VOL. 20, NO. 15, AUGUST 2025 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

ENHANCED DETECTION AND SEGMENTATION OF RETINAL HARD EXUDATES IN DIABETIC RETINOPATHY USING AN FPN WITH CNN EFFICIENTNET-B0 ENCODER

D. Praneeth, E. Aravind, Selvamuthukumaran N., Mohan Dholvan, Shetty Sravanthi, and Battula Balnarsaiah
Department of Computer Science and Engineering, Chaitanya (Deemed to be University), Telangana, India
E-Mail: praneeth040@gmail.com

ABSTRACT

This paper explores the application of a Feature Pyramid Network (FPN) with an EfficientNet-B0 Encoder for detecting and segmenting exudates in retinal images, a critical task in Diabetic Retinopathy assessment. Accurate exudate evaluation is vital as it provides insights into retinopathy progression. Our two-stage approach first detects and removes the optic disc to reduce false positives, followed by exudate detection and segmentation. Leveraging EfficientNet-B0 (LENB0) within the FPN architecture enables multi-scale feature representation, which is crucial for handling exudates of varying sizes and shapes. Evaluated on standard datasets, the proposed model achieves superior pixel accuracy and mIoU performance, demonstrating its potential as a reliable clinical tool.

Keywords: diabetic retinopathy detection, hard exudates detection, FPN, CNN, efficientNet-B0, optic disc detection, and segmentation.

Manuscript Received 3 May 2025; Revised 11 October 2025; Published 15 November 2025

1. INTRODUCTION

The original RGB retinal images are preprocessed by separating them into 3 distinct channels: the R-assignment, the G-graph, and the B-graph. Each of these channels is then amplified threefold. Subsequently, these amplified images are input into a Hybrid-Centric CNN classifier, categorizing them into Grade 0 and Grade 1. Following classification, a specialized method for segmenting exudate lesions is employed to further detect the presence of exudates in Diabetic Retinopathy (DR) images. A comparative analysis of various Deep Convolutional Neural Network (DCNN) architectures aimed at reducing the parameters of a residual Convolutional Neural Network (CNN) with residual skip connections. DCNN was utilized for the semantic segmentation of exudates in retinal images [1]. To enhance the performance of the VGG-19 network architecture, an appropriate image augmentation technique was applied. Additionally, a modernized capsule network was introduced in research work (8) for the detection and classification of DR [2]. This network uses convolution and primary capsule layers to extract features from fundus images, while the class capsule layer and softmax function determine the likelihood of the image belonging to a specific class. The evaluation of this approach was conducted using four performance metrics derived from the Messidor dataset.

A novel DR screening method (9) was proposed in the literature, leveraging DL asymmetry features and asymmetric DL feature extraction. In another study by C. Wan Fu, *et al* [3], a new Neural Network (NN) called Residual Multi-scale feature fusion Channel Attention RMCAU-net was introduced for the segmentation of hard exudates in fundus images [4]. This network combines a U-shaped architecture with a residual structure to capture subtle details of hard exudates, demonstrating superior performance compared to the traditional U-net. Despite these advancements, existing methodologies for DR image analysis face several limitations. For instance, while the EAD-Net has shown success in lesion segmentation, its generalizability across diverse datasets and robustness against varying image quality levels remain unexplored [5]. Aruna Priya Latha *et al* proposed a comprehensive DR classification and segmentation approach, which met experimental requirements but lacked scalability on larger datasets and adaptability to other retinal diseases [6].

Explore focused on minimizing parameters for exudate segmentation did not address the impact of different augmentation techniques on various network architectures [7], [8], [9]. Furthermore, the capsule network for DR detection and DL asymmetry features has not been validated on diverse datasets or complex image transformations. Although the RMCA U-net was effective for segmenting hard exudates [10],[11], its applicability to other lesion types and performance under varying image quality conditions were not investigated. Additionally, the computational efficiency and practical utility of these advanced methods for real-time clinical use, particularly in resource-limited settings, require further validation. To address these limitations, the following steps are recommended: First, a detailed comparative analysis of existing methods should be conducted through implementation. It will provide a comprehensive understanding of the drawbacks inherent in these models. This process should be applied to both DR classification and segmentation methods. Once the limitations are identified, tailored DL models should be developed to overcome these challenges. By addressing these weaknesses, the proposed model is expected to achieve superior results compared to existing approaches.

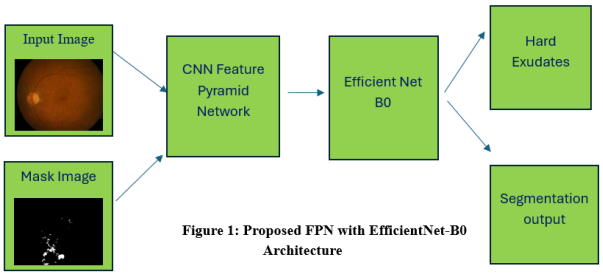


Figure 1: Proposed FPN with EfficientNet-B0 Architecture

2. METHODOLOGY

This study introduces a novel method for detecting and segmenting exudates in retinal images using an FPN with EfficientNet-B0 (ENB0) as the encoder. The proposed approach comprises two main components: (i) locating and removing the Optic Disc (OD) from retinal images, and (ii) accurately identifying and segmenting exudates. The OD, a bright, disk-shaped region in the retina where the optic nerve enters the eye, often appears prominently in retinal images. Due to its high intensity and contrast, it can be mistaken for exudates, leading to false positives during analysis. To address this, the method employs an FPN architecture with an ENB0 encoder to model the spatial hierarchies of retinal images [9]. The FPN captures contextual information across different regions, which is crucial for distinguishing the OD from other retinal structures. Once the OD is visually identified, a morphological transformation technique is applied to isolate and remove it from the image. This preprocessing step enhances precision, ensuring that the exudates remain the refined target area for analysis. A visual representation of the FPN with the ENB0 encoder is illustrated in Figure-1 [12].

After the removal of the optical disc, the second step involves detecting and segmenting exudate in the image [13], [14]. The FPN with an EfficientNet-B0 encoder is highly useful, FPN processes multi-scale feature representation that is useful for detecting. Following the removal of the OD, the second step focuses on detecting and segmenting exudates in the image. Here, the FPN with the ENB0 encoder proves highly effective. The FPN processes multi-scale feature representations, enabling the detection of exudates of varying sizes and shapes, as commonly observed in retinal images. The filtered images are processed through multiple network layers, each corresponding to a different scale of feature detection. This multi-scale approach allows the network to detect prominent and subtle exudates across the retinal surface. During segmentation, the method employs pixel-wise classification, which assigns a probability to each pixel, indicating its likelihood of belonging to an exudate. A final softmax layer is used to interpret these probabilities, generating a precise segmentation map. The combination of ENB0 and FPN enables the network to learn complex features and enhances computational efficiency, a critical factor for clinical applications requiring accurate and rapid results.

2.1 FPN Architecture

FPN is a widely used architecture in computer vision, particularly well-suited for medical imaging tasks such as object detection and semantic segmentation. In medical imaging, FPN is an optimal choice for exudate detection and segmentation in diabetic retinopathy because it can capture multi-scale spatial hierarchies [13]. This capability is crucial for identifying entities of varying sizes, as seen with exudates. FPN enhances convolutional network performance by constructing a feature pyramid at multiple resolution levels. Specifically, it fuses low-resolution, semantically rich features with high-resolution, semantically weaker features through a top-down pathway and lateral connections, enabling improved detection and segmentation accuracy.

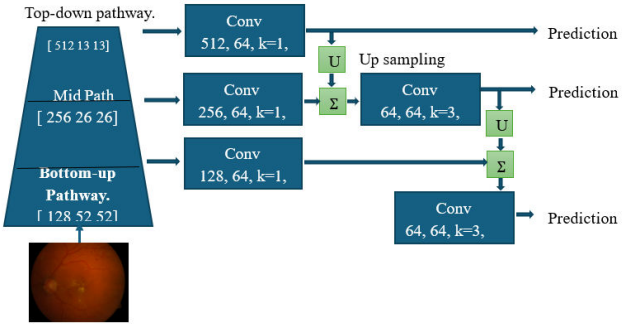


Figure 2: Feature Pyramid Network Architecture

Figure-2. Illustrates the architecture of the feature FPN.

The key components of the FPN architecture are described below:

Input image: The process begins with an input image, which serves as the raw data for feature extraction. This image is fed into the network to initiate the analysis.

Bottom-up pathway: This pathway represents the initial phase of feature extraction. The network progressively captures more complex features as the image data passes through the convolutional layers (depicted as blue feature maps). During this process, the semantic richness of the features increases while the spatial resolution decreases due to pooling operations.

Feature maps: The network generates feature maps at various layers, capturing different levels of information. Early feature maps retain higher resolution but carry less semantic meaning, while deeper layers extract more meaningful features with reduced resolution.

Top-down pathway: After the bottom-up pathway processes the image through its deepest layer, up sampling begins to refine the highest-level feature map. This step ensures that the feature maps maintain high semantic relevance and improved resolution.

Lateral connections with 1×1 convolution: To effectively combine low-level features from the bottom-up pathway with the up sampled high-level features, lateral

connections are introduced. These connections utilize 1×1 convolutions to ensure compatibility in channel dimensions.

Predictions: The network generates predictions at each level of the top-down pathway. The resulting feature maps incorporate both semantic and spatial information, enabling the detection of objects at multiple scales. This multiscale classification capability allows FPN to effectively identify objects of varying sizes.

FPN proves highly effective in detecting and segmenting small, irregularly shaped exudates commonly found in DR. Its multi-scale architecture enhances the model's robustness against pathological variations in retinal images, accounting for differences in size and appearance. After detection and segmentation, the results can be further refined using post-processing techniques like thresholding and morphological procedures before being evaluated.

2.2 EfficientNet-B0

The ENB0 architecture is the foundational model in the EfficientNet family, designed to provide an optimized and scalable network structure for image classification. As the baseline model, ENB0 serves as the starting point for the series, with subsequent variants (B1 to B7) representing scaled-up versions. These variations provide effective performance across a range of processing capacities by proportionally adjusting the network's depth, breadth, and resolution using a set of scaling coefficients.

Input: The network takes an input image with a size of 224 × 224 pixels and three colour channels (RGB).

Conv 3×3 Layer: The first layer is a 3×3 convolutional layer, responsible for initial feature extraction from the input image. This layer captures basic patterns such as edges and textures.

Mobile inverted bottleneck convolution (MBConv): A key component of EfficientNet-B0 is the MBConv block, an advanced version of the inverted residual blocks used in MobileNetV2. This block consists of a 1×1 convolution to expand input parameters, a 3×3 depthwise convolution for spatial feature extraction, and another 1×1 convolution to reduce the number of activations. Additionally, it integrates squeeze-and-excitation optimization to enhance feature representation.

Squeeze-and-excitation blocks: These blocks within the MBConv structure help the network focus on the most essential features. Adaptively recalibrating channel-wise feature responses improves the model's ability to capture meaningful patterns.

Swish activation function: ENB0 employs the Swish activation function, defined as f(x)=x×sigmoid(x)f(x) = x \times \text{sigmoid}(x)f(x)=x×sigmoid(x). Swish is a smooth, Non-monotonic function that has been shown to outperform traditional ReLU activation in DL models.

Downsampling: As data moves through the MBConv blocks, spatial dimensions (height and width) are progressively reduced while the number of channels (depth) increases. This allows the network to process high-level features at lower resolutions, reducing computational costs by avoiding unnecessary low-level feature extraction.

Output: The latter layers usually consist of a dense layer with a softmax activation function for classification and a global average pooling layer. Based on the number of target categories, the result gives the expected class probabilities.

3. RESULTS AND DISCUSSIONS

This section presents the simulation results for exudate detection and segmentation using the FPN with an ENB0 encoder. The methodology focused on two key stages: detecting and removing the optic disc, followed by detecting and segmenting exudates. The proposed system's effectiveness was validated through comparisons with other deep learning models. Five of the 89 color fundus images in the sample were verified by all reviewing experts to be normal fundus images, empty of DR, whereas the remaining 84 showed at least mild Non-Proliferative DR (NPDR). Despite differences in imaging characteristics, all images were taken with a digital fundus camera that had a 50-degree field of view. Overall, the dataset provides an optimal yet non-standard practical scenario for calibration, where image consistency allows for accurate screening of primary diagnostic methods. As such, this dataset can be categorized as calibration level 1 fundus images. Sample images from the dataset are displayed in Figure-3.

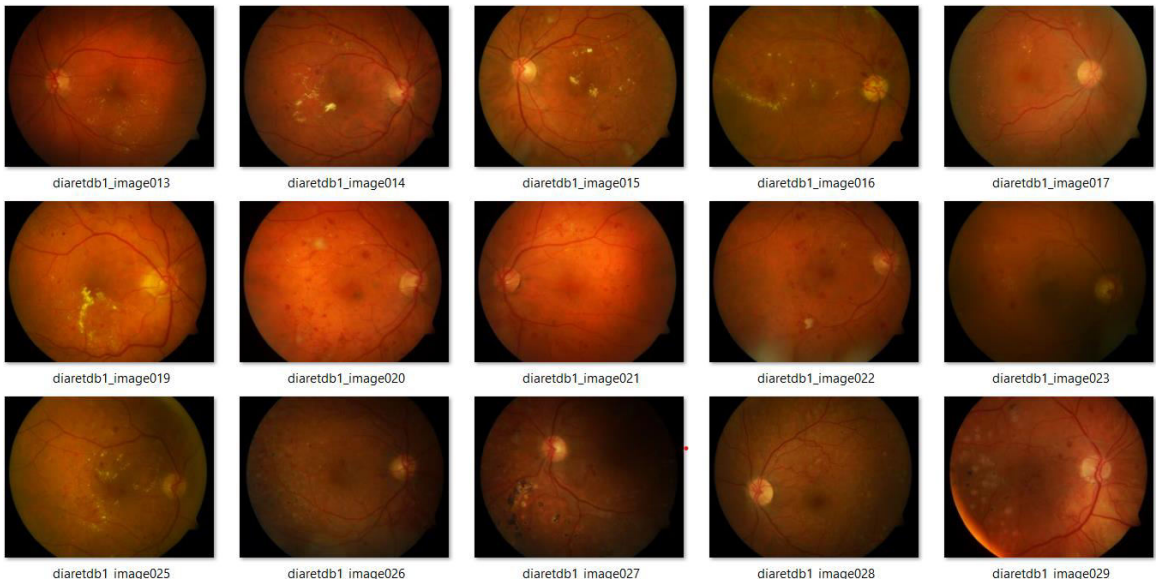


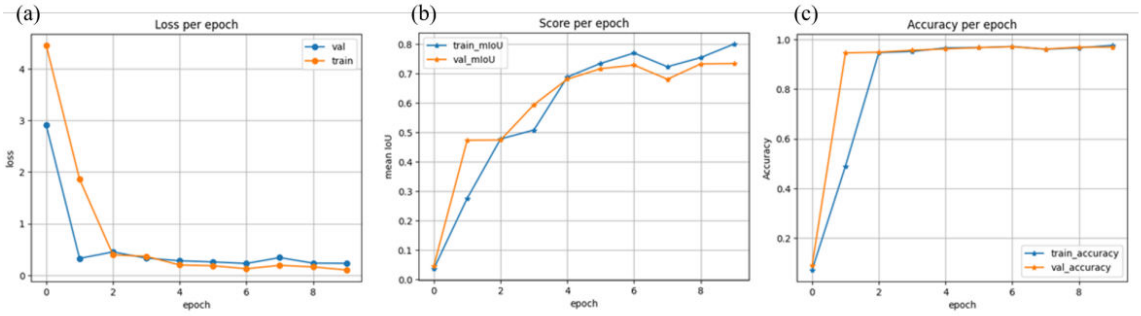
Figure-3. Sample of the database DIARETDB1.

The FPN serves as the backbone of this model, enabling multi-scale feature extraction for detailed retinal image analysis. The ENB0 encoder acts as the initial processing filter, utilizing compound scaling to analyze images at various resolutions. This ensures efficient feature extraction, distinguishing critical OD features from the surrounding retina despite variations in size, shape, and luminance. The multi-level feature integration typical of FPN further refines feature representation, capturing both global and local context for accurate segmentation of the OD.

3.1 Training and Validation Performance

Figure-4(a) presents the training and validation loss of the model over multiple epochs. Initially, both losses are high, with the training loss significantly exceeding the validation loss, indicating an intense learning phase or potential overfitting. However, as training progresses, both losses rapidly decline and stabilize, demonstrating the model's ability to generalize without overfitting. By the eighth epoch, the losses converge, suggesting the model has reached its learning capacity with the given data and configuration.

Figure-5(b) illustrates the mean Intersection over Union (mIoU) score for the training and validation sets across multiple epochs. mIoU is a key metric for assessing segmentation quality. Initially, both scores were low, reflecting the early learning phase. With each epoch, mIoU improves significantly, indicating enhanced segmentation performance. By the sixth epoch, the training mIoU stabilizes, while the validation mIoU exhibits minor fluctuations but follows a similar upward trend. Around the eighth epoch, the two scores converge, confirming a balance between training and generalization.



(a) (b) (c)

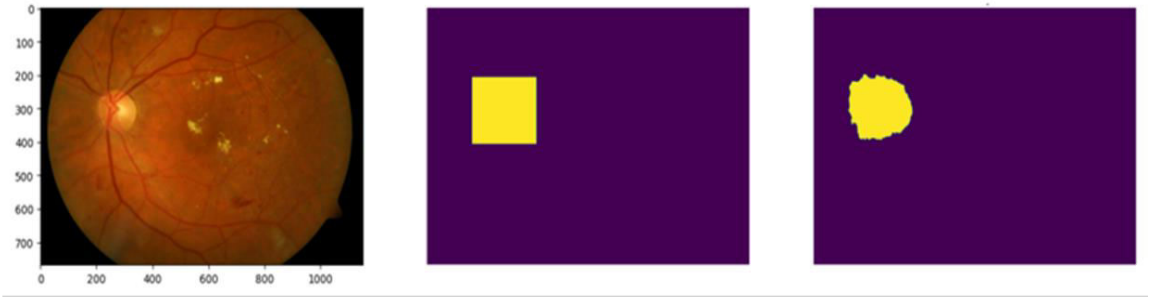
Figure-4. (a-c) Performance metric graphs of the proposed model.

Figure-4(c) displays the accuracy trends for both training and validation sets. Accuracy improves rapidly in the early epochs, with training accuracy rising more steeply than validation accuracy. The curves run parallel from epoch 4 onward, indicating neither overfitting nor underfitting. By the later epochs, accuracy plateaus at a

high level, signifying optimal performance given the model's architecture and hyperparameters.

3.2 Optic Disc Segmentation and Removal

Figure-5 presents the model's segmentation results for detecting and removing the optic disc. The first image is a retinal fundus photograph showing the retina with lesions. The second image, labeled "ground truth," highlights the optic disc region in yellow. The third image represents the FPN-based segmentation output using the EfficientNet-B0 backbone. Following successful optic disc segmentation, the model proceeds to the removal stage. This step is crucial as the optic disc's high intensity can introduce artifacts that mimic pathologies or obscure disease markers. Removal is performed using inpainting techniques that intelligently fill the region once occupied by the optic disc. These algorithms leverage contextual and textural information from adjacent retinal areas to reconstruct a realistic and reliable background. EfficientNet-B0's feature-rich encoding aids in differentiating the OD from the surrounding retina, enabling effective inpainting.



(a) Original Input Image (b) Ground truth Image (c) Segmented Output Image

Figure-5. Segmentation performance.

3.3 Exudate Detection and Segmentation

Once the optic disc is removed, the model focuses on exudate detection and segmentation. Images are preprocessed by resizing to a uniform scale, normalizing pixel values, and enhancing contrast to make exudates more distinguishable. The FPN architecture, with its ENB0 encoder, systematically processes input images to extract and downscale features, generating a rich, multi-scale feature map. ENB0's balanced approach to network width, depth, and resolution ensures both efficiency and accuracy in feature extraction. FPN establishes a top-down structure through lateral connections, combining high-level semantic information from deep layers with fine-grained details from shallow layers. This results in a feature pyramid that represents multiple scales, making it well-suited for detecting exudates of varying sizes and shapes. Using the feature maps produced by the FPN, the network performs pixel-wise classification to determine whether each pixel belongs to an exudate or the background. The classified pixels are then grouped into lesions, defining their boundaries.

Post-processing techniques further refine segmentation results, including:

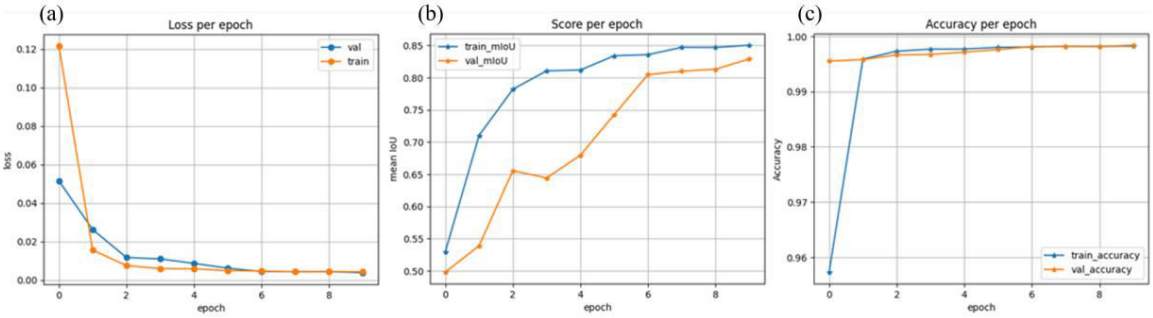
- Thresholding to remove small, isolated predictions
- Morphological operations to smooth lesion borders
- Connectivity analysis to eliminate non-coherent regions

3.4 Training and Evaluation

The FPN-ENB0 model is trained on a dataset of annotated retinal images with known exudate ground truth. Training is optimized using backpropagation, where the network's weights are adjusted to minimize loss. A separate validation set assesses the model's generalization capability. The training performance for exudate detection and segmentation is illustrated in Figure-6, demonstrating the model's effectiveness in learning and accurately identifying exudates in retinal images. The training curve of the suggested model is depicted in Figure-6(a), which shows the loss values for the training and validation sets throughout several epochs. At first, the training and validation losses both drop off quickly, showing that the model is picking up knowledge from the data quickly. Around the fourth epoch, the difference between the training and validation losses narrows, and the rate of decline slows down as training goes on.

The model's training performance throughout epochs for both the training and validation sets is shown in Figure-6(b) using the mean Intersection over Union (mIoU) measure. The overlap between the anticipated object area and the ground truth is measured by the mIoU, which is frequently employed in segmentation tasks. The mIoU scores for training and validation begin at a moderate level and significantly improve in the early epochs. The accuracy of the suggested model throughout epochs is displayed in Figure-6(c) for both the training and validation sets. The accuracy of the model starts high and rises quickly before plateauing after a few epochs. This rapid rise in accuracy indicates that the model quickly

learns to classify the training data. The training and validation accuracies closely track each other, with the validation accuracy slightly lower but still closely mirroring the training accuracy. This consistency suggests that the model generalizes well to unseen data and is not overfitting. The high accuracy of the validation set indicates that the model is stable and well-suited for deployment.



(a) (b) (c)

Figure-6. Performance metric graphs of the proposed segmented model.

Figure-7 displays three images: the first is a color photograph of the retina showing the fundus details, which may include exudates; the second image, labeled "Ground Truth," presents annotated lesions marked in bright colors against a dark background and serves as the reference for evaluating the segmentation accuracy; the third image shows the output from an FPN with an ENB0 encoder. The pattern of the "Ground Truth" and the FPN-ENB0 output are closely aligned, demonstrating the model's strong segmentation capabilities.

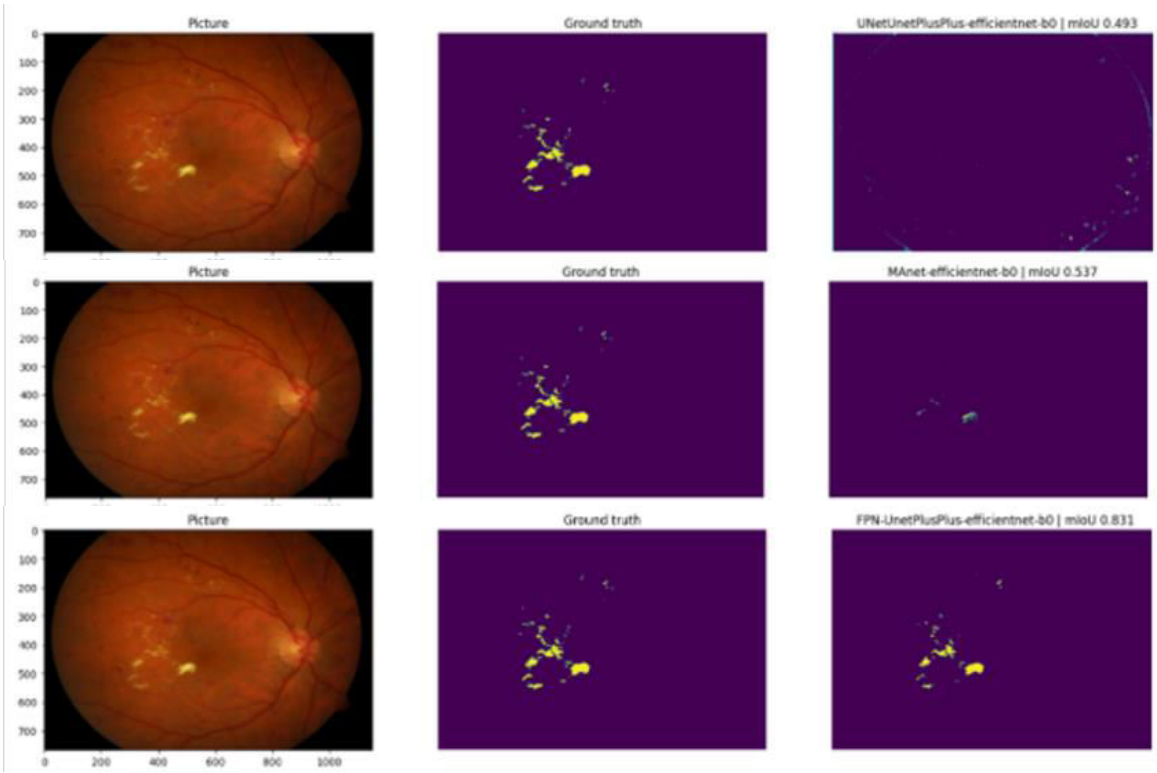


Figure-7. Comparison of segmentation results for other state-of-the-art models.

In Figure-7, the performance of the suggested model is contrasted with that of the most advanced models. The FPN with ENB0 encoder performs better and produces better segmentation results when compared to more conventional deep learning models like U-Net, U-Net++, and MANet. The model's performance on the test set is evaluated using two key metrics: Test Set mIoU and Test Set Pixel Accuracy, as reported in Table-1.

Table-1 compares the performance of various semantic segmentation models using the EfficientNet-B0 backbone and different architectural designs. The models compared include U-Net with EfficientNet-B0, U-Net++, MANet with ENB0, and the proposed FPN with EfficientNet-B0. Test Set Pixel Accuracy measures the percentage of pixels correctly classified by the model. All models exhibit high pixel accuracy, with values above 98%. MANet performs slightly better than the others, but the proposed FPN with ENB0 achieves the highest pixel accuracy of 98.88%, indicating that it correctly classifies the most pixels compared to the other models.

Table-1. Comparison results.

Model Name	Test Set Pixel Accuracy	Test Set mIoU
Unet + EfficientNet-B0	0.9908	0.4943
UnetPlusPlus + EfficientNet-B0	0.9958	0.4831
Manet + EfficientNet-B0	0.9815	0.5774
Proposed FPN + EfficientNet-B0	0.9888	0.8172

Test Set mIoU, which measures the overlap between the predicted and ground truth regions, reveals even greater differences in performance. U-Net and U-Net++ show similar performance with mIoU below 50%, while MANet achieves an mIoU of 53.74%. The proposed FPN with ENB0 shows a significant improvement, achieving a mIoU of 81.72%, which indicates superior performance in terms of both precision and recall. This suggests that the model not only classifies pixels accurately but also effectively segments relevant regions of the image, closely matching the ground truth.

The FPN with the ENB0 encoder outperforms all other state-of-the-art methods due to several advantages. The FPN architecture excels at capturing multi-scale spatial hierarchies, which is crucial for detecting exudates of varying sizes and shapes. The model's ability to distinguish between classes like the optic disc and exudates results in fewer false positives. Additionally, the ENB0 encoder maintains a balance between network depth, width, and input resolution, allowing the model to achieve optimal training efficiency while reducing optimization time. The preprocessing step of optic disc removal further improves exudate detection accuracy by eliminating the high-intensity optic disc, which reduces artifacts that could mimic pathological features. Together, these innovations contribute to the model's superior pixel accuracy and mIoU scores compared to traditional U-Net architectures and more advanced models such as U-Net++ and MANet.

4. CONCLUSIONS AND FUTURE SCOPE

In exploring a novel approach for exudate detection and segmentation of retinal images by combining an FPN with an ENB0encoder. The use of FPN enables multi-scale feature representation, while ENB0 enhances feature extraction efficiency, contributing significantly to the superior performance of the model compared to existing methods. A key innovation of this work is the integration of a preprocessing step for optic disc removal, which improves exudate detection by reducing false positives. The experimental results demonstrate the proposed model's exceptional performance, achieving a pixel accuracy of 98.80% and a mean Intersection over Union (mIoU) of 81.72%, outperforming state-of-the-art models like U-Net, U-Net++, MANet, and ResNet. These impressive metrics highlight the model's effectiveness in both pixel-level classification and accurate segmentation.

This research represents a significant step forward in retinal image analysis and offers great potential for further development. The proposed method can be expanded and applied to other retinal diseases or medical imaging tasks. Future work could involve using this model as a benchmark to train other models with diverse image datasets. Additionally, incorporating this technique into clinical diagnostic systems has the potential to enhance the sensitivity and specificity of DR screening, leading to better patient outcomes and management.

REFERENCES

[1] M. Mateen, J. Wen, N. Nasrullah, S. Sun, and S. Hayat. 2020. Exudate detection for diabetic retinopathy using pre-trained convolutional neural networks. Complexity. 2020: 1.

[2] P. K. Jena et al. 2003. A novel approach for diabetic retinopathy screening using asymmetric deep learning features. Big Data Cogn. Comput. 7(1): 25.

[3] C. Wan et al. 2021. EAD-net: a novel lesion segmentation method in diabetic retinopathy using neural networks. Dis. Markers. 2021: 2021.

[4] G. Latha, P. A. Priya, and V. K. Smitha. 2024. Enhanced diabetic retinopathy detection and exudates segmentation using deep learning: A promising

approach for early disease diagnosis. Multimedia Tools Appl. 2024: 1-24.

[5] R. Romero-Oraá, M. García, J. Oraá-Pérez, M. I. López-Gálvez y R. Hornero. 2020. Effective fundus image decomposition for the detection of red lesions and hard exudates to aid in the diagnosis of diabetic retinopathy. Sensors. 20(22): 6549.

[6] N. Theera-Umpon, I. Poonkasem, S. Auephanwiriyakul, and D. Patikulsila. 2020. Hard exudate detection in retinal fundus images using supervised learning. Neural Comput. Appl. 32(17): 13079-13096.

[7] J. Kaur et al. 2023. Automated detection and segmentation of exudates for the screening of background retinopathy. J. Healthc. Eng. 2023: 2023.

[8] S. K. Ghosh and A. Ghosh. 2021. A novel retinal image segmentation using an R-SVM boosted convolutional neural network for exudates detection. Biomed. Signal Process. Control. 68: 102785.

[9] M. A. Manan et al. 2022. A Residual Encoder-Decoder Network for Segmentation of Retinal Image-Based Exudates in Diabetic Retinopathy Screening. arXiv preprint arXiv:2201.05963.

[10] Y. Fu et al. 2023. RMCA U-net: hard exudates segmentation for retinal fundus images. Expert Syst. Appl. 234: 120987.

[11] Z. Si, D. Fu, Y. Liu, and Z. Huang. 2021. Hard exudate segmentation in retinal image with attention mechanism. IET Image Process. 15(3): 587-597.

[12] G. Sivapriya, V. Praveen, P. Gowri, S. Saranya, S. Sweetha, and K. Shekar. 2022. Segmentation of Hard Exudates for the detection of Diabetic Retinopathy with RNN-based semantic features using fundus images. Mater. Today: Proc. 64: 693-701.

[13] K. R. Remya and M. N. Giriprasad. 2022. An automated exudate detection scheme supporting diabetic retinopathy screening using spatial-spectral-statistical feature maps. Multimedia Tools Appl. 81(7): 9829-9853.

[14] N. J. Mohan, R. Murugan, T. Goel, and P. Roy. 2022. Fast and robust exudate detection in retinal fundus images using extreme learning machine autoencoders and modified Kaze features. J. Digit. Imaging. 35(3): 496-513.