iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79563

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

A Comprehensive Review of Datasets for Diabetic Retinopathy Detection Using Deep Learning

Shikha Methil[1], Dr. Mukul Shrivastava[2]
[1]*Research Scholar,* [2]*Assistant Professor, Department of Electronic and Communication Engineering, Bansal Institute of Science and Technology Bhopal*

Abstract: ***Diabetic Retinopathy (DR) is one of the most prevalent microvascular complications of diabetes and remains a leading cause of preventable blindness worldwide. Early diagnosis and timely intervention are critical to reducing vision loss; however, traditional screening methods are often resource-intensive and dependent on expert ophthalmologists. In recent years, deep learning-based automated systems have emerged as powerful tools for DR detection and grading, primarily utilizing retinal fundus images. Despite their promising performance, the effectiveness and generalizability of these models are significantly influenced by the datasets used for training and evaluation.This review paper presents a comprehensive and systematic analysis of publicly available datasets for diabetic retinopathy detection. The study emphasizes dataset characteristics such as size, image quality, annotation type, and class distribution, which play a crucial role in model performance. Furthermore, datasets are categorized based on their scale (small, medium, and large), annotation levels (image-level, lesion-level, and pixel-level), and clinical applicability. The paper also highlights key challenges, including class imbalance, variability in imaging conditions, and limited availability of high-quality annotated data. By providing a detailed comparative evaluation, this review aims to guide researchers in selecting appropriate datasets and identifying gaps for future dataset development in DR research.***

Keywords: Diabetic Retinopathy, Deep Learning, Medical Imaging, Retinal Fundus Images, Dataset Analysis, Image Classification, Image Segmentation.

I. INTRODUCTION

Diabetic Retinopathy (DR) is a progressive microvascular complication of diabetes mellitus that affects the retinal blood vessels and is recognized as one of the leading causes of vision impairment and blindness worldwide. The condition typically develops over time due to prolonged hyperglycemia, resulting in structural and functional damage to the retina. Early detection and timely treatment are essential to prevent irreversible vision loss; however, conventional screening methods rely heavily on manual examination by ophthalmologists, which can be time-consuming, costly, and inaccessible in resource-limited settings.In recent years, the integration of deep learning techniques, particularly Convolutional Neural Networks (CNNs), has significantly advanced the field of automated DR detection. These models are capable of analyzing retinal fundus images with high accuracy, enabling scalable and efficient screening solutions. Despite these advancements, the success of deep learning models is fundamentally dependent on the availability and quality of datasets used for training and validation.Datasets play a crucial role in determining the robustness, generalizability, and reliability of DR detection systems. High-quality datasets with diverse samples, accurate annotations, and balanced class distributions facilitate effective feature learning and improve model performance. Conversely, datasets with poor image quality, limited size, or class imbalance can lead to overfitting and biased predictions. Therefore, a comprehensive understanding of existing DR datasets—their characteristics, strengths, and limitations—is essential for developing reliable and clinically applicable automated diagnostic systems.

Figure 1: Introduction to Diabetic Retinopathy Detection Using Deep Learning

©IJRASET: All Rights are Reserved

II. IMPORTANCE OF DATASETS IN DEEP LEARNING FOR DR

Deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable capability in the automated detection and classification of Diabetic Retinopathy (DR). However, their performance is highly dependent on the availability of large, well-annotated, and diverse datasets. Unlike traditional machine learning methods, deep learning models learn hierarchical feature representations directly from data, making the quality and structure of the dataset a critical factor in determining model effectiveness.One of the primary aspects of dataset importance is data diversity, which ensures that the model can generalize well across different populations, imaging devices, and clinical conditions. Diverse datasets expose models to variations in illumination, noise, and pathological features, thereby enhancing robustness. Additionally, annotation quality plays a vital role in determining prediction accuracy, as incorrect or inconsistent labels can significantly degrade model performance. Class balance is another crucial factor, as imbalanced datasets—where normal cases dominate over severe DR cases—can lead to biased predictions and poor sensitivity for minority classes.Furthermore, image resolution directly influences the model's ability to capture fine-grained retinal features such as microaneurysms and hemorrhages. Large-scale datasets like EyePACS have significantly contributed to advancements in DR detection by providing real-world variability and sufficient data volume for training deep learning models effectively.

III. CLASSIFICATION OF DR DATASETS

Datasets used for diabetic retinopathy (DR) detection play a fundamental role in the design, training, and evaluation of deep learning models. Given the diversity in dataset characteristics, it is essential to classify them systematically to better understand their applicability for different research objectives. DR datasets can be broadly categorized based on three primary criteria: dataset size, type of annotation, and clinical relevance. Each classification provides unique insights into how datasets contribute to various deep learning tasks such as classification, segmentation, detection, and clinical decision support.

A. Based on Size

The size of a dataset is one of the most critical factors influencing the performance and generalization capability of deep learning models. DR datasets are typically categorized into small-scale, medium-scale, and large-scale datasets based on the number of images they contain.

Small-scale datasets, usually comprising fewer than 1,000 images, are often used for preliminary studies, algorithm validation, and proof-of-concept implementations. While these datasets may offer high-quality annotations, they are limited in diversity and may lead to overfitting when used for training complex models.Medium-scale datasets, containing between 1,000 and 10,000 images, strike a balance between data availability and computational feasibility. These datasets are commonly used for benchmarking and comparative studies, providing sufficient variability while remaining manageable for training purposes. In contrast, large-scale datasets, consisting of more than 10,000 images, are essential for training deep learning models capable of achieving high accuracy and robustness. Such datasets expose models to real-world variability, including differences in imaging conditions, patient demographics, and disease severity. However, large-scale datasets often present challenges such as class imbalance, inconsistent labeling, and increased computational requirements. Therefore, selecting an appropriate dataset size depends on the specific research objective, available computational resources, and desired level of model generalization.

B. Based on Annotation Type

Another important classification criterion is the type of annotation provided within the dataset, which directly determines the type of deep learning task that can be performed. DR datasets can include image-level labels, pixel-level annotations, and lesion-level annotations. Image-level labels are the most common form of annotation and are typically used for classification tasks, where each retinal image is assigned a severity grade based on standardized scales. These datasets are relatively easier to construct, as they require less detailed labeling, but they may lack fine-grained information about lesion localization.Pixel-level annotations provide detailed information about the exact location and boundaries of pathological features such as microaneurysms, hemorrhages, and exudates. These annotations are essential for segmentation tasks, enabling models to identify and delineate specific lesions within retinal images. However, generating pixel-level annotations is time-consuming and requires expert ophthalmological input, making such datasets relatively scarce. Lesion-level annotations, which lie between image-level and pixel-level annotations, provide information about the presence and approximate location of lesions. These datasets are particularly useful for object detection tasks and offer a balance between annotation complexity and informational richness. The choice of annotation type depends on the intended application, with more detailed annotations generally enabling more advanced and interpretable models.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Based on Clinical Relevance

DR datasets can also be classified based on their clinical relevance, which reflects how closely they align with real-world medical applications. Screening datasets are designed for early detection and typically include a large number of images with varying quality and minimal annotations. These datasets are used to develop models capable of identifying DR in large populations, particularly in mass screening programs. Diagnostic datasets, on the other hand, are more refined and include high-quality images with detailed annotations, making them suitable for precise disease grading and clinical decision-making.In addition, multi-modal datasets have emerged as an advanced category that combines different types of medical imaging data, such as fundus images and Optical Coherence Tomography (OCT) scans. These datasets provide complementary information, enabling more comprehensive analysis and improved diagnostic accuracy. Multi-modal datasets are particularly valuable for developing sophisticated deep learning models that can integrate multiple sources of information. However, they also introduce challenges related to data integration, standardization, and increased computational complexity. Overall, classifying datasets based on clinical relevance helps researchers select appropriate data sources that align with their specific application goals, whether for large-scale screening, precise diagnosis, or advanced multimodal analysis.

IV. MAJOR PUBLICLY AVAILABLE DATASETS

A. EyePACS Dataset

The EyePACS dataset is one of the largest and most widely used datasets in diabetic retinopathy research, containing over 35,000 retinal fundus images labeled on a severity scale ranging from 0 to 4. These images are collected under diverse real-world conditions, including variations in illumination, focus, and imaging devices, making the dataset highly representative of practical screening environments. Its large size enables effective training of deep learning models and enhances their robustness. However, the dataset lacks pixel-level annotations, limiting its use in segmentation tasks, and suffers from class imbalance, which can bias model predictions toward majority classes.

B. APTOS 2019 Dataset

The APTOS 2019 dataset consists of 3,662 retinal fundus images collected in India and labeled according to internationally accepted DR severity levels. It provides real-world clinical data with relatively balanced class distributions, making it suitable for classification and grading tasks. The dataset is particularly useful for evaluating model performance in practical scenarios. However, compared to larger datasets like EyePACS, its smaller size limits deep learning scalability. Additionally, variations in image quality, including differences in brightness and clarity, may affect model training and require preprocessing techniques to ensure consistency and improve performance.

C. Messidor Dataset (and Messidor-2)

The Messidor dataset and its extended version, Messidor-2, are widely used as benchmark datasets for evaluating diabetic retinopathy detection systems. Messidor-2 contains 1,748 images from 874 examinations and is known for its high-quality fundus images and clinically validated annotations. These characteristics make it highly reliable for performance comparison and validation studies. However, the dataset is relatively small and lacks diversity in terms of imaging conditions and patient demographics. As a result, models trained solely on this dataset may not generalize well to real-world clinical environments with greater variability.

D. IDRiD Dataset (Indian Diabetic Retinopathy Image Dataset)

The IDRiD dataset contains 516 high-resolution retinal images and provides both image-level grading and detailed lesion annotations. It is particularly valuable for segmentation and lesion detection tasks, as it includes pixel-level annotations for pathological features such as microaneurysms and exudates. This makes it highly suitable for developing interpretable deep learning models. However, the dataset is limited in size, which restricts its use for training large-scale models. Additionally, its limited diversity may reduce the generalizability of models when applied to broader populations or different imaging conditions.

E. DDR Dataset (DeepDR)

The DDR dataset is a medium-scale dataset containing over 13,000 retinal images, designed to support both classification and lesion detection tasks. It offers a balance between dataset size and annotation detail, making it suitable for multi-task learning approaches in deep learning. The dataset enables models to simultaneously perform grading and lesion localization, improving diagnostic performance.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

However, one of its main limitations is restricted public accessibility, which can hinder widespread research and reproducibility. Additionally, variations in annotation standards may introduce inconsistencies when compared with other commonly used datasets.

F. Combined and Preprocessed Datasets

Combined datasets are created by merging multiple publicly available datasets such as EyePACS, APTOS 2019, and Messidor to form larger datasets comprising up to approximately 92,000 images. These datasets aim to overcome limitations related to data scarcity and class imbalance by increasing both volume and diversity. They are particularly useful for training robust deep learning models with improved generalization. However, combining datasets introduces challenges such as domain shift, differences in imaging conditions, and inconsistent labeling standards, which may negatively impact model performance if not properly addressed through normalization and preprocessing techniques.

G. Comparative Analysis of Datasets

A comparative analysis of widely used diabetic retinopathy (DR) datasets is essential to understand their relative strengths, limitations, and suitability for different deep learning applications. Each dataset differs in terms of size, annotation type, image quality, and clinical relevance, which directly impacts model performance and generalization. Large-scale datasets such as EyePACS provide sufficient data for training robust models, while smaller datasets like IDRiD offer detailed annotations useful for segmentation tasks. Similarly, benchmark datasets such as Messidor-2 are valuable for evaluation purposes. Therefore, comparing these datasets helps researchers select appropriate data sources based on specific research objectives.

Table 1: Comparative Analysis of Major Diabetic Retinopathy Datasets.

Dataset	No. of Images	Annotation Type	Strength	Limitation
EyePACS	~35,000+	Image-level	Large-scale	Class imbalance
APTOS 2019	3,662	Image-level	Real-world data	Smaller size
Messidor-2	1,748	Image-level	High quality	Limited diversity
IDRiD	516	Pixel + Image-level	Detailed annotations	Very small
DDR	~13,000	Multi-level	Multi-task learning	Limited access

V. CHALLENGES IN EXISTING DATASETS

A. Class Imbalance

Class imbalance is one of the most significant challenges in diabetic retinopathy (DR) datasets, where the number of normal or mild cases often far exceeds the number of severe cases. This imbalance can lead to biased model training, as deep learning algorithms tend to favor majority classes during optimization. Consequently, models may achieve high overall accuracy but perform poorly in detecting minority classes, particularly severe or proliferative DR, which are clinically critical. This limitation reduces the sensitivity and reliability of automated systems in real-world screening scenarios. Addressing class imbalance typically requires techniques such as data augmentation, resampling strategies, or cost-sensitive learning; however, these methods may introduce additional complexity and may not fully replicate real clinical distributions.

B. Variability in Imaging Conditions

Datasets such as EyePACS include retinal images captured using different cameras, settings, and acquisition protocols, resulting in significant variability in imaging conditions. Variations in illumination, contrast, focus, and field of view can introduce noise and inconsistencies in the dataset, making it challenging for deep learning models to learn robust and generalized features. This variability may lead to reduced model performance when applied to new or unseen data from different clinical environments. To mitigate this issue, preprocessing techniques such as image normalization, contrast enhancement, and quality assessment are commonly employed. However, achieving complete standardization across diverse datasets remains a difficult task.

C. Limited Annotated Data

The availability of high-quality annotated data is a major constraint in DR research, particularly for tasks requiring detailed pixel-level annotations. Creating such annotations involves expert ophthalmologists manually identifying and labeling pathological features such as microaneurysms, hemorrhages, and exudates, which is both time-consuming and resource-intensive. As a result, datasets like IDRiD that provide detailed annotations are limited in size.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

This scarcity restricts the development of advanced deep learning models for segmentation and lesion detection. Furthermore, inconsistencies in annotation standards across datasets can introduce variability, affecting model training and evaluation.

D. Domain Shift

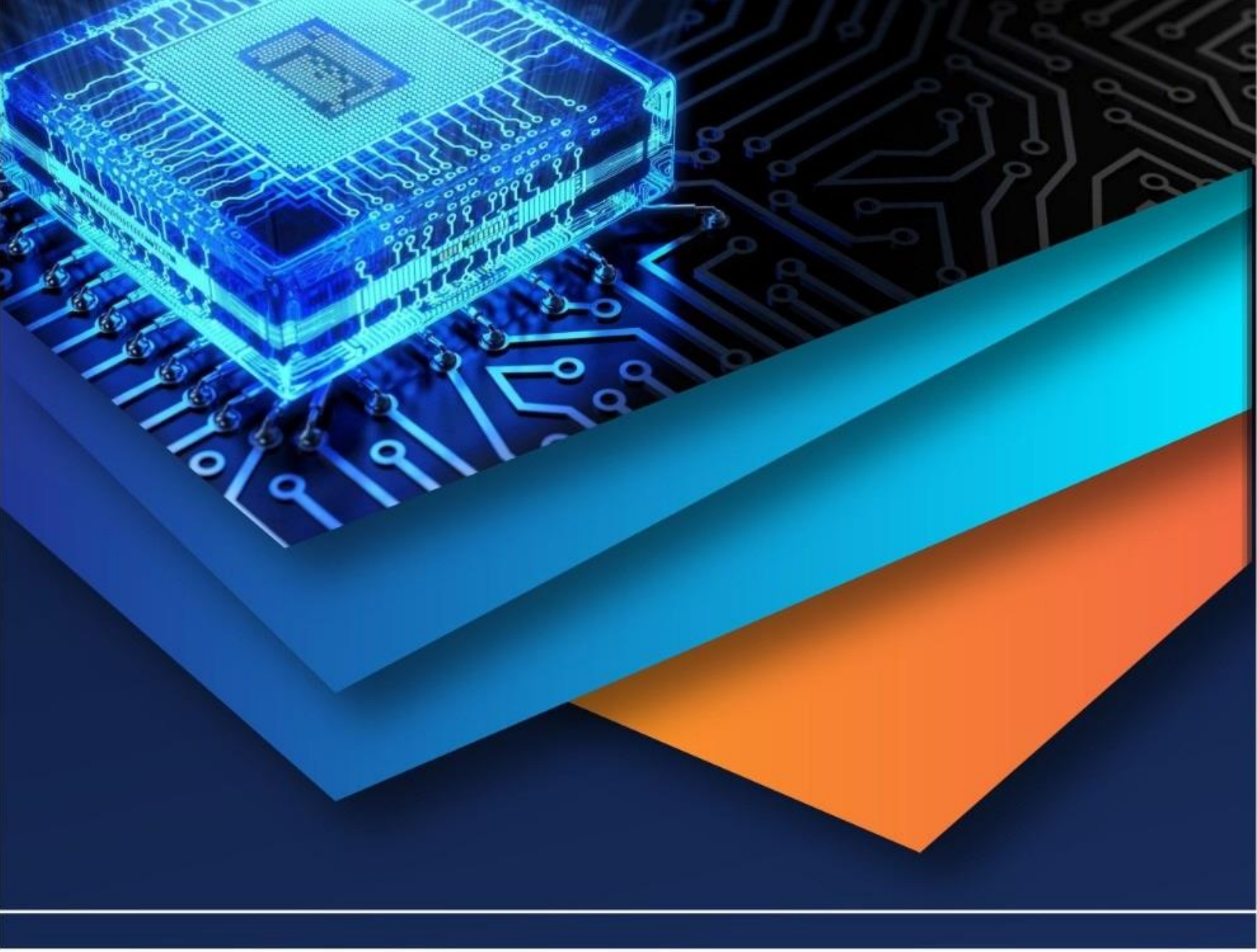
Domain shift arises when datasets from different sources are combined or when models trained on one dataset are applied to another with different characteristics. Variations in image resolution, annotation protocols, and clinical labeling standards contribute to this issue, leading to discrepancies in data distribution. Such inconsistencies can significantly degrade model performance, as the learned features may not generalize well across domains. Domain shift is particularly problematic in multi-dataset training, where differences in imaging devices and population demographics further complicate model adaptation. Addressing this challenge requires advanced techniques such as domain adaptation, transfer learning, and dataset standardization, although achieving complete consistency across datasets remains an ongoing research challenge.

VI. CONCLUSION

Datasets play a fundamental and indispensable role in the development and performance of deep learning models for diabetic retinopathy (DR) detection. The success of automated diagnostic systems is highly dependent on the quality, diversity, and scale of the data used during training and evaluation. Large-scale datasets such as EyePACS enable the development of robust and generalizable models by providing extensive variability in imaging conditions and disease severity. In contrast, smaller datasets like IDRiD offer detailed pixel-level annotations that are particularly valuable for tasks such as lesion segmentation and interpretability.However, no single dataset is sufficient to address all aspects of DR detection, including classification, grading, and lesion localization. Each dataset presents its own strengths and limitations, such as class imbalance, limited diversity, or restricted accessibility. As a result, combining multiple datasets has emerged as a promising approach to enhance data diversity and improve model performance. Nevertheless, this approach introduces challenges such as domain shift and inconsistencies in labeling standards.Future research should focus on developing large-scale, standardized, and well-annotated datasets that incorporate diverse populations and imaging conditions. Such advancements will be critical for improving the reliability, fairness, and clinical applicability of deep learning-based DR detection systems in real-world healthcare settings.

REFERENCES

[1] Bhulakshmi, D., & Rajput, D. S. (2024). A systematic review on diabetic retinopathy detection and classification based on deep learning techniques. PeerJ Computer Science. (PeerJ)
[2] Sebastian, A. (2023). A survey on deep-learning-based diabetic retinopathy diagnosis from fundus images. Diagnostics. (MDPI)
[3] Bappi, M. D. I. et al. (2025). Deep learning-based diabetic retinopathy recognition and grading: Challenges and gaps. ICT Express. (ScienceDirect)
[4] Gong, W. et al. (2025). Deep learning for enhanced prediction of diabetic retinopathy. Frontiers in Medicine. (PMC)
[5] Dejene, F. M. (2025). Machine learning and deep learning in diabetic retinopathy screening: A review. (PMC)
[6] Dai, L. et al. (2024). DeepDR Plus: Predicting DR progression using large-scale datasets. Nature Medicine. (Nature)
[7] Brant, A. et al. (2025). Performance evaluation of deep learning DR screening across multi-site datasets. (PMC)
[8] Akhtar, S. et al. (2025). Deep learning-based DR grading using Kaggle datasets. Scientific Reports. (Nature)
[9] Mutawa, A. M. et al. (2024). Deep learning model for DR detection using EyePACS and Messidor datasets. Applied Sciences. (MDPI)
[10] Karthik, S. A. et al. (2025). Early detection and severity classification of DR using combined datasets. (Springer Link)
[11] Chakour, E. M. et al. (2025). Mobile-based DR detection using APTOS and EyePACS datasets. (ScienceDirect)
[12] Sen, C. et al. (2025). Deep learning model using data augmentation for DR datasets. (ScienceDirect)
[13] Nadda, R. et al. (2025). Ensemble learning for DR detection using retinal datasets. (Springer Link)
[14] Mofreh, A. (2025). CNN-based DR detection using fundus datasets. (erurj.journals.ekb.eg)
[15] Sabo, A. G. (2024). Scalable deep learning system for DR detection. (Cureus Journals)
[16] Shukla, A. et al. (2024). HybridFusionNet: Vision transformer-based DR detection. (MDPI)
[17] ResearchGate Study (2025). Multimodal data fusion approaches in DR detection. (ResearchGate)
[18] VR-FuseNet (2025). Fusion of heterogeneous DR datasets for classification. (arXiv)
[19] DRAC Challenge (2023). Ultra-wide OCTA dataset for DR analysis. (arXiv)
[20] DRStageNet Study (2023). Multi-dataset domain adaptation for DR detection. (arXiv)
[21] Al-Kamachy, I. et al. (2024). Pre-trained models for DR classification using Kaggle datasets. (arXiv)
[22] Bappi et al. (2025). Benchmarking DR datasets with attention-based models. (ScienceDirect)
[23] Waboke, W. R. (2025). Analysis of commonly used DR datasets in deep learning. (slujst.com.ng)
[24] Mushtaq, G. et al. (2021/2023 cited works). Deep learning methodologies for DR detection. (Sage Journals)
[25] IJSDR Study (2025). DR detection using Kaggle dataset and CNN models. (ijsdr.org)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)