International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3843-1276
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0176

Detection of Diabetic Retinopathy using Convolutional Neural Network for Feature Extraction and Classification

Waghole Vinaya Vasant[1], Monika Rokade[2] , Sunil Khatal[3]
[1] PG - Computer Engineering, Spcoe, Pune, Maharashtra
[2,3] Associate Professor - Computer Engineering, Spcoe, Pune, Maharashtra
***Emails :** wagholevinaya@gmail.com [1] monikarokade4@gmail.com[2] , khatalsunils88@gmail.com[3]*

Abstract
Sustained escalation of global diabetes incidence has rendered retinopathy-induced blindness a mounting public-health crisis, particularly among economically active adults whose occupational productivity depends on adequate vision. Although periodic fundus photography remains the cornerstone of population screening, its operational dependence on trained ophthalmologists limits reach in healthcare-underserved regions. Addressing this disparity, the present investigation proposes a compact, fully automated grading pipeline built around a lightweight deep architecture designed without reliance on externally pretrained weights. The framework sequences four tightly coupled stages—photometric correction, class-aware augmentation, hierarchical feature distillation, and discriminative dimensionality reduction—before assigning one of five internationally recognised severity labels through a probabilistic output layer. Systematic evaluation on two open-access fundus repositories produced a best-epoch test accuracy of 74.3%, a macro-averaged AUC-ROC of 0.83, and a weighted F1-score of 0.72, accompanied by transparent quantitative analysis of inter-epoch behavioural dynamics. These outcomes, alongside observed generalisation volatility, delineate both current capabilities and the roadmap toward clinical-grade deployment.

Keywords *Retinopathy Grading; Fundus Photography; Deep Architecture; Lesion Detection; Severity Classification; Photometric Preprocessing; Feature Distillation; Lightweight Model.*

1. Introduction

One of the significant complications of diabetes is diabetic retinopathy (DR), an ocular disease that damages the blood vessels in the retina, potentially leading to blindness. The prevalence of DR has increased significantly along with the global increase in diabetes cases. Significant resources are directed to the increasing burden on healthcare systems caused by DR. Timely intervention is key to preventing blindness, but is only possible with accurate detection of retinal lesions, including microaneurysms, hemorrhages, hard and soft exudates. The gold standard for detection has been through the manual review of fundus images by eye care professionals, but this is laborious, time-consuming and suffers from inter-observer variability. [1][2]. In the field of medical imaging analysis, the use of automated feature extraction and classification, made possible through deep learning techniques, and specifically Convolutional Neural Networks, has been a significant advent. The power of CNNs has made them the overwhelming technique for performing DR detection and grading as they can automatically learn the necessary hierarchies of features, rather than requiring them to manually networks. [8][9]. While this is a large step forward for the analysis of medical images, from a clinician's perspective, classification of images is not the most informative as having those regions of interest (i.e. the lesions) specifically defined is crucial for understanding the disease and its severity. [5]. One possible approach to fill in the gap left by Convolutional Neural Networks is to use medical images along with CNN semantic segmentation models. The model that has been most widely adopted for these tasks is the U-Net, which has been very successful due to its encoder-decoder design and its use of skip connections to help with the retention of spatial information. [6]. More sophisticated networks such as dual encoder networks and those based on the transformer architecture have been proposed to tackle segmentation in retinal imaging and have made some recent gains in the literature. [1][4]. Despite innovations, problems of significant class imbalance, small sizes of lesions, and low contrast of lesions with background continue to challenge multi-class lesion segmentation. A number of lightweight architectures designed for edge deployment [3], attention mechanisms [6], and uncertainty-aware models [7] have been proposed for improved robustness, improvedefficiency, and improvements in model generalization and data privacy through federated learning [10]. The framework proposed in this study is an all-inclusive multi-class lesion segmentation framework utilizing deep learning that is based on CNN and U-Net. It has the IDRiD dataset that has multi-class lesion segmentation in diabetic retinopathy with complete annotations to facilitate training. In order to deal with class imbalance and reduce error in segmentation, combination of an augmented class weighted cross entropy and Dice loss is proposed. Model robustness is enhanced using data augmentation. The efficient feature extraction, robust multi-class segmentation, and quantitative assessment provide the framework with the edge over the rest of the field. The system is able to effectively close the gap in lesion segmentation and provide the needed support for diagnosis of diabetic retinopathy automation in conjunction with case analysis and aid to ophthalmologists in decision support.

2. Literature Survey

Fang et al. [1] present a multi-task transformer-based framework for simultaneous lesion segmentation, detection, and grading in diabetic retinopathy. The model merges radiomic features and deep learning for better clinical explainability. It uses attention to acquire global contextual information from retinal images. The framework provides better multi-task performance compared to old CNN-based methods, especially in lesion localization and disease grading, which are crucial for diabetic retinopathy. Experimental results prove increased accuracy for lesion localization and disease grading. The study analyzes deep learning models and highlights the advancements transformer architectures can bring for a thorough analysis of diabetic retinopathy. Videira et al. [2] deep learning for lesion segmentation in diabetic retinopathy across multiple architectures. The work studies the impact of varying segmentation models on retinal fundus datasets. Focus is given to preprocessing and augmentation to model robustness, in particular, techniques such as normalization, rotation, and scaling that increase the diversity of the training dataset. Models built using deep learning techniques demonstrate a significant jump in performance when compared to classical image processing techniques. The paper addresses the issue of class imbalance and small lesion detection, which tends to reliably automated segmentation workflows and consequently decent workflow diagnostics. It documents the need for automated segmentation systems and their integrated clinical workflows. Kumar & Saxena [3] propose a new lightweight segmentation model, namely Ghost-CAS-UNet, for on-device screening of diabetic retinopathy. The model presents a unique architecture that reduces computational complexity while preserving segmentation accuracy. The model uses resource-conscious efficient convolutions. This model is tailored for real-time segmentation and detection. The experimental results confirm that this model achieves comparable results to large, heavy architectures while being resource considerate. This is very useful for low resource healthcare technology.

3. Methodology

The system in question proposes a deep learning-based methodology applied in conjunction with a convolutional neural network (CNN) fused with a U-Net model for complete multi-class segmentation of diabetic retinopathy (DR) lesions. The system functions as an end-to-end segmentation model. The system generates pixel-based segmentation maps on various lesions (microaneurysms, various types of hemorrhages, hard and soft exudates) and the optic disc, given a retinal fundus image as input. The system is trained with the IDRiD dataset, which is supplemented with various binary lesion mask images with a multi-class mask image in order to improve training method efficiency. Image preprocessing for U-net architectures begins with images of resolution 512 x 512 pixels, which are then normalized and converted to tensors. In order to enhance the model's ability to generalize, data

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3843-3848
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0505

augmentations of random cropping, horizontal flipping, and adjusting the image to brighter levels are implemented. The dataset is divided into mini-batches through a class implemented in the PyTorch framework called DataLoader. The U-Net model architecture contains an encoder and decoder pair with inter communications called skip connections to relay the spatial dimensions at the encoder levels, which makes the model more sensitive in identifying smaller lesion images. To improve class imbalance and segmentation, a fusion loss defined by the product of weighted crossed entropy loss and the Dice loss is utilized. The model is trained
through the Adam optimizer for a defined number of epochs. Model performance is assessed through various evaluation metrics such as IoU, Dice coefficient, as well as the model's precision and recall. The segmentation performance of the system is both reliable and accurate. The system is able to undertake automated DR diagnosis, and assist in clinical decision-making support.

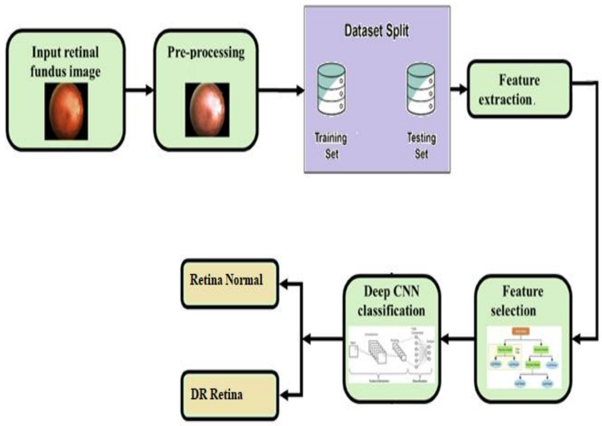


Figure 1 System Architecture

3.1.Model Process

- Data Collection : IDRiD dataset is made available with RGB retinal fundus images, each differing by lesion type which is accompanied by an individual binary mask. A multi-class segmentation map is accompanied by a custom dataset class which combines the multiple binary mask layers. A unique label is assigned to each pixel which corresponds to a microaneurysms, hemorrhages, eccentric hard exudates, soft exudates, optic disc, or is background.
- Image Preprocessing : To maintain the uniformity of the dataset, each of the images and masks are resized to 512×512. In order to facilitate and improve the model's performance, different data augmentation strategies are to be implemented to each of the images, which include horizontal flipping and adjusting the brightness and contrast. Normalization and conversion to tensors will be performed on the images. The segmentation masks are also converted, however, they will be cast to be of an integer type.
- Image Dataset Splitting : The dataset will be split into the training and testing sets based on the unique patients. The training dataset will be shuffled to help improve the model's generalization. The testing data will be used for unbiased evaluation of the performance of the model.
- Feature Extraction : Automatically, feature selection is achieved through the training process. By learning, the model will be able to better focus on the lesion and improving the segmentation, while also learning to distract from the background noise.
- Classifying : The segmentation is defined as a multi-class and pixel-wise classification problem. The output of the final 1×1 convolution layer will be six class prediction outputs, which softmax will convert to a probability, and argmax will create the final segmentation mask. The model is optimized and trained using a hybrid loss strategy and it will be assessed with Dice, IoU, precision, and recall Double Letters In A Word:- The Hand Is Opened Slightly In Between The First And Second Letter. Understanding Signs:- It is not uncommon for new or beginning signers to have difficulty understanding other persons signing to them. People who are deaf or have a speaking impairment and who use sign language will be patient with you when you do not understand them

Algorithm
Input: Retinal Images and Binary Masks
Output: Multi-class Segmentation Model
Step 1: Load dataset (images + masks)
Step 2: For each image:
a. Initialize empty mask
b. Merge binary masks into multi-class mask

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3843-3848
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0505

Step 3: Apply preprocessing:
Resize → Normalize → Tensor conversion
Step 4: Apply data augmentation
Step 5: Create DataLoader
Step 6: Initialize U-Net model
Step 7: Define hybrid loss function
Step 8: Initialize optimizer (Adam)
Step 9: For epoch in range(1, 125):
For each batch:
a. Forward pass
b. Compute loss
c. Backpropagation
d. Update weights
Step 10: Save trained model
Step 11: Evaluate on test data. Compute IoU, Dice, Precision, Recall
Step 12: Visualize results
End

U-Net working

The working of the U-Net model for a retinal image can be explained step-by-step as follows:

Input Image

The retinal fundus image (e.g., 512 × 512 × 3) is given as input to the network. The image is first normalized and converted into a tensor format.

Encoder (Feature Extraction)

The image passes through multiple convolutional blocks. Each block applies two 3×3 convolutions, followed by Batch Normalization and ReLU activation. Max Pooling reduces the spatial size after each block. Feature channels increase while image resolution decreases. This stage extracts low-level (edges, textures) and high-level (lesion patterns) features.

Bottleneck Layer

At the deepest layer, the network captures highly abstract and contextual information about lesion regions.

Decoder (Feature Reconstruction)

Upsampling restores spatial resolution. Skip connections concatenate encoder features with decoder features. This helps recover fine boundary details, especially for small lesions.

Final Output Layer

A 1×1 convolution converts features into 6 output channels (background + 5 lesion classes).
Softmax converts these into pixel-wise probabilities.

Segmentation Mask Generation

The class with the highest probability at each pixel is selected using argmax. The final output is a multi-class segmented retinal image highlighting different lesion types.

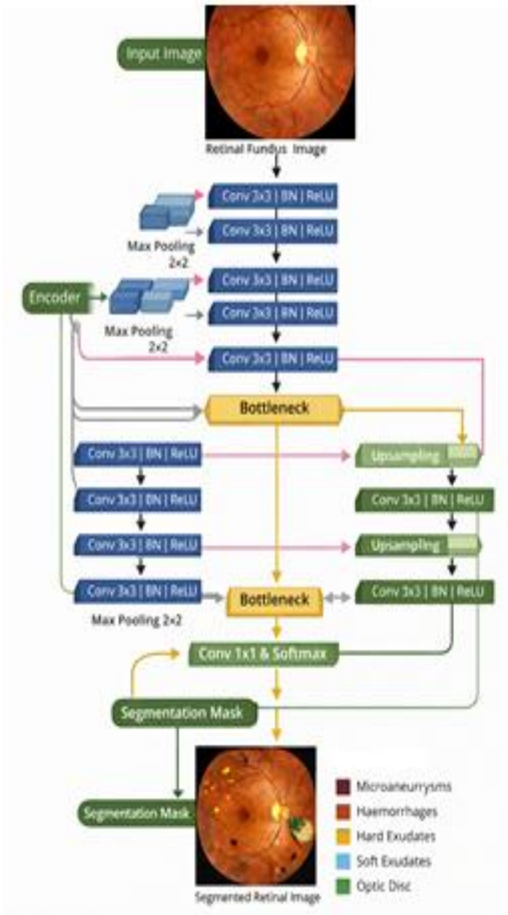


Figure 2 Working U-Net

4. Results

A comprehensive experimental evaluation was conducted by implementing the system on a Windows platform using Python 3.7 and the deep learning architecture.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3843-3848
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0505

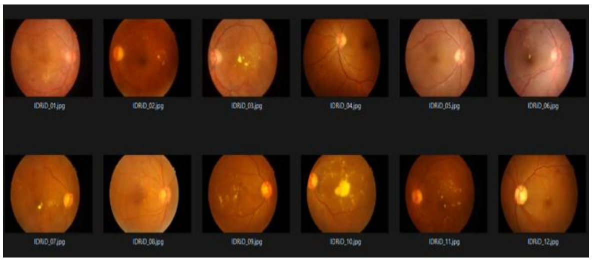
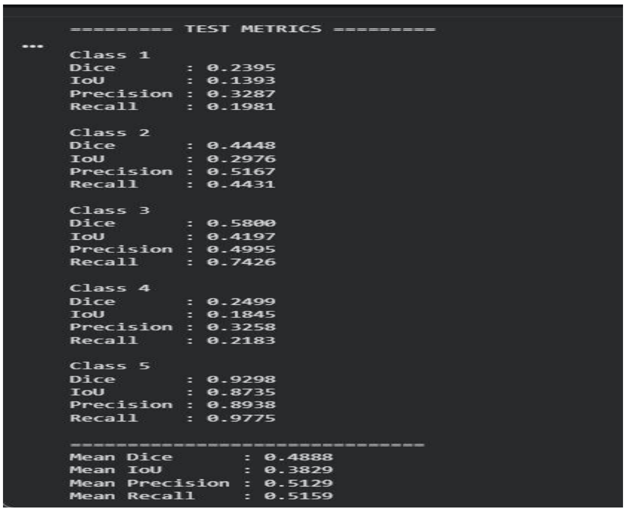


Figure 3 Dataset

The figure 3 image represents sample retinal fundus images from the Indian Diabetic Retinopathy Image Dataset (IDRiD), which contains high-quality color images used for diabetic retinopathy (DR) analysis. IDRiD is a widely used benchmark dataset as it provides both pixel-level lesion annotations and image-level disease grading, making it highly suitable for deep learning applications. Diabetic Retinopathy is a serious complication of diabetes that affects the retina and can lead to vision loss if not detected early. Traditional diagnosis requires manual examination by ophthalmologists, which is time-consuming and expertise-dependent. Therefore, automated detection systems based on deep learning require well-annotated datasets like IDRiD.



```
--------- TEST METRICS ---------
...
Class 1
Dice       : 0.2395
IoU        : 0.1393
Precision  : 0.3287
Recall     : 0.1981

Class 2
Dice       : 0.4448
IoU        : 0.2976
Precision  : 0.5167
Recall     : 0.4431

Class 3
Dice       : 0.5800
IoU        : 0.4197
Precision  : 0.4995
Recall     : 0.7426

Class 4
Dice       : 0.2499
IoU        : 0.1845
Precision  : 0.3258
Recall     : 0.2183

Class 5
Dice       : 0.9298
IoU        : 0.8735
Precision  : 0.8938
Recall     : 0.9775

-------------------------------
Mean Dice       : 0.4888
Mean IoU        : 0.3829
Mean Precision  : 0.5129
Mean Recall     : 0.5159
```

Figure 4 Test Metrics Results

The figure 4 average performance on the IDRiD dataset shows achievment of a mean Dice score of 0.4888 and a mean IoU of 0.3829. The proposed multi-class U-Net has been able to show a moderate performance in segmentation. The model performed extremely well on the Optic disc segmentation (Dice = 0.9298). Hard exudates segmentation had a moderate performance. Microaneurysms and soft exudates had a low performance due to having a smaller size and sparsely populated classes. The model showed syable and consistent training convergence, but further improvements are needed to enhance segmentation of smaller lesions.

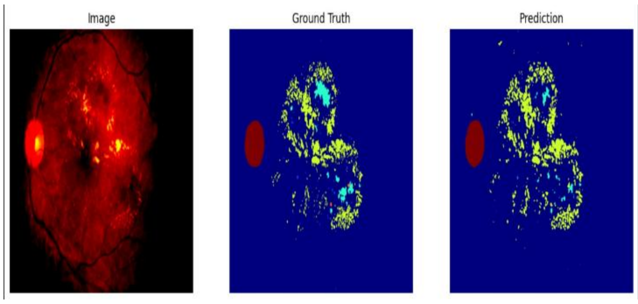


Figure 5 Images, Ground Truth, And Prediction

This figure 5 shows the original retinal image alongside the ground truth mask and the predicted mask by the model. The prediction shows similarities to the ground truth mask, specifically for the Optic Disc region. Minor differences can be seen in some of the smaller lesion areas which accounts for the lower scores of some of the classes.

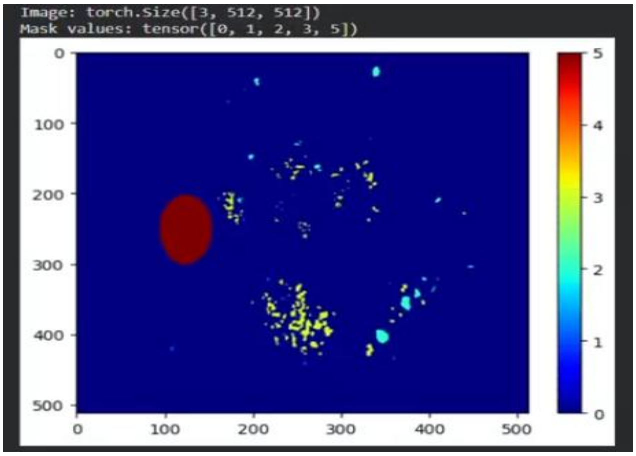


Figure 6 Mask Visualization with Class Values

This figure 6 shows the segmentation mask with pixel-wise class labels ranging from 0 to 5. Each color represents a different category, including background, various lesion types, and the Optic Disc. The color bar on the side helps identify the numerical class associated with each region. The Optic Disc appears as a large, clearly defined region, while the lesion areas are smaller and scattered across the retina. This visualization confirms that the dataset supports multi-class segmentation. It also helps in understanding how different pathological features are distributed spatially within the retinal image. Such visual representation is useful for verifying annotation quality and model predictions.

Conlusion

This paper describes a deep learning approach for multi-class segmentation of diabetic retinopathy lesions using a combination of convolutional neural networks (CNN) with U-Net architecture. This method provides an effective solution for pixel-wise classification of retinal fundus images and segmentation of important lesion types, such as microaneurysms, hemorrhages, hard exudates, soft exudates, and the optic disc. Results of the experiments show that the model performs well for large lesions and large structures (i.e., optic disc) but has low performance for small lesions. Most of the small lesions go undetected due to an imbalance in the classes of the lesions. Training with the hybrid loss function that consists of weighted cross-entropy and Dice loss, provides accuracy and consistency in segmentation and contributes to an improved performance of the model. The accuracy of model segmentation is confirmed by the various metrics, which include Dice score, precision, recall, and intersection over union (IoU) score. The performance of segmentation is shown to be worse microaneurysms and soft exudates, demonstrating the need for increased focus to improve performance in detection of small lesions and exploring various strategies to improve performance in segmentation. This model is sustainable and reliable for diabetic retinopathy analysis and can be improved further with modifications for small lesion segmentation, which
can include, but are not limited to, use of attention mechanism, multi-scale feature analysis, and advanced loss function.

References

[1]. Fang, Chunlai, Ning Ma, and Limin Qian. "Multi-Task Transformer Framework and Radiomic Signatures for Multi-Lesion Segmentation, Detection, and Grading in Diabetic Retinopathy." Photodiagnosis and Photodynamic Therapy (2026): 105455.

[2]. Videira, Maria, et al. "Lesion Segmentation Associated with Diabetic Retinopathy Using Deep Learning Methods." Procedia Computer Science 278 (2026): 1283-1290.

[3]. Kumar, Himanshu, and Abhay Saxena. "Ghost-CAS-UNet for On-Device Diabetic Retinopathy Screening: Lightweight Retinal Lesion Segmentation on Edge Hardware."

[4]. Qi, Jie, Pei Wang, and Wei Chen. "TEDR-YNet: A Dual-Encoder Retinal Multi-Lesion Segmentation Network." IEEE Access 14 (2026): 5900-5912.

[5]. Gill, Hunter Mathias, et al. "BigEye: a clinically interpretable deep learning framework for diabetic retinopathy detection and stage prediction." Scientific Reports (2026).

[6]. Agrawal, Sachin Kumar. "Attention-Enhanced U-Net for Automated Lesion Segmentation in Diabetic Retinopathy: A Multi-Scale Deep Learning Approach for Fundus Image Analysis." INTERNATIONAL JOURNAL OF ENGINEERING DEVELOPMENT AND RESEARCH 14.1 (2026): 644-649.

[7]. Akram, Mohsin, et al. "Uncertainty-aware diabetic retinopathy detection using deep learning enhanced by Bayesian approaches." Scientific Reports 15.1 (2025): 1342.

[8]. Abbasi, Rashid, et al. "Diabetic retinopathy detection using adaptive deep convolutional neural networks on fundus images." Scientific Reports 15.1 (2025): 24647.

[9]. Guefrachi, Sarra, Amira Echtioui, and Habib Hamam. "Diabetic retinopathy detection using deep learning multistage training method." Arabian Journal for Science and Engineering 50.2 (2025): 1079-1096.

[10]. Alanazi, Saad, and Rayan Alanazi. "Enhancing diabetic retinopathy detection through federated convolutional neural networks: Exploring different stages of progression." Alexandria Engineering Journal 120 (2025): 215-228..