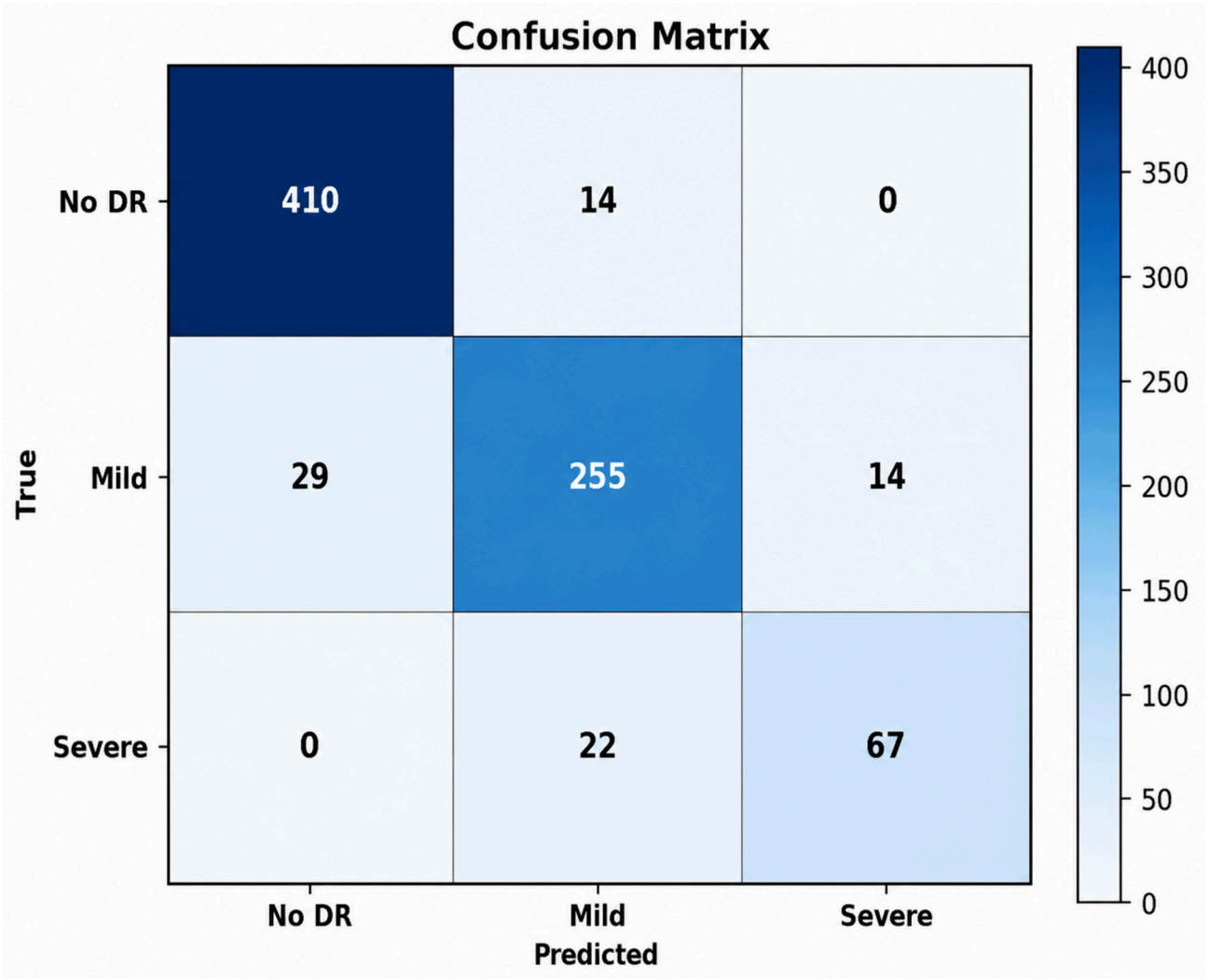


FIGURE 2: Confusion matrix of the proposed lesion-guided domain-adaptive framework on the test dataset

DR, Diabetic Retinopathy

As shown in Figure *2*, the confusion matrix corresponds to a test set of 811 retinal images, of which 732 were correctly classified, resulting in an overall accuracy of 90.26%. The proposed framework correctly classified 410 of 424 No DR images (96.7%), demonstrating excellent discrimination between healthy and pathological retinal images. Most misclassifications occurred in the Mild DR category, where subtle retinal lesions occasionally overlapped with characteristics of adjacent severity levels. Similarly, a few severe DR cases were misclassified as mild DR, reflecting the gradual progression of DR and the visual similarity between neighboring stages. Most classification errors occurred between adjacent severity classes, indicating that the proposed ordinal learning strategy effectively preserved the natural progression of disease severity. Overall, these results demonstrate that the integration of lesion-guided feature learning, domain adaptation, and ordinal severity modeling provides accurate and clinically meaningful DR grading across heterogeneous retinal image datasets.

Uncertainty and Explainability Analysis

Apart from diagnostic accuracy, the practicality of a machine learning-based automated grading scheme for DR requires generating reliable estimates of certainty levels and explainable decisions. To improve the transparency of the model, the proposed pipeline incorporates the use of MC Dropout and Grad-CAM methods.

To utilize MC Dropout during inference, the algorithm was run on each input sample for $T = 20$ stochastic forward passes. As a result, uncertainty-aware model evaluation yielded an accuracy of 89.89% and a QWK of 88.99%. Predictive uncertainty was quantified using the predictive entropy of the mean softmax distribution, whose theoretical range for the three-class classification problem is from 0 to $\ln(3)\ (\approx 1.099)$. The mean predictive entropy on the test set was 0.2571, corresponding to approximately 23% of the maximum possible entropy, indicating a relatively low level of predictive uncertainty. Samples with predictive entropy greater than 0.5 were classified as high-uncertainty. Based on this criterion, 156 of the 811 test retinal images (19.2%) were identified as high-uncertainty cases. These cases typically corresponded to borderline stages of DR, subtle retinal lesions, reduced image quality, or overlapping characteristics between adjacent severity classes. Consequently, the proposed uncertainty estimation approach effectively identifies ambiguous cases that may benefit from additional expert review before a final clinical diagnosis is made.

Figure *3* shows a histogram describing the distribution of uncertainty scores estimated via MC dropout. It can be seen that correctly classified samples are mostly located in the low uncertainty range, whereas incorrectly predicted retinas demonstrate a considerably higher value of this parameter. This indicates a close connection between prediction certainty and classification accuracy. Therefore, the usefulness of the implemented uncertainty estimation procedure is validated since it is able to differentiate between certain and ambiguous instances.

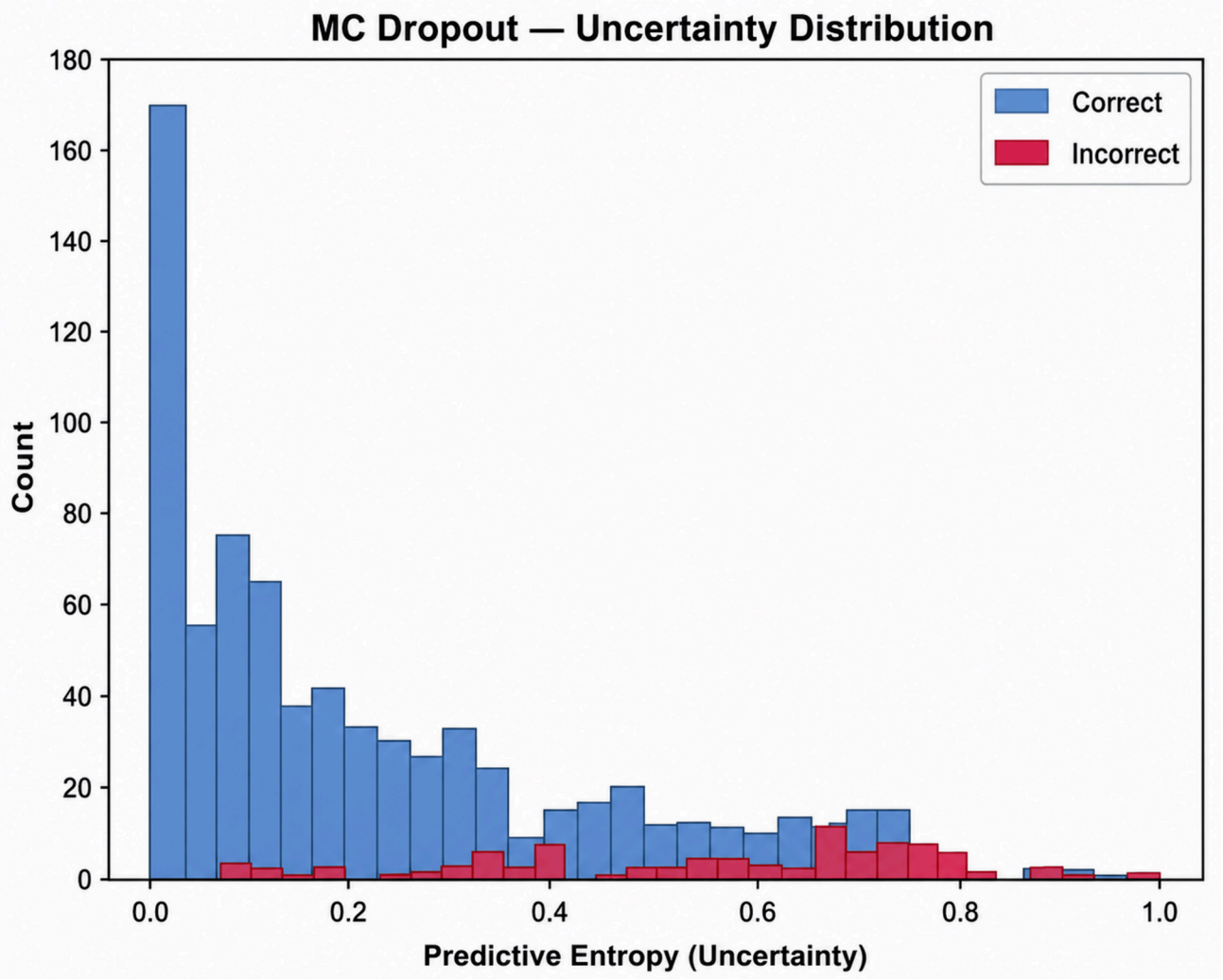


FIGURE 3: Distribution of uncertainty scores obtained using MC dropout-based uncertainty estimation

MC, Monte Carlo

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

Beyond uncertainty estimation, Grad-CAM was employed to visualize the retinal regions that contributed most strongly to the predicted DR grade. Qualitative inspection of representative Grad-CAM overlays indicated that the model frequently attended to retinal regions containing lesion-like structures commonly associated with DR, including areas consistent with microaneurysms, retinal hemorrhages, and distinct exudates, rather than predominantly highlighting irrelevant background regions. These observations suggest that the proposed lesion-guided attention mechanism encourages the network to focus on pathologically plausible retinal features. However, the present Grad-CAM analysis is qualitative in nature and should not be interpreted as formal clinical validation of explanation correctness. Quantitative evaluation using expert lesion annotations or localization metrics was not performed in this study and remains an important direction for future work.

Together, MC dropout-based uncertainty estimation and Grad-CAM-based visual explanations improve the transparency and interpretability of the proposed DR grading framework. While uncertainty estimation identifies predictions that may require additional clinical review, Grad-CAM provides visual evidence of the retinal regions influencing the model's decisions. These techniques are intended to support human-in-the-loop clinical decision-making rather than serve as standalone evidence of lesion localization accuracy.

Ablation Study

To examine the importance of each component in our lesion-guided domain-adaptive approach, we performed an ablative analysis in which each component was successively removed while all other experimental conditions remained constant. In particular, we considered the effect of lesion-guided attention, domain adaptation, and ordinal EMD loss. The performance outcomes are presented in Table *9*.

Configuration	Accuracy (%)	F1-score (%)	QWK (%)
Full Model (Proposed)	90.26	90.14	89.28
Without Lesion-Guided Attention	89.40	89.18	88.10
Without Domain Adaptation	88.95	88.70	87.65
Without Ordinal EMD Loss	89.10	88.82	87.90

TABLE 9: Ablation study results of the proposed framework

QWK, Quadratic Weighted Kappa; EMD, Earth Mover's Distance

According to Table *9*, the designed framework was sensitive to the presence of each architectural component. Removal of the lesion-guided attention network led to decreased accuracy, F1 score, and QWK, which shows that attention-based feature refining is useful for the accurate localization and representation of the pathological structures of the retina. By focusing on relevant areas and ignoring unnecessary features, the attention helps improve feature discrimination.

Furthermore, removing the ordinal EMD loss caused deterioration in classification and ordinal evaluation metrics. As a result, one can conclude that explicit ordinal modeling of the progression of DR helped achieve consistent grading by punishing higher ordinal misclassification errors.

Figure *4* shows the effect of individual components. It is interesting to note that even without the use of domain adaptation, the model reached satisfactory accuracy levels on the combined data. However, adversarial domain adaptation still plays an important role in reducing data-specific bias and improving feature correspondence between heterogenous retinal images.

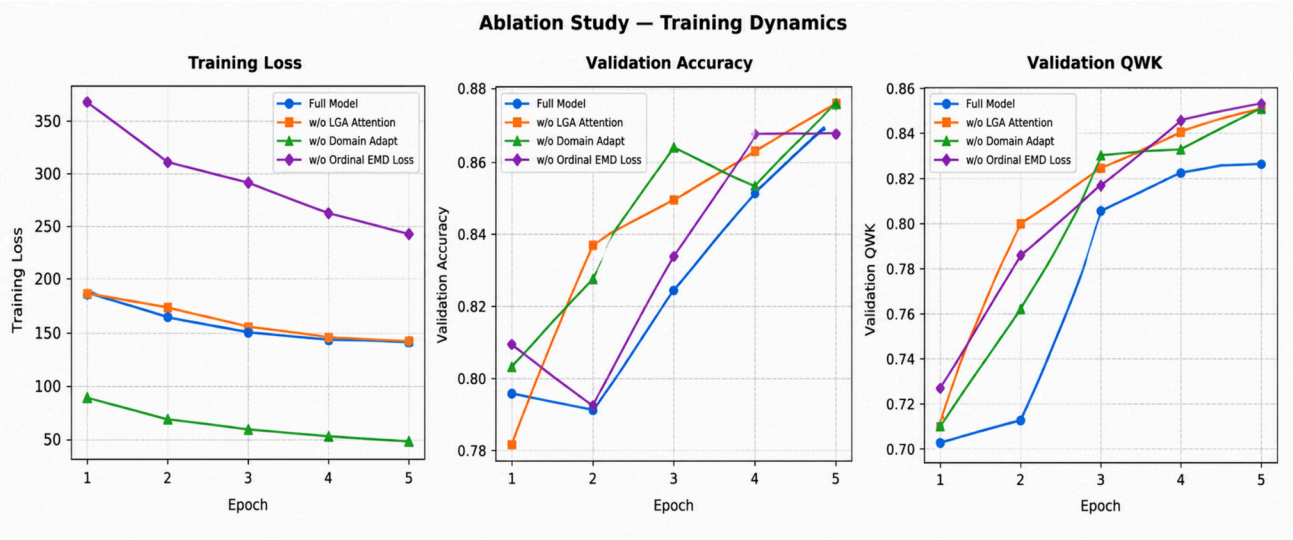


FIGURE 4: Contribution of individual framework components to diabetic retinopathy grading performance

QWK, Quadratic Weighted Kappa

Backbone Comparison

A comparative analysis was performed using four commonly used convolutional neural networks - EfficientNetV2-S, ResNet-50, ResNet-101, and DenseNet-121 - to determine the most suitable backbone for retinal feature extraction. For this comparison, all backbone networks were evaluated under the same baseline training configuration prior to integrating the proposed lesion-guided attention, adversarial domain adaptation, and ordinal EMD learning modules. The backbone comparison results are presented in Table *10*. After selecting EfficientNetV2-S as the backbone, the complete optimized framework incorporating all proposed components was trained, and its overall performance is reported in the primary experimental results.

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

Backbone	Accuracy (%)	F1-score (%)	QWK (%)
EfficientNetV2-S (Selected)	87.18	87.04	84.48
ResNet-50	85.04	84.68	81.51
ResNet-101	87.04	86.79	85.14
DenseNet-121	85.19	84.79	81.53

TABLE 10: Performance comparison of different backbone architectures

QWK, Quadratic Weighted Kappa

As presented in Table *10*, EfficientNetV2-S achieved the highest classification accuracy (87.18%) and F1-score (87.04%) among the evaluated backbone architectures, whereas ResNet-101 achieved the highest QWK score (85.14%), marginally outperforming EfficientNetV2-S (84.48%). Although ResNet-101 demonstrated slightly better ordinal agreement, EfficientNetV2-S consistently provided superior overall classification performance while maintaining competitive ordinal grading consistency. Furthermore, compared with ResNet-50 and DenseNet-121, EfficientNetV2-S exhibited stronger feature extraction capability, enabling more effective learning of both global retinal structures and fine-grained pathological features through its compound scaling strategy. Considering its favorable balance between classification performance, ordinal grading consistency, and computational efficiency, EfficientNetV2-S was selected as the backbone network for the proposed lesion-guided domain-adaptive framework. After integrating the lesion-guided attention module, adversarial domain adaptation, ordinal EMD learning, and subsequent hyperparameter optimization, the complete framework achieved the overall performance reported in the primary experimental results.

Figure *5* presents the performance comparison of the evaluated backbone architectures. While ResNet-101 achieved the highest QWK score, EfficientNetV2-S consistently delivered the best overall performance in terms of classification accuracy and F1 score while offering greater computational efficiency. These characteristics make EfficientNetV2-S more suitable for practical DR screening applications.

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

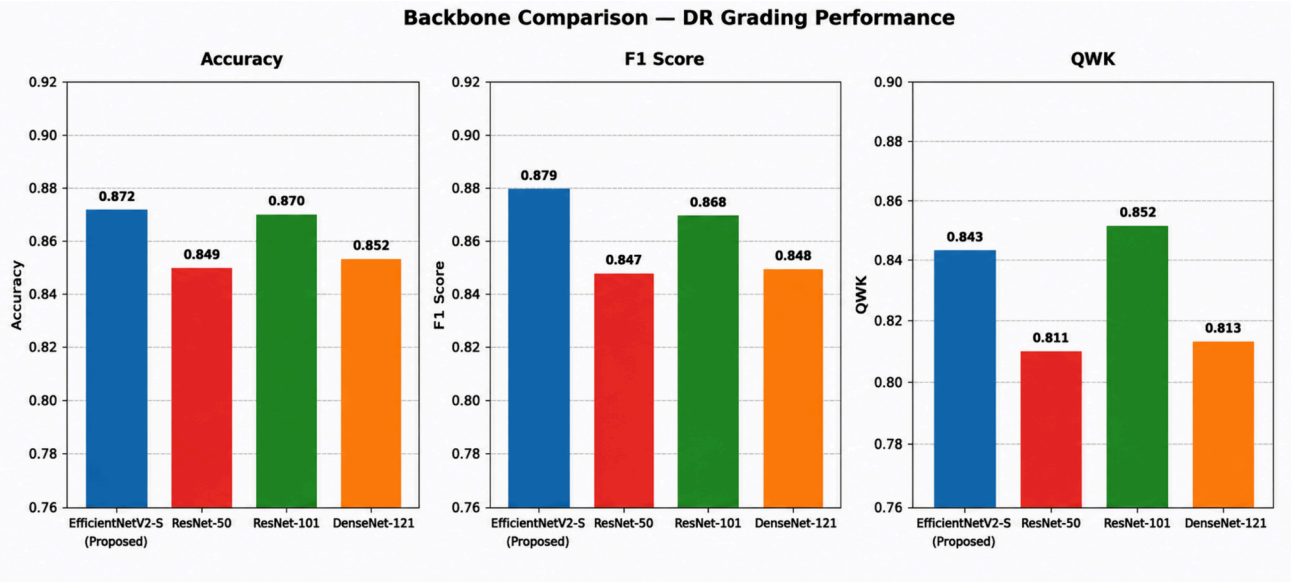


FIGURE 5: Performance comparison of different backbone architectures for diabetic retinopathy grading

QWK, Quadratic Weighted Kappa; DR, Diabetic Retinopathy

Based on the results presented in Table *10* and Figure *5*, EfficientNetV2-S was selected as the feature extraction backbone for the proposed lesion-guided domain-adaptive framework due to its favorable balance between classification performance, ordinal grading consistency, and computational efficiency.

Comparison with State-of-the-Art Methods

To assess the effectiveness of the proposed framework, a comparative analysis was performed against several recently reported deep learning-based DR grading methods. The comparison results are presented in Table *11*.

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

Method	Attention	Domain Adaptation	Ordinal Learning	Explainability	Uncertainty Estimation
MS-DNN [17]	No	No	No	No	No
CNN-RNN [18]	Yes	No	No	No	No
CNN-DenseNet [24]	No	No	No	No	No
AG-CNN [20]	Yes	No	No	Yes	No
Transfer Learning [19]	No	No	No	No	No
Feature Fusion CNN [21]	No	No	No	No	No
Texture Attention Net [45]	Yes	No	No	No	No
Proposed Framework	Yes	Yes	Yes	Yes	Yes

TABLE 11: Comparison of methodological capabilities among recent diabetic retinopathy grading approaches

CNN, Convolutional Neural Network; RNN, Recurrent Neural Network; DNN, Deep Neural Network

As presented in Table *11*, there are many research studies that obtain high classification accuracies on standard DR datasets. Nevertheless, almost all these studies mainly focus on enhancing the classification task and do not address domain shift, the ordinality of disease progression, uncertainty of predictions, and interpretability of the models.

On the contrary, the proposed approach combines a lesion-guided attention mechanism, domain adversarial training, ordinal learning, uncertainty estimation using MC Dropout, and Grad-CAM-based explainability into a unified framework. Compared with the representative methods summarized in Table *11*, the proposed framework is distinctive in jointly integrating all five methodological components. This comparison is limited to the representative methods included in Table *11* and does not imply a claim of uniqueness across the broader DR literature.

While achieving an accuracy at 90.26%, the framework reached a QWK value of 89.28%, which demonstrates its ability to correlate well with the severity labels assigned by experts and preserve the ordinality of the disease progression. Additionally, evaluating the framework on the combination of the APTOS2019 and MESSIDOR datasets can help test it under more diverse imaging techniques and patient samples.

Therefore, the main contribution of the proposed framework consists not just in obtaining high classification accuracy but in being robust, reliable, interpretable, and clinically applicable as well.

Limitations and Future Work

However, there are some limitations worth mentioning for the proposed framework for automated DR grading based on lesion awareness, cross-domain generalization capacity, and uncertainty quantification. First of all, the proposed method was evaluated on only two publicly available retinal image databases (APTOS2019 and MESSIDOR) to demonstrate its effectiveness. As described above, the use of both retinal image databases brought diversity to the model training process; however, further application of the proposed framework for DR grading requires its prospective validation on larger multicenter datasets. Future work will also include external validation on independent clinical cohorts acquired with different imaging devices and clinical settings to further assess the generalizability and clinical applicability of the proposed framework.

The second limitation of the presented study is the reformulation of the DR grading problem into a three-class classification problem. By grouping the original DR severity grading scale into three groups, we can formulate the problem better. However, this method may negatively affect the granularity of disease severity assessment, which may restrict further disease progression analysis and clinical decision-making. Future research on DR grading should consider the implementation of a fine-grained five-class grading solution to solve this problem.

The third disadvantage of the presented study is the retrospective nature of the evaluation of the proposed framework based on the use of retinal images collected in the past. Future studies are needed to evaluate the potential of the framework to be integrated into the routine ophthalmological examination workflow, including the assessment of its performance on different imaging devices. The presented study is limited so far to color retinal fundus photographs, but other ophthalmic imaging procedures are also considered. Furthermore, thorough computational benchmarking, including inference time, computational complexity, and robustness against low-quality retinal images, will be performed to advocate real-world clinical applicability.

Some ways to improve the developed framework should be considered. Such improvements can be associated with the emergence of vision transformers and hybrid CNN and transformer-based architectures. We will also compare with recent transformer-based and foundation-model approaches for retinal image analysis in future work. In addition, statistical significance analysis and confidence interval estimation will be included to provide some additional evidence of the robustness of the reported performance improvements. Moreover, future research can focus on the application of federated learning to train models across multiple medical institutions while preserving the anonymity of patients. Another way to improve the framework is through the combination of multimodal DR assessment, i.e., using retinal images together with metadata. We used Grad-CAM to provide a qualitative interpretation of the model. However, future work will focus on quantitative validation of the quality of explanations using expert ophthalmologist annotations. Finally, the use of advanced uncertainty-aware learning techniques may be beneficial for further studies.

Despite the limitations of the research, the proposed framework provides a solid foundation for the implementation of robust, lesion-aware, domain-adaptive, and uncertainty-guided DR grading systems.

Conclusions

DR grading is difficult due to the presence of small retinal lesions, high variability among imaging data, ordinal nature of retinopathy progression, and the requirement for reliable predictions. To tackle these issues, a lesion-guided domain-adaptive framework was proposed in this work. The proposed solution includes the use of the EfficientNetV2-S feature extractor, a lesion-aware attention mechanism, adversarial domain adaptation, ordinal learning, and explainability techniques. With the help of heterogeneous retinal images taken from the APTOS2019 and MESSIDOR datasets, the proposed framework was able to learn robust retinal features, improve generalizability by overcoming domain shift, and preserve the ordinal properties of retinopathy progression. Additionally, uncertainty analysis and the application of Grad-CAM helped make model predictions more transparent and reliable, which can be used for decision-making purposes. Comparative experiments, backbone analysis, and ablation experiments confirmed the usefulness of each component included in the proposed approach and demonstrated that lesion-aware modeling, domain adaptation, and ordinal

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

learning should be used in conjunction. Overall, results indicated that DR grading could be done in an automated way, provided that a proper set of tools is applied for both learning discriminative retinal features and improving prediction quality and interpretability.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Vijay Kalmani, Shilpa R. Bhagat

Acquisition, analysis, or interpretation of data: Vijay Kalmani, Amol C. Adamuthe, Shilpa R. Bhagat

Drafting of the manuscript: Vijay Kalmani, Shilpa R. Bhagat

Critical review of the manuscript for important intellectual content: Vijay Kalmani, Amol C. Adamuthe, Shilpa R. Bhagat

Supervision: Vijay Kalmani, Amol C. Adamuthe

Disclosures

Human subjects: Consent was obtained or waived by all participants in this study. Not Applicable (Publicly Available De-identified Dataset Study) issued approval Not Applicable. This study was conducted using publicly available, de-identified retinal fundus image datasets (APTOS2019 and MESSIDOR) and involved only secondary analysis of anonymized data. No direct patient recruitment, intervention, or access to identifiable personal information was involved. Therefore, Institutional Review Board (IRB)/Ethics Committee approval was not required for this study. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

References

1. Yang M, Liu Y, Ma Y, Chen Z, Duan R, Jiang L, Li T: Global, regional, and national burden of blindness due to diabetic retinopathy, 1990-2021. Ophthalmology and Therapy. 2025, 14:2599-2615. 10.1007/s40123-025-01230-y

2. Yang M, Liu Y, Duan R, et al.: Global, regional, and national epidemiology of vision impairment due to diabetic retinopathy among working-age population, 1990-2021. Journal of Diabetes. 2025, 17:e70121. 10.1111/1753-0407.70121

3. Pan Y, Li Y, Cui M, He G, Wang G: Global, regional and national burden of blindness and vision loss attributable to diabetic retinopathy, 1990-2021: A systematic analysis for the Global Burden of Disease Study 2021. Diabetes, Obesity and Metabolism. 2025, 27:5464-5477. 10.1111/dom.16588

4. Li X, Tan T-E, Wong TY, Sun X: Diabetic retinopathy in China: Epidemiology, screening and treatment trends—A review. Clinical and Experimental Ophthalmology. 2023, 51:607-626. 10.1111/ceo.14269

5. Benhamza M, Dahlui M, Said MA: Determining direct, indirect healthcare and social costs for diabetic retinopathy management: a systematic review. BMC Ophthalmology. 2024, 24:424. 10.1186/s12886-024-03665-6

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

6. Rudnicka AR, Shakespeare R, Chambers R, et al.: Automated retinal image analysis systems to triage for grading of diabetic retinopathy: a large-scale, open-label, national screening programme in England. The Lancet Digital Health. 2025, 7:100914. 10.1016/j.landig.2025.100914

7. Jagadesh BN, Karthik MG, Siri D, Khaja Shareef SK, Mantena SV, Vatambeti R: Segmentation using the IC2T model and classification of diabetic retinopathy using the rock hyrax swarm-based coordination attention mechanism. IEEE Access. 2023, 11:124441-124458. 10.1109/access.2023.3330436

8. Bilal A, Zhu L, Deng A, Lu H, Wu N: AI-based automatic detection and classification of diabetic retinopathy using u-net and deep learning. Symmetry. 2022, 14:1427-1427. 10.3390/sym14071427

9. Kim HJ, Eesaar H, Chong KT: Transformer-enhanced retinal vessel segmentation for diabetic retinopathy detection using attention mechanisms and multi-scale fusion. Applied Sciences. 2024, 14:10658. 10.3390/app142210658

10. Soomro TA, Ubaid FB, Zardari S, Shah AA, Alansari MM, Nagro SA: The state of retinal image analysis: Deep learning advances and applications. IEEE Access. 2025, 13:65066-65093. 10.1109/access.2025.3556765

11. Rezaee K, Farnami F: Innovative approach for diabetic retinopathy severity classification: An AI-powered tool using cnn-transformer fusion. Journal of Biomedical Physics and Engineering. 2025, 15:137-158. 10.31661/jbpe.v0i0.2408-1811

12. Ara T, Mishra VP, Bali M, Yenkikar A: Hybrid quantum-classical deep learning framework for balanced multiclass diabetic retinopathy classification. MethodsX. 2025, 15:103605. 10.1016/j.mex.2025.103605

13. Akhtar S, Aftab S, Ali O, Ahmad M, Khan MA, Abbas S, Ghazal TM: A deep learning based model for diabetic retinopathy grading. Scientific Reports. 2025, 15:3763. 10.1038/s41598-025-87171-9

14. Tajudin NMA, Kipli K, Mahmood MH, et al.: Deep learning in the grading of diabetic retinopathy: A review. IET Computer Vision. 2022, 16:667-682. 10.1049/cvi2.12116

15. Huang X, Wang H, She C, et al.: Artificial intelligence promotes the diagnosis and screening of diabetic retinopathy. Frontiers in Endocrinology. 2022, 13:946915. 10.3389/fendo.2022.946915

16. Bhimavarapu U, Battineni G: Deep learning for the detection and classification of diabetic retinopathy with an improved activation function. Healthcare. 2023, 11:97. 10.3390/healthcare11010097

17. Mustafa H, Ali SF, Bilal M, Hanif MS: Multi-stream deep neural network for diabetic retinopathy severity classification under a boosting framework. IEEE Access. 2022, 10:113172-113183. 10.1109/access.2022.3217216

18. Sushith M, Sathiya A, Kalaipoonguzhali V, Sathya V: A hybrid deep learning framework for early detection of diabetic retinopathy using retinal fundus images. Scientific Reports. 2025, 15:15166. 10.1038/s41598-025-99309-w

19. Vallukappully S, van der Linde I, Chakraborty A: Early detection and classification of diabetic retinopathy by transfer learning of NASNet-large and ResNet-50 convolutional neural networks. Informatics in Medicine Unlocked. 2025, 58:101688. 10.1016/j.imu.2025.101688

20. Moustari AM, Brik Y, Attallah B, Bouaouina R: Two-stage deep learning classification for diabetic retinopathy using gradient weighted class activation mapping. Automatika. 2024, 65:1284-1299. 10.1080/00051144.2024.2363692

21. Ali G, Dastgir A, Iqbal MW, Anwar M, Faheem M: A hybrid convolutional neural network model for automatic diabetic retinopathy classification from fundus images. IEEE Journal of Translational Engineering in Health and Medicine. 2023, 11:341-350. 10.1109/jtehm.2023.3282104

22. Suedumrong C, Phongmoo S, Akarajaka T, Leksakul K: Diabetic retinopathy detection using convolutional neural networks with background removal and data augmentation. Applied Sciences. 2024, 14:8823. 10.3390/app14198823

23. Mutawa AM, Hemalakshmi GR, Prakash NB, Murugappan M: Randomization-driven hybrid deep learning for diabetic retinopathy detection. IEEE Access. 2025, 13:38901-38913. 10.1109/access.2025.3546359

24. Yasashvini R, Sarobin MVR, Panjanathan R, Jasmine SG, Anbarasi LJ: Diabetic retinopathy classification using CNN and hybrid deep convolutional neural networks. Symmetry. 2022, 14:1932. 10.3390/sym14091932

25. Yang B, Li T, Xie H, Liao Y, Chen YPP: Classification of diabetic retinopathy severity based on GCA attention mechanism. IEEE Access. 2022, 10:2729-2739. 10.1109/access.2021.3139129

26. Ashraf MH, Mehmood MN, Ahmed M, et al.: HIRD-Net: An explainable CNN-Based framework with attention mechanism for diabetic retinopathy diagnosis using CLAHE-D-DoG enhanced fundus images. Life. 2025, 15:1411. 10.3390/life15091411

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4

27. Niranjan KR, Srinivasa Rao N, Rangahanumaiah N, Ullala Mata BN, Anantharamaiah V: A hybrid vision transformer and graph neural network model with attention mechanisms for diabetic retinopathy detection. Journal of Artificial Intelligence and Technology. 2025, 5:10.37965/jait.2025.0645

28. Sushith M, Lakkshmanan A, Saravanan M, Castro S: Attention dual transformer with adaptive temporal convolutional for diabetic retinopathy detection. Scientific Reports. 2025, 15:7694. 10.1038/s41598-025-92510-x

29. Wang Y, Shen S, Tong N: JFSP-ViT: A joint frequency-spatial perception vision transformer for diabetic retinopathy grading. IEEE Access. 2025, 13:150348-150364. 10.1109/access.2025.3600453

30. Liu Y, Yao D, Ma Y, et al.: STMF-DRNet: A multi-branch fine-grained classification model for diabetic retinopathy using Swin-TransformerV2. Biomedical Signal Processing and Control. 2025, 103:107352. 10.1016/j.bspc.2024.107352

31. Hu J, Wang H, Wang L, Lu Y: Graph adversarial transfer learning for diabetic retinopathy classification. IEEE Access. 2022, 10:119071-119083. 10.1109/access.2022.3220776

32. Men Y, Fhima J, Celi LA, Ribeiro LZ, Nakayama LF, Behar JA: Deep learning generalization for diabetic retinopathy staging from fundus images. Physiological Measurement. 2025, 46:015001. 10.1088/1361-6579/ada86a

33. Huang Y, Lyu J, Cheng P, Tam R, Tang X: SSiT: Saliency-guided self-supervised image transformer for diabetic retinopathy grading. IEEE Journal of Biomedical and Health Informatics. 2024, 28:2806-2817. 10.1109/jbhi.2024.3362878

34. Qian P, Zhao Z, Chen C, Zeng Z, Li X: Two eyes are better than one: Exploiting binocular correlation for diabetic retinopathy severity grading. 3rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico. 2021, 2115-2118. 10.1109/embc46164.2021.9630812

35. Farag MM, Fouad M, Abdel-Hamid AT: Automatic severity classification of diabetic retinopathy based on densenet and convolutional block attention module. IEEE Access. 2022, 10:38299-38308. 10.1109/access.2022.3165193

36. Chilukoti SV, Shan L, Tida VS, Maida AS, Hei X: A reliable diabetic retinopathy grading via transfer learning and ensemble learning with quadratic weighted kappa metric. BMC Medical Informatics and Decision Making. 2024, 24:37. 10.1186/s12911-024-02446-x

37. Alghamdi HS: Towards explainable deep neural networks for the automatic detection of diabetic retinopathy. Applied Sciences. 2022, 12:9435. 10.3390/app12199435

38. Zhang P, Zhao J, Liu Q, Liu X, Li X, Gao Y, Li W: Fundus image generation and classification of diabetic retinopathy based on convolutional neural network. Electronics. 2024, 13:3603. 10.3390/electronics13183603

39. Yaman Atcı Ş, Güneş A, Zontul M, Arslan Z: Identifying diabetic retinopathy in the human eye: A hybrid approach based on a computer-aided diagnosis system combined with deep learning. Tomography. 2024, 10:215-230. 10.3390/tomography10020017

40. Kansal S, Mishra BK, Sethi S, Vinayak K, Kansal P, Narayan J: Retinodeep: Leveraging deep learning models for advanced retinopathy diagnostics. Sensors. 2025, 25:5019. 10.3390/s25165019

41. Saporta A, Gui X, Agrawal A, et al.: Benchmarking saliency methods for chest X-ray interpretation. Nature Machine Intelligence. 2022, 4:867-878. 10.1038/s42256-022-00536-x

42. Akram M, Adnan M, Ali SF, Ahmad J, Yousef A, Alshalali TAN, Shaikh ZA: Uncertainty-aware diabetic retinopathy detection using deep learning enhanced by Bayesian approaches. Scientific Reports. 2025, 15:1342. 10.1038/s41598-024-84478-x

43. Spaide T, Rajesh AE, Gim N, et al.: Estimating uncertainty of geographic atrophy segmentations with Bayesian deep learning. Ophthalmology Science. 2024, 5:100587. 10.1016/j.xops.2024.100587

44. Alahmadi MD: Texture attention network for diabetic retinopathy classification. IEEE Access. 2022, 10:55522-55532. 10.1109/access.2022.3177651

45. Das PK, Pumrin S: Diabetic retinopathy classification: Performance evaluation of pre-trained lightweight CNN using imbalance dataset. Engineering Journal. 2024, 28:13-25. 10.4186/ej.2024.28.7.13

46. Canipek AS, Çakan M, Aktuğ A: Eyepacs, Aptos, Messidor Diabetic Retinopathy. Kaggle. 2024,

How to cite this article:
Adamuthe A C, Bhagat S R, Kalmani V (August 17, 2026) Toward Reliable Diabetic Retinopathy Severity Assessment via Cross-Domain Retinal Representation Learning. Cureus J Comput Sci 3 : es44389-026-00226-4. DOI https://doi.org/10.7759/s44389-026-00226-4