

# State-of-the-Art Deep Learning Methodologies for the Detection, Segmentation, and Grading of Diabetic Macular Edema (DME): A Comprehensive Multi-Decadal Survey

Kiran Kadakuntla<sup>1</sup>

<sup>1</sup> Lecturer, Government Polytechnic, Karnataka, India

<sup>1</sup> Research Scholar, Department of E&C Engineering, SDM CET,  
Dharwad, Karnataka, India

## 1. INTRODUCTION

Diabetes mellitus is a rapidly increasing chronic disease worldwide and is associated with several complications, including diabetic retinopathy (DR), a major cause of vision impairment and blindness. Among its complications, Diabetic Macular Edema (DME) is one of the most critical causes of vision loss in diabetic patients. It occurs due to leakage of fluid from damaged retinal blood vessels into the macula, leading to swelling and progressive loss of central vision. Since early stages of DME may be asymptomatic, timely detection is essential to prevent irreversible visual damage.

Traditionally, DME diagnosis is performed using retinal fundus images and Optical Coherence Tomography (OCT) through manual examination by ophthalmologists. Although clinically reliable, this process is time-consuming, subjective, and challenging to scale due to increasing patient load, especially in resource-limited settings.

To overcome these limitations, automated image analysis methods have been widely explored. Early approaches relied on handcrafted features and conventional machine learning techniques to detect lesions such as exudates and retinal thickening. However, their performance was limited by dependency on feature design and image variability.

The emergence of deep learning has significantly improved retinal disease analysis. Convolutional Neural Networks (CNNs) can automatically learn hierarchical feature representations from images, eliminating the need for manual feature extraction. Recent methods using CNNs, attention mechanisms, U-Net variants, transformer-based models, and hybrid architectures have achieved improved performance in DME detection and grading.

In recent years, research in deep learning-based DME analysis has expanded rapidly, covering tasks such as detection, segmentation, and severity classification using diverse datasets and evaluation protocols. However, the diversity of methodologies makes it difficult to obtain a unified view of progress in this domain.

This review provides a comprehensive overview of deep learning approaches for DME detection, segmentation, and grading. It summarizes the evolution from traditional machine learning to advanced deep learning frameworks, discusses commonly used datasets and evaluation metrics, and highlights emerging trends such as explainable AI, self-supervised learning, federated learning, and foundation models. The study aims to provide a consolidated understanding of current progress, limitations, and future research directions in automated DME diagnosis.

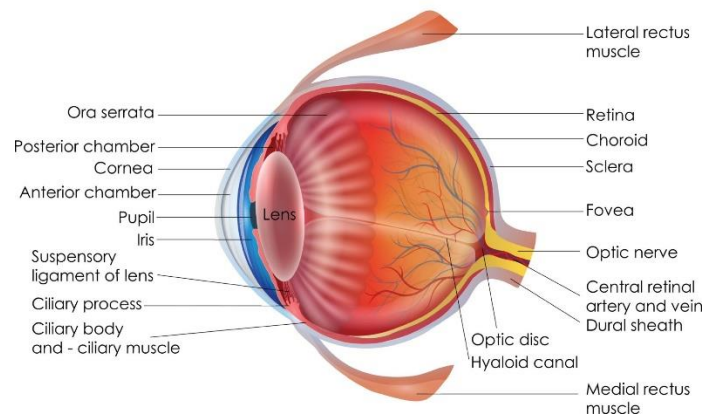
## 2. BACKGROUND AND CLINICAL OVERVIEW OF DIABETIC MACULAR EDEMA

### 2.1 Anatomy of the Retina

The retina is a thin, light-sensitive neural tissue located at the posterior segment of the eye. It plays a vital role in vision by converting incoming light into electrical signals that are transmitted to the brain through the optic nerve. The retina consists of multiple layers and specialized structures that collectively support visual perception.

Among these structures, the macula is responsible for central vision and enables activities requiring fine visual detail, such as reading, driving, and facial recognition. At the center of the macula lies the fovea, which contains the highest concentration of photoreceptor cells and provides maximum visual acuity. The retina also contains an extensive vascular network that supplies oxygen and nutrients to retinal tissues while maintaining the integrity of the blood-retinal barrier.

## Anatomy of the Human Eye



**Figure 1.** Anatomical structure of the retina highlighting the optic disc, macula, fovea, retinal blood vessels, and peripheral retina.

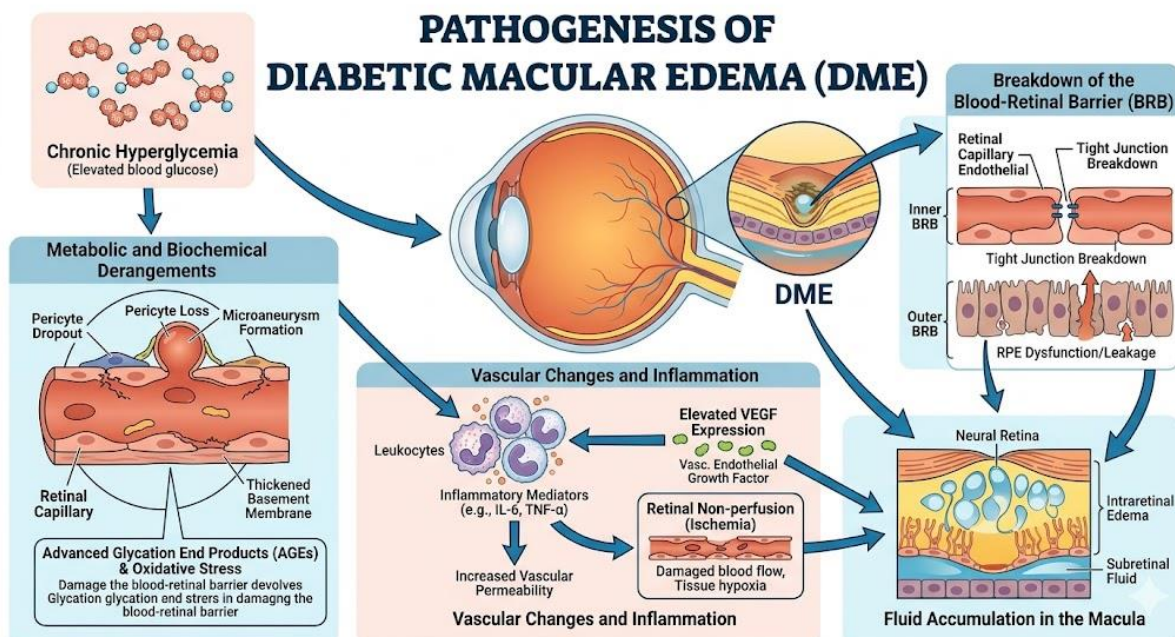
Any disruption to the retinal vascular system due to prolonged hyperglycemia can lead to pathological changes, including diabetic retinopathy and diabetic macular edema.

### 2.2 Diabetic Retinopathy and Development of DME

Diabetic retinopathy (DR) is a microvascular complication of diabetes mellitus caused by prolonged elevation of blood glucose levels. Chronic hyperglycemia damages retinal capillaries, resulting in increased vascular permeability, microaneurysm formation, hemorrhages, and ischemic changes within retinal tissues.

One of the most vision-threatening manifestations of diabetic retinopathy is Diabetic Macular Edema (DME). DME occurs when the blood-retinal barrier becomes compromised, allowing plasma fluid and lipoproteins to leak into the macular region. This leakage causes retinal thickening and swelling of the macula, ultimately affecting central vision.

Unlike several other retinal abnormalities that primarily occur in advanced diabetic retinopathy, DME may develop at various stages of the disease and can significantly impair vision if left untreated. Therefore, early diagnosis and continuous monitoring are essential for preventing irreversible vision loss.



**Figure 2.** Pathogenesis of diabetic macular edema illustrating the progression from diabetes mellitus to retinal vascular damage, fluid leakage, retinal thickening, and DME formation.

### 2.3 Clinical Significance of DME

DME is one of the leading causes of vision impairment among working-age adults worldwide. In its early stages, patients may experience few or no noticeable symptoms. As the disease progresses, visual disturbances such as blurred vision, reduced contrast sensitivity, image distortion, and central vision loss become increasingly apparent.

Early detection of DME is critical because timely therapeutic interventions can significantly reduce the risk of severe vision loss. Current treatment strategies include anti-vascular endothelial growth factor (anti-VEGF) therapy, corticosteroid injections, and laser photocoagulation. The success of these treatments largely depends on the stage at which the disease is diagnosed.

Consequently, effective screening and monitoring programs are essential for improving patient outcomes and reducing the burden of diabetes-related blindness.

### 2.4 Clinical Classification and Grading of DME

The severity of DME is commonly classified according to the location and extent of retinal thickening and hard exudates relative to the macular center. Clinical grading systems help ophthalmologists evaluate disease severity and determine appropriate treatment strategies.

Generally, DME is categorized into three levels:

- **Mild DME:** Retinal abnormalities located away from the macular center.
- **Moderate DME:** Lesions approaching the macular center.
- **Severe DME:** Retinal thickening or exudates involving the macular center and posing a substantial risk to vision.

Automated grading systems developed using artificial intelligence aim to mimic this clinical assessment process and provide consistent severity classification.

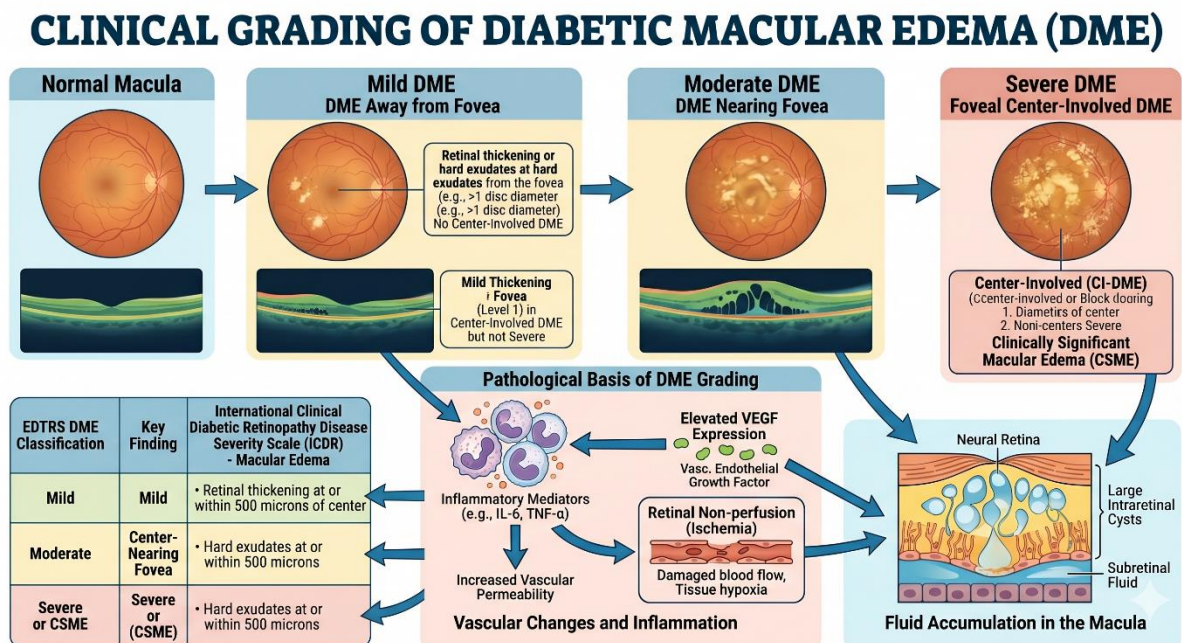


Figure 3. Clinical grading categories of diabetic macular edema showing mild, moderate, and severe disease progression.

### 2.5 Retinal Imaging Modalities for DME Assessment

Advances in ophthalmic imaging technologies have greatly enhanced the diagnosis and management of DME. Retinal fundus photography remains one of the most widely adopted imaging techniques due to its affordability, accessibility, and suitability for large-scale screening programs. Fundus images provide valuable information regarding retinal lesions such as hard exudates, microaneurysms, and hemorrhages.

Optical Coherence Tomography (OCT) has emerged as the gold standard for DME assessment because it provides high-resolution cross-sectional images of retinal layers. OCT enables direct visualization of fluid accumulation and retinal thickening, making it highly effective for disease diagnosis and progression monitoring.

Retinal thickness is an important biomarker used to assess DME severity and is commonly calculated as:

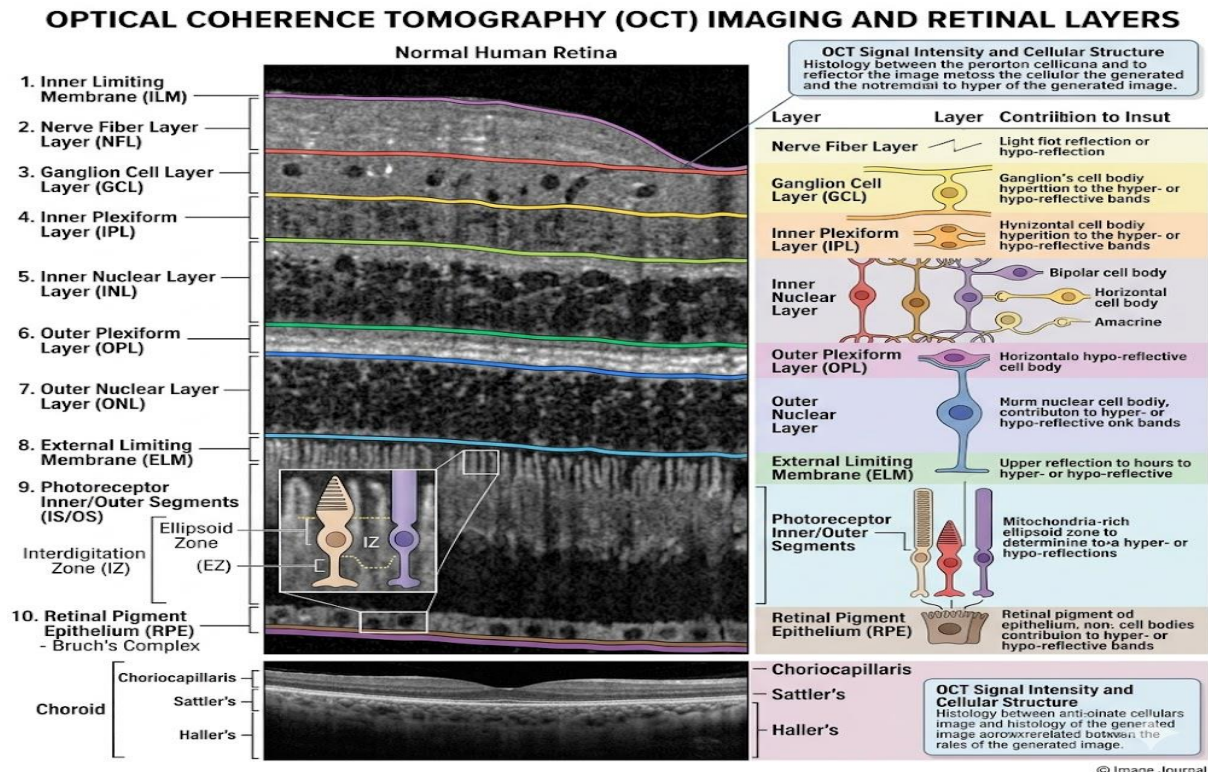
**Equation (1): Retinal Thickness Measurement**

$$RT = ILM - RPE$$

where:

- RT denotes retinal thickness,
- ILM represents the Internal Limiting Membrane,
- RPE represents the Retinal Pigment Epithelium.

This measurement quantifies the distance between the inner and outer retinal boundaries and is frequently used in OCT-based analysis.



**Figure 4.** Example OCT image illustrating retinal layers and retinal thickness measurement between the ILM and RPE boundaries.

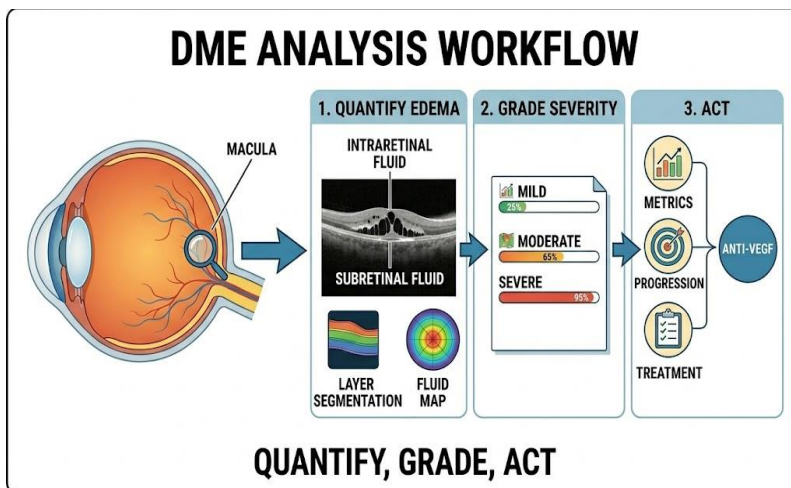
Recent studies have also explored multimodal imaging approaches that combine fundus photography, OCT, and OCT angiography (OCTA) to provide complementary structural and vascular information for comprehensive DME assessment.

**2.6 Need for Automated DME Analysis**

The rapid growth of the diabetic population has substantially increased the demand for retinal screening services. Manual interpretation of retinal images requires specialized expertise and can be both time-consuming and resource-intensive. Additionally, variations in clinical experience may introduce inconsistencies in diagnosis and grading.

Automated image analysis systems powered by artificial intelligence offer a promising solution to these challenges. By leveraging advanced deep learning techniques, these systems can automatically detect pathological features, segment retinal lesions, grade disease severity, and assist clinicians in making informed treatment decisions.

A typical automated DME analysis framework consists of image acquisition, preprocessing, feature extraction, deep learning-based analysis, and clinical decision support.



**Figure 5.** General workflow of an automated DME analysis system showing image acquisition, preprocessing, deep learning analysis, and clinical decision support.

The growing success of deep learning models in retinal image analysis has led to significant improvements in the accuracy and reliability of automated DME diagnosis. Consequently, deep learning has become a major research focus in the development of next-generation ophthalmic screening systems.

### 2.7 Quantitative Measures Used in DME Assessment

The assessment of Diabetic Macular Edema (DME) relies on both clinical observations and quantitative measurements obtained from retinal imaging modalities. Various mathematical metrics are used to evaluate retinal thickness, disease severity, lesion segmentation performance, and classification accuracy. These quantitative measures provide objective criteria for disease diagnosis, progression monitoring, and evaluation of automated computer-aided diagnostic systems.

#### 2.7.1 Retinal Thickness Measurement

Retinal thickness is one of the most important biomarkers used in Optical Coherence Tomography (OCT)-based DME assessment. Increased retinal thickness is generally associated with fluid accumulation and macular swelling.

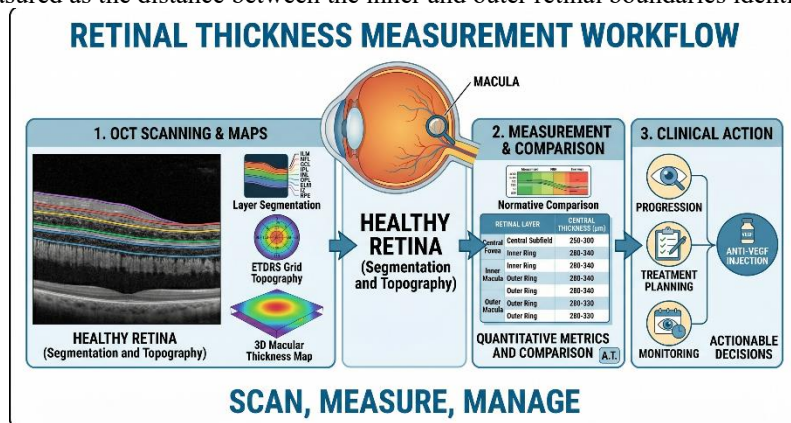
#### Equation (1): Retinal Thickness

$$RT = ILM - RPE$$

Where:

- $RT$  represents retinal thickness.
- $ILM$  denotes the Internal Limiting Membrane.
- $RPE$  denotes the Retinal Pigment Epithelium.

The retinal thickness is measured as the distance between the inner and outer retinal boundaries identified in OCT scans.



**Figure 6.** OCT-based retinal thickness measurement between the Internal Limiting Membrane (ILM) and Retinal Pigment Epithelium (RPE).

#### 2.7.2 Accuracy

Accuracy is widely used to evaluate the overall performance of DME classification models.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- $TP$ = True Positives
- $TN$ = True Negatives
- $FP$ = False Positives
- $FN$ = False Negatives

Accuracy indicates the proportion of correctly classified retinal images among all examined samples.

### 2.7.3 Sensitivity

Sensitivity, also known as Recall or True Positive Rate, measures the ability of a model to correctly identify DME cases.

$$Sensitivity = \frac{TP}{TP + FN}$$

A higher sensitivity value indicates that fewer diseased cases are missed, which is particularly important in medical screening applications.

### 2.7.4 Specificity

Specificity measures the ability of a model to correctly identify healthy or non-DME cases.

$$Specificity = \frac{TN}{TN + FP}$$

High specificity reduces the number of false alarms and unnecessary clinical referrals.

### 2.7.5 Dice Similarity Coefficient

The Dice Similarity Coefficient (DSC) is one of the most frequently used metrics for evaluating lesion segmentation performance in retinal images.

$$Dice = \frac{2 | A \cap B |}{| A | + | B |}$$

Where:

- $A$  represents the ground truth segmentation.
- $B$  represents the predicted segmentation.

The Dice coefficient ranges from 0 to 1, with values closer to 1 indicating better overlap between the predicted and actual lesion regions.

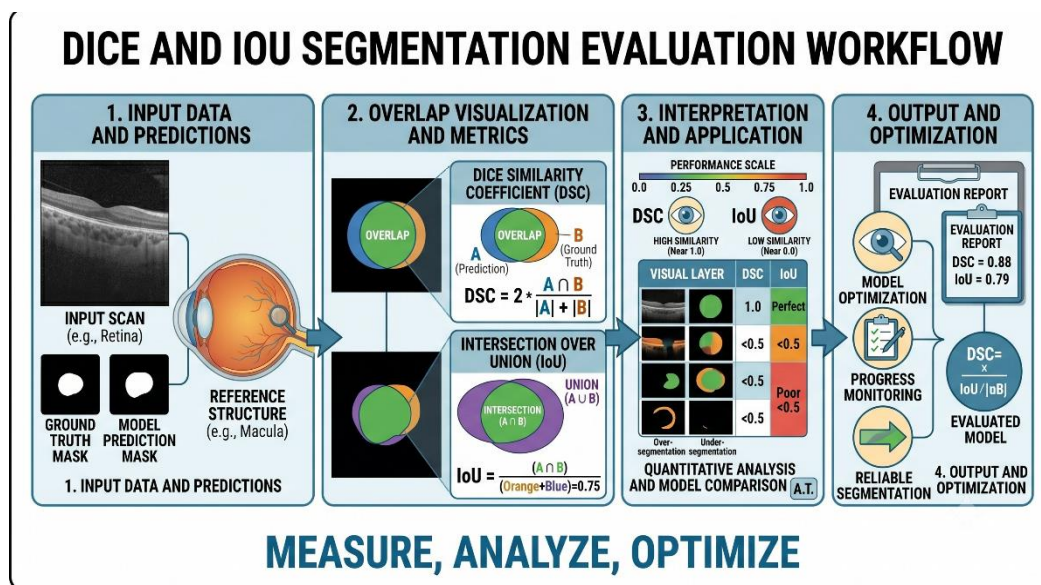


Figure 7. Illustration of Dice Similarity Coefficient showing overlap between ground truth and predicted lesion regions.

### 2.7.6 Intersection over Union (IoU)

Intersection over Union (IoU), also known as the Jaccard Index, is another widely used segmentation evaluation metric.

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

Where:

- $A$  denotes the ground truth lesion area.

- $B$  denotes the predicted lesion area.

Higher IoU values indicate more accurate lesion localization and segmentation performance.

### 3. EVOLUTION OF AUTOMATED DME ANALYSIS: FROM TRADITIONAL MACHINE LEARNING TO DEEP LEARNING

The increasing prevalence of diabetes has led to growing demand for automated retinal screening systems for Diabetic Macular Edema (DME). Over the past two decades, DME analysis has evolved from traditional image processing techniques to advanced deep learning-based systems capable of near-expert performance. This progression highlights the improvements in accuracy, robustness, and clinical applicability over time.

#### 3.1 Traditional Image Processing Approaches

Early DME analysis methods relied on classical image processing techniques to enhance retinal images and detect lesions such as hard exudates and hemorrhages. Common operations included preprocessing steps like contrast enhancement, noise reduction, histogram equalization, and color space conversion, followed by rule-based lesion detection using thresholding, edge detection, morphological operations, and clustering.

These methods were simple and computationally efficient but heavily dependent on handcrafted rules and image quality, resulting in poor generalization across diverse clinical datasets.

#### 3.2 Machine Learning-Based DME Detection

With the availability of retinal datasets, machine learning approaches were introduced. These systems followed a pipeline of preprocessing, handcrafted feature extraction, feature selection, and classification. Features such as texture descriptors, shape features, wavelet transforms, LBP, GLCM, and HOG were widely used.

Classifiers including SVM, KNN, Random Forest, Decision Trees, and ANN improved performance compared to rule-based methods. However, their effectiveness remained limited by the quality of manually engineered features.

#### 3.3 Emergence of Deep Learning

Deep learning significantly transformed retinal image analysis by enabling automatic feature learning from raw data. This shift was supported by large datasets, GPU computing, and improved neural network designs. Among deep learning models, Convolutional Neural Networks (CNNs) became the most widely used due to their ability to learn hierarchical representations from retinal images.

A typical CNN learns progressively complex features ranging from edges and textures to disease-specific patterns through convolutional, pooling, activation, and fully connected layers.

#### 3.4 CNN-Based DME Detection Systems

Between 2015 and 2020, CNN-based models became the dominant approach for DME detection. Architectures such as AlexNet, VGGNet, GoogLeNet, ResNet, DenseNet, and Inception were widely adopted using both custom designs and transfer learning.

Transfer learning was especially effective for medical imaging due to limited datasets, allowing pre-trained networks to be fine-tuned for retinal disease classification with improved accuracy.

#### 3.5 Deep Learning for DME Segmentation

Segmentation models play a key role in identifying lesion regions such as fluid accumulation and exudates at pixel level. Encoder-decoder architectures, particularly U-Net and its variants (U-Net++, Attention U-Net, Residual U-Net, SegNet, and DeepLab), have shown strong performance in retinal lesion localization and quantitative assessment.

#### 3.6 Attention Mechanisms and Hybrid Networks

To improve focus on clinically relevant regions, attention mechanisms were introduced into CNN-based models. These mechanisms enhance feature weighting for key lesion areas such as exudates, macular edema, and abnormal retinal structures. Hybrid CNN-attention models have improved both accuracy and interpretability.

#### 3.7 Transformer-Based DME Analysis

Recently, transformer architectures have been applied to retinal imaging. Vision Transformer (ViT), Swin Transformer, TransUNet, and SegFormer use self-attention to capture global contextual relationships across images, improving both classification and segmentation performance compared to CNN-only models.

### 3.8 Emerging Trends

Current research is moving toward advanced paradigms such as explainable AI, self-supervised learning, federated learning, vision-language models, foundation models, and multimodal learning using fundus and OCT images. These approaches aim to improve interpretability, reduce annotation requirements, and enable scalable clinical deployment.

## 4. PUBLICLY AVAILABLE DATASETS AND BENCHMARK RESOURCES FOR DME ANALYSIS

The performance of deep learning models for Diabetic Macular Edema (DME) analysis strongly depends on the availability of high-quality annotated retinal imaging datasets. These datasets form the basis for training, validation, and benchmarking of automated diagnostic systems. Over time, several public datasets have been developed for diabetic retinopathy (DR), DME detection, lesion segmentation, and disease grading. However, variations in imaging protocols, annotations, and patient populations often affect model generalization, making dataset selection an important factor in research evaluation.

### 4.1 Fundus Image Datasets

Color fundus photography is widely used for large-scale DME screening due to its cost-effectiveness. Several benchmark datasets support detection and grading tasks.

#### IDRiD-Dataset

The Indian Diabetic Retinopathy Image Dataset (IDRiD) contains high-resolution fundus images with pixel-level annotations for lesions such as microaneurysms, hemorrhages, hard exudates, and soft exudates. It is widely used for segmentation and grading tasks. However, it has a relatively small sample size and class imbalance.

#### MESSIDOR-Dataset

MESSIDOR is one of the earliest and most widely used datasets for diabetic eye disease research. It contains fundus images with multiple severity levels and is commonly used for classification and DR/DME screening. However, it lacks detailed pixel-level annotations, limiting its use for segmentation.

#### DIARETDB1-Dataset

DIARETDB1 is designed for lesion detection and includes expert annotations for diabetic retinopathy abnormalities. It is frequently used for evaluating early detection methods but has limited dataset size and fewer advanced-stage cases.

#### e-Ophtha-Dataset

The e-Ophtha dataset provides lesion-specific annotations, particularly for microaneurysms and exudates, making it useful for lesion detection and segmentation tasks in DME research.

#### DDR-Dataset

The DeepDR (DDR) dataset is a large-scale dataset containing fundus images across multiple severity levels and diverse populations. It is widely used for training deep learning models and evaluating generalization performance.

### 4.2 OCT-Based Datasets

Optical Coherence Tomography (OCT) provides cross-sectional retinal imaging and is considered the gold standard for DME diagnosis due to its ability to visualize retinal thickness and fluid accumulation.

OCT datasets offer high-resolution structural information and enable accurate disease severity assessment. However, their usage is limited by data scarcity, annotation complexity, and high storage requirements.

### 4.3 Multimodal Retinal Imaging Datasets

Recent research has focused on multimodal datasets combining fundus images, OCT scans, OCT angiography, and clinical metadata. These datasets improve diagnostic accuracy and disease characterization by integrating complementary information.

Despite their advantages, multimodal datasets are difficult to construct due to high cost, complex annotation requirements, and limited availability.

### 4.4 Dataset Challenges in DME Research

Despite the availability of several benchmark datasets, key challenges remain:

- **Class imbalance:** Fewer DME-positive samples compared to normal cases lead to biased learning.
- **Limited annotations:** Pixel-level labeling requires expert ophthalmologists and is time-consuming.
- **Dataset heterogeneity:** Differences in imaging devices, illumination, and populations reduce model generalization.
- **Small dataset size:** Retinal datasets are relatively small, increasing overfitting risk in deep models.

- **Lack of standard protocols:** Inconsistent evaluation methods make cross-study comparison difficult.

#### 4.5 Comparative Analysis of Publicly Available Datasets

To facilitate dataset selection for future research, Table 1 summarizes the most widely used datasets in DME analysis.

Dataset	Imaging Modality	Number of Images	Annotation Type	Application
IDRiD	Fundus	58	Lesion Masks	Detection, Segmentation
MESSIDOR	Fundus	862	Disease Grades	Classification
DIARETDB1	Fundus	421	Lesion Labels	Detection
e-Ophtha	Fundus	150	Exudate Annotations	Segmentation
DDR	Fundus	52	Severity Grades	Classification
OCT Datasets	OCT	156	Layer Annotations	DME Assessment

**Table 1.** Comparative summary of publicly available retinal imaging datasets used for DME detection, segmentation, and grading.

The availability of publicly accessible retinal imaging datasets has played a crucial role in advancing automated DME analysis. These datasets have enabled researchers to develop increasingly sophisticated machine learning and deep learning models while providing standardized benchmarks for performance evaluation. Nevertheless, challenges related to data quality, annotation availability, and dataset diversity continue to motivate the development of larger, more representative datasets for future research.

### 5. DEEP LEARNING METHODOLOGIES FOR DME DETECTION

The introduction of deep learning has significantly improved the automated detection of Diabetic Macular Edema (DME). Unlike traditional machine learning approaches that rely on handcrafted features, deep learning models can automatically learn complex and discriminative representations directly from retinal images. This capability has enabled more accurate identification of pathological features associated with DME, leading to substantial improvements in diagnostic performance.

#### 5.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) were among the first deep learning architectures successfully applied to retinal image analysis. CNNs learn hierarchical image features through convolutional operations, allowing them to detect retinal abnormalities such as hard exudates and macular changes associated with DME.

The convolution operation is defined as:

$$F(i, j) = (I * K)(i, j)$$

where  $I$  represents the input image,  $K$  denotes the convolution kernel, and  $F(i, j)$  is the resulting feature map.

Popular CNN architectures used in DME studies include AlexNet, VGGNet, ResNet, DenseNet, and Inception networks.

#### 5.2 Transfer Learning Approaches

Due to the limited availability of annotated retinal datasets, transfer learning has become a widely adopted strategy in DME detection. In this approach, models pre-trained on large image datasets are fine-tuned using retinal images, reducing training time and improving performance.

Commonly used pre-trained models include ResNet, DenseNet, EfficientNet, and InceptionV3.

#### 5.3 Attention-Based Models

Attention mechanisms enable deep learning models to focus on clinically relevant retinal regions while reducing the influence of background information. By highlighting important lesion areas, attention-based models often achieve improved detection accuracy and interpretability.

#### 5.4 Transformer-Based Architectures

Recent advances in computer vision have led to the adoption of transformer-based models for retinal image analysis. Unlike CNNs, transformers utilize self-attention mechanisms to capture long-range dependencies and global contextual information across the entire image.

Popular architectures include Vision Transformer (ViT), Swin Transformer, and hybrid CNN-transformer models.

#### 5.5 Explainable Artificial Intelligence

Although deep learning models achieve high diagnostic accuracy, their decision-making process is often difficult to interpret. Explainable Artificial Intelligence (XAI) techniques such as Grad-CAM, SHAP, and LIME help visualize the retinal regions influencing model predictions, thereby increasing clinician trust and supporting clinical adoption.

## 5.6 Summary

The evolution of deep learning has transformed automated DME detection from conventional CNN-based systems to sophisticated attention and transformer-based architectures. These advancements have improved detection accuracy, robustness, and clinical applicability. Nevertheless, challenges related to data availability, model interpretability, and cross-dataset generalization continue to drive ongoing research in this area.

Method	Key Advantage	Limitation
CNN	Automatic feature extraction	Limited global context
Transfer Learning	Effective with small datasets	Domain dependency
Attention Networks	Better lesion localization	Increased complexity
Transformers	Global feature learning	High computational cost
XAI Methods	Improved interpretability	Additional processing

**Table 2.** Summary of major deep learning methodologies used for DME detection.

## 6. Deep Learning Methodologies for DME Segmentation

DME segmentation aims to accurately identify and delineate retinal lesions such as hard exudates and fluid accumulation regions. Accurate segmentation provides valuable information regarding lesion location, extent, and disease progression. Deep learning architectures, particularly U-Net and its variants, have become the dominant approaches for automated retinal lesion segmentation due to their ability to perform pixel-level classification.

### 6.1 U-Net and Advanced Segmentation Models

U-Net, U-Net++, Attention U-Net, and transformer-based models such as TransUNet and Swin-UNet are widely used for DME segmentation. These architectures combine feature extraction and spatial localization to achieve accurate lesion boundary detection.

### 6.2 Segmentation Evaluation Metrics

The performance of segmentation models is commonly evaluated using overlap-based metrics.

#### Dice Similarity Coefficient

$$Dice = \frac{2 | A \cap B |}{| A | + | B |}$$

where:

- $A$  = Ground truth segmentation
- $B$  = Predicted segmentation

A higher Dice score indicates better overlap between the predicted and actual lesion regions.

#### Intersection over Union (IoU)

$$IoU = \frac{| A \cap B |}{| A \cup B |}$$

where:

- $A$  = Ground truth lesion area
- $B$  = Predicted lesion area

IoU measures the ratio between the intersection and union of segmented regions.

#### Segmentation Loss Function

A commonly used loss function in DME segmentation is Dice Loss:

$$L_{Dice} = 1 - \frac{2 | A \cap B |}{| A | + | B |}$$

Minimizing Dice Loss improves segmentation accuracy by maximizing overlap with the ground truth.

## 6.3 Summary

Recent advances in deep learning have significantly improved DME lesion segmentation. U-Net-based architectures remain the most widely adopted models, while attention and transformer-based approaches continue to enhance segmentation accuracy and robustness.

Model	Application	Common Metrics
U-Net	Lesion Segmentation	Dice, IoU
U-Net++	Exudate Segmentation	Dice, IoU
Attention U-Net	Lesion Localization	Dice, IoU
TransUNet	DME Segmentation	Dice, IoU
Swin-UNet	Retinal Segmentation	Dice, IoU

**Table 3.** Summary of deep learning segmentation methods used in DME analysis.

## 7. DEEP LEARNING METHODOLOGIES FOR DME GRADING

DME grading focuses on classifying the severity of the disease into different categories, such as mild, moderate, and severe. Accurate grading is essential for clinical decision-making, treatment planning, and patient prioritization. Unlike detection, which determines the presence of DME, grading assesses the extent of disease progression.

Deep learning models, particularly CNNs, attention-based networks, and transformer architectures, have demonstrated promising performance in automatically learning disease-specific patterns from retinal images and assigning appropriate severity grades.

### 7.1 Classification Metrics

The performance of DME grading models is commonly evaluated using classification metrics.

#### Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy measures the proportion of correctly classified samples.

#### Precision

$$Precision = \frac{TP}{TP + FP}$$

Precision indicates the reliability of positive predictions.

#### Recall (Sensitivity)

$$Sensitivity = \frac{TP}{TP + FN}$$

Recall measures the ability of the model to correctly identify DME cases.

#### F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

F1-score provides a balanced assessment of precision and recall.

### 7.2 Loss Function for DME Grading

For multi-class DME grading, Cross-Entropy Loss is commonly employed:

$$L = - \sum_{i=1}^C y_i \log(\hat{y}_i)$$

where:

- $C$  is the number of classes,
- $y_i$  is the true label,
- $\hat{y}_i$  is the predicted probability.

The objective is to minimize the classification error during model training.

### 7.3 Summary

Deep learning-based grading systems have improved the consistency and accuracy of DME severity assessment. Modern architectures can effectively distinguish between different disease stages and support clinicians in making timely treatment decisions.

Method	Application	Common Metrics
--------	-------------	----------------

CNN	Severity Classification	Accuracy, F1
Transfer Learning	Multi-class Grading	Accuracy, Recall
Attention Networks	Severity Assessment	Precision, F1
Transformers	Advanced Grading	Accuracy, AUC

The integration of deep learning into DME grading has enhanced automated disease assessment and laid the foundation for intelligent clinical decision-support systems.

## 8. COMPARATIVE ANALYSIS AND PERFORMANCE EVALUATION OF DEEP LEARNING MODELS

The rapid advancement of deep learning has led to the development of numerous models for DME detection, segmentation, and grading. While most approaches report high performance, direct comparison remains challenging due to differences in datasets, evaluation protocols, and model architectures. Therefore, a comprehensive comparative analysis is essential to identify the strengths and limitations of existing methodologies.

### 8.1 Performance Metrics

Different tasks require different evaluation measures. Detection and grading models are generally evaluated using classification metrics, whereas segmentation models rely on overlap-based measures.

#### Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

#### F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

#### Dice Coefficient

$$Dice = \frac{2 | A \cap B |}{| A | + | B |}$$

#### Intersection over Union (IoU)

$$IoU = \frac{| A \cap B |}{| A \cup B |}$$

### 8.2 Comparative Discussion

CNN-based models provide strong feature extraction capabilities and have achieved high performance in DME detection tasks. Attention-based architectures further improve lesion localization by focusing on clinically relevant retinal regions. More recently, transformer-based models have demonstrated superior capability in capturing global image context, resulting in improved detection and grading performance.

For segmentation tasks, U-Net and its variants remain the most widely adopted architectures due to their simplicity and effectiveness. Transformer-based segmentation models have shown promising results but often require larger datasets and greater computational resources.

Method	Strengths	Limitations
CNN	Efficient feature extraction	Limited global context
Transfer Learning	Works well with small datasets	Domain dependency
Attention Models	Better lesion localization	Increased complexity
U-Net Variants	Accurate segmentation	Limited contextual awareness
Transformers	Global feature representation	High computational cost

**Table 5.** Comparative analysis of deep learning methodologies for DME analysis.

### 8.3 Summary

Overall, deep learning has significantly improved the accuracy and reliability of automated DME analysis. While CNN-based approaches continue to serve as strong baselines, attention and transformer-based architectures are increasingly becoming the

preferred choice for advanced retinal image analysis. The selection of an appropriate model ultimately depends on the available dataset, computational resources, and clinical application requirements.

## 9. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

Despite the remarkable progress achieved by deep learning in DME detection, segmentation, and grading, several challenges continue to limit its widespread clinical adoption. Addressing these issues is essential for developing reliable and scalable diagnostic systems.

### 9.1 Current Challenges

#### Limited Annotated Data

Deep learning models require large amounts of labeled data, but obtaining expert annotations for retinal images is expensive and time-consuming.

#### Class Imbalance

Many datasets contain significantly fewer DME-positive samples than normal images, which may lead to biased model performance.

#### Lack of Explainability

Most deep learning models operate as black boxes, making it difficult for clinicians to understand the reasoning behind predictions.

#### Generalization Issues

Models trained on one dataset often experience performance degradation when applied to images acquired from different devices or populations.

#### Computational Complexity

Advanced architectures such as transformers require substantial computational resources, limiting their deployment in resource-constrained healthcare settings.

### 9.2 Future Research Directions

#### Explainable Artificial Intelligence (XAI)

Techniques such as Grad-CAM, SHAP, and LIME can improve model transparency and increase clinician trust in AI-assisted diagnosis.

#### Federated Learning

Federated learning enables collaborative model training across multiple healthcare institutions while preserving patient privacy.

#### Self-Supervised Learning

Self-supervised approaches can reduce the dependence on large annotated datasets by learning meaningful representations from unlabeled retinal images.

#### Multimodal Learning

Combining fundus photographs, OCT images, and clinical information can provide a more comprehensive understanding of disease progression.

#### Foundation Models

Large-scale vision foundation models have the potential to improve generalization and support multiple retinal analysis tasks within a single framework.

#### Lightweight Models

Developing efficient models for mobile and edge devices can facilitate real-time DME screening in remote and underserved regions.

### 9.3 Summary

Future research should focus on developing interpretable, data-efficient, and clinically deployable AI systems. The integration of explainable AI, federated learning, multimodal imaging, and foundation models is expected to drive the next generation of intelligent DME diagnostic solutions and improve accessibility to retinal healthcare worldwide.

## 10. CONCLUSION

Diabetic Macular Edema remains one of the leading causes of vision impairment among individuals with diabetes, highlighting the need for timely and accurate diagnosis. This review surveyed the evolution of automated DME analysis from traditional image processing techniques to modern deep learning methodologies for detection, segmentation, and grading.

Deep learning architectures, particularly CNNs, attention-based networks, U-Net variants, and transformer models, have significantly improved diagnostic performance and automated retinal image analysis. Publicly available datasets and advances in computational resources have further accelerated research in this field.

Despite these achievements, challenges related to data availability, model interpretability, generalization, and computational requirements remain. Emerging technologies such as explainable AI, federated learning, self-supervised learning, multimodal systems, and foundation models offer promising directions for overcoming these limitations.

Overall, deep learning continues to transform DME diagnosis and management, providing new opportunities for early screening, precise disease assessment, and improved clinical decision support. Continued research in these areas is expected to contribute toward more accurate, reliable, and accessible ophthalmic healthcare solutions.

## REFERENCES

- [1] Early Treatment Diabetic Retinopathy Study Research Group, "Photocoagulation for diabetic macular edema: Early Treatment Diabetic Retinopathy Study report no. 1," *Archives of Ophthalmology*, vol. 103, no. 12, pp. 1796–1806, 1985.
- [2] Early Treatment Diabetic Retinopathy Study Research Group, "Early photocoagulation for diabetic retinopathy: ETDRS report no. 9," *Ophthalmology*, vol. 98, no. 5, pp. 766–785, 1991.
- [3] R. Klein, B. E. Klein, S. E. Moss, M. D. Davis, and D. L. DeMets, "The Wisconsin epidemiologic study of diabetic retinopathy," *Archives of Ophthalmology*, vol. 102, no. 4, pp. 520–526, 1984.
- [4] J. W. Yau et al., "Global prevalence and major risk factors of diabetic retinopathy," *Diabetes Care*, vol. 35, no. 3, pp. 556–564, 2012.
- [5] T. Y. Wong and C. Sabanayagam, "Strategies to tackle the global burden of diabetic retinopathy: From epidemiology to artificial intelligence," *Ophthalmologica*, vol. 243, no. 1, pp. 9–20, 2020.
- [6] A. Sopharak, B. Uyyanonvara, and S. Barman, "Automatic exudate detection from non-dilated diabetic retinopathy retinal images using mathematical morphology methods," *Computerized Medical Imaging and Graphics*, vol. 32, no. 8, pp. 720–727, 2008.
- [7] M. Niemeijer, B. van Ginneken, J. Staal, M. S. A. Suttorp-Schulten, and M. D. Abramoff, "Automatic detection of red lesions in digital color fundus photographs," *IEEE Transactions on Medical Imaging*, vol. 24, no. 5, pp. 584–592, 2005.
- [8] A. D. Fleming et al., "Automated microaneurysm detection using local contrast normalization and local vessel detection," *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1223–1232, 2006.
- [9] D. Sánchez, A. M. López, J. Poza, and R. Hornero, "Retinal image analysis based on texture descriptors for diabetic retinopathy screening," *Medical Engineering & Physics*, vol. 31, no. 6, pp. 745–752, 2009.
- [10] M. D. Abramoff, M. K. Garvara, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010.
- [11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 1097–1105.
- [12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. International Conference on Learning Representations (ICLR)*, 2015.
- [13] C. Szegedy et al., "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.
- [16] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [17] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, 2017.
- [18] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia Computer Science*, vol. 90, pp. 200–205, 2016.
- [19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.
- [20] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018, pp. 3–11.
- [21] O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.
- [22] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [23] L. C. Chen et al., "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818.
- [24] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7132–7141.
- [25] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.
- [26] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7794–7803.
- [27] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. International Conference on Learning Representations (ICLR)*, 2021.
- [28] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2021, pp. 10012–10022.
- [29] J. Chen et al., "TransUNet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021.
- [30] E. Xie et al., "SegFormer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12077–12090, 2021.

- [31] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626.
- [32] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.
- [33] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2017, pp. 4765–4774.
- [34] B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artificial Intelligence and Statistics (AISTATS)*, 2017, pp. 1273–1282.
- [35] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. International Conference on Machine Learning (ICML)*, 2020, pp. 1597–1607.
- [36] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 9729–9738.
- [37] J. Grill et al., "Bootstrap your own latent: A new approach to self-supervised learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21271–21284, 2020.
- [38] A. Radford et al., "Learning transferable visual models from natural language supervision," in *Proc. International Conference on Machine Learning (ICML)*, 2021.
- [39] A. Kirillov et al., "Segment Anything," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2023, pp. 4015–4026.
- [40] M. D. Abramoff et al., "Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices," *npj Digital Medicine*, vol. 1, no. 39, pp. 1–8, 2018.