ICJK
Journal of Artificial Intelligence in Bioinformatics
http://dx.doi.org/10.62762/JAIB.2025.322062

RESEARCH ARTICLE

RetinoNet: An Efficient MobileNetV3-Based Model for Diabetic Retinopathy Detection Using Multi-Scale Feature Fusion

Muhammad Usman Saeed[1,2,*], **Aqsa Dastgir**[1], **Muhammad Ahmad Nawaz Ul Ghani**[3] and **Arslan Manzoor**[4]

[1] School of Computer Science and Engineering, Central South University, Changsha 410017, China
[2] School of Computer Science, Harbin Institute of Technology, Shenzhen 518055, China
[3] School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[4] Department of Mathematics and Computer Science, University of Catania, 95131 Catania, Italy

Abstract

Diabetic retinopathy (DR) is a leading cause of blindness globally, requiring timely detection and classification to prevent vision loss. Deep learning techniques offer significant potential for automating DR detection by analyzing retinal fundus images with high precision. This paper proposes a RetinoNet model that consists of MobileNetV3, Convolutional Block Attention Module (CBAM), Atrous Spatial Pyramid Pooling (ASPP), and Feature Pyramid Network (FPN). MobileNetV3 provides a lightweight and efficient foundation for feature extraction, while CBAM emphasizes critical spatial and channel information, enabling the detection of subtle retinal abnormalities. ASPP captures multi-scale contextual information through atrous convolutions, improving the model's ability to identify lesions of varying sizes and shapes. FPN combines hierarchical features from multiple network levels, ensuring both fine-grained details and high-level semantics are leveraged for accurate classification. The model was trained on the APTOS dataset. Evaluation metrics such as accuracy, precision, recall, and F1 score demonstrate the effectiveness of the proposed model in achieving state-of-the-art performance for DR detection and classification across five severity levels. This approach addresses computational challenges and improves generalization, making it suitable for both clinical and remote healthcare applications.

Keywords: diabetic retinopathy, feature fusion, bio-informatics, multi-scale.

1 Introduction

Diabetic retinopathy (DR) is a sight-impaired disease that affects the back of the eye, specifically the sensitive

Submitted: 16 August 2025
Accepted: 03 September 2025
Published: 25 October 2025

Vol. 1, **No.** 2, 2025.
10.62762/JAIB.2025.322062

***Corresponding author:**
✉ Muhammad Usman Saeed
usmansaeed@csu.edu.cn

Citation
Saeed, M. U., Dastgir, A., Ghani, M. A. N. U., & Manzoor, A. (2025). RetinoNet: An Efficient MobileNetV3-Based Model for Diabetic Retinopathy Detection Using Multi-Scale Feature Fusion. *Journal of Artificial Intelligence in Bioinformatics*, 1(2), 58–67.

layer of the eye called the retina, where images are captured [1]. DR is associated with sustained high blood sugar levels and, therefore, is an injury to the eye vessels in the retina, which can lead to exudation of fluids, bleeding within the eye, and, in a later stage, the formation of pathological neovascularization. In this progressive disorder, a deficiency of proper treatment can lead to loss of vision [2]. It is noted that DR occurs at one of the highest levels, causing blindness in the working-age population, and thus requires regular screening and treatment. Advances in medical imaging and artificial intelligence are helping clinicians diagnose and monitor DR more effectively, offering promising tools for preventing vision loss in diabetic patients [3].

The detection of DR using machine learning and deep learning has become an area of significant interest in medical imaging [4, 5], owing to its potential to enhance diagnostic accuracy and enable early intervention [6, 7]. Traditional diagnostic methods, which rely on retinal examination by trained specialists, can be time-consuming, subjective, and inaccessible in regions with limited healthcare resources. Machine learning and deep learning techniques, particularly convolutional neural networks (CNNs), offer powerful tools for analyzing retinal fundus images, identifying subtle signs of DR with high sensitivity and specificity. These methods can automate DR grading, from mild to severe stages, by learning patterns associated with retinal abnormalities, such as microaneurysms, hemorrhages, and neovascularization. Recent advances in deep learning architectures and access to large annotated datasets have further propelled the development of DR detection models, showing promise in both clinical and remote settings to support ophthalmologists and improve patient outcomes.

Deep learning, while highly effective for detecting DR, presents significant drawbacks in terms of computational cost. Deep learning models, especially CNNs, commonly used for analyzing retinal images, require substantial computational power due to the need for extensive data processing and high-dimensional parameter training. These models often depend on large datasets of high-resolution fundus images, demanding significant memory and specialized hardware like GPUs or TPUs to manage the intensive computations. Training deep learning models for DR can be time-consuming as it requires processing and learning from millions of parameters over numerous epochs, which can drive up energy consumption and operational costs. The deployment of these models in real-time diagnostic systems, especially in resource-constrained settings, remains challenging due to their high computational requirements. To address these issues, researchers are exploring optimized architectures, model compression techniques, and transfer learning to reduce the computational burden without sacrificing the accuracy needed for reliable DR detection.

The proposed methodology for detecting diabetic retinopathy is built on the efficient MobileNetV3 architecture, which is enhanced with CBAM, ASPP, and FPN modules to improve the extraction and classification accuracy of characteristics. MobileNetV3 serves as the foundation, leveraging its lightweight design and depth-wise separable convolutions for efficient computations. CBAM is incorporated into select bottleneck layers to prioritize significant spatial and channel features, allowing the model to focus on subtle retinal anomalies. ASPP is utilized in deeper layers to capture multiscale contextual information via atrous convolutions with varying dilation rates, aiding in the detection of lesions of diverse sizes and shapes. The FPN module strengthens the network by merging hierarchical features from different levels, enabling the model to combine detailed local features with broader semantic information. This approach ensures a precise and effective classification of stages of diabetic retinopathy, addressing variations in the patterns and severity of the lesions. Here are the main technical contributions of the RetinoNet model.

- **Integration of CBAM:** Enhances spatial and channel attention within bottleneck layers to focus on critical retinal features indicative of diabetic retinopathy.
- **Incorporation of ASPP**: Introduces multi-scale context awareness through atrous convolutions with varying dilation rates, enabling the detection of lesions of different sizes and shapes.
- **Utilization of FPN:** Combines multi-level features from the network, ensuring both fine-grained details and high-level semantics are utilized for robust classification.
- **Lightweight Base Architecture:** Leverages MobileNetV3s efficient design to maintain low computational requirements while achieving high feature extraction performance.
- **Optimized for Retinal Images:** Tailors the network to effectively handle the unique

challenges of diabetic retinopathy, such as diverse lesion patterns and severity stages, ensuring improved detection accuracy and stage classification.

2 Related Work

Guefrachi et al. [8] discuss a deep learning approach for detecting and classifying diabetic retinopathy (DR) using convolutional neural networks (CNNs) with a multistage training method. It evaluates various CNN architectures, including InceptionResnetV2, VGG16, VGG19, DenseNet121, MobileNetV2, and EfficientNet2, on a dataset of retinal fundus images. The study employs data augmentation techniques to enhance model resilience and reduce overfitting, achieving high classification accuracy of 96.61% for DR stages. The research highlights the importance of model evaluation on external datasets to ensure robustness and generalizability, emphasizing the potential of deep learning in improving early diagnosis and treatment of diabetic retinopathy. Key metrics such as recall, precision, and F1 score were analyzed, indicating significant potential for clinical applications. Kurup et al. [9] discuss the development of an automated system for detecting and classifying diabetic retinopathy (DR) using a pretrained Inception-v3 deep learning model. The study utilizes the APTOS 2019 blindness detection dataset, which contains retinal images classified into five stages of DR. The model achieved approximately 82% accuracy and a Cohens weighted Kappa score of 0.72. The objectives outlined include data pre-processing, model selection, and the creation of a user-friendly interface for image uploads and results display. The literature review highlights various methodologies and advancements in DR detection using deep learning techniques.

Bodapati et al. [10] present a model for predicting the severity levels of diabetic retinopathy (DR) using deep convolution feature aggregation from the pre-trained VGG-16 model. By extracting features from multiple convolution blocks, the authors enhance the representation of retinal images. The model, evaluated on the Kaggle APTOS 2019 dataset, achieved an accuracy of 84.31%. The study emphasizes the superiority of deep features over handcrafted features and the effectiveness of feature aggregation for DR classification. Mohanty et al. [11] discuss advancements in diabetic retinopathy (DR) detection using deep learning techniques, particularly focusing on two models: a hybrid network combining VGG16 and XGBoost Classifier and the DenseNet 121 network. Utilizing the APTOS 2019 Blindness Detection dataset, the study addresses class imbalance and reports that the DenseNet 121 model achieved a high accuracy of 97.30%, significantly outperforming the hybrid model, which achieved 79.50%. The findings indicate that deep learning can enhance the efficiency and accuracy of DR diagnosis. Additionally, the paper mentions the E-DenseNet model, which combines Eyenet and DenseNet architectures, achieving an average accuracy of 91.2% across four datasets. The study emphasizes the effectiveness of deep learning models over traditional methods and highlights future work aimed at developing applications for early DR detection to assist healthcare professionals and patients.

Nahiduzzaman et al. [12] present a novel automated technique for detecting diabetic retinopathy (DR) using a combination of a lightweight parallel convolutional neural network (CNN) for feature extraction and an extreme learning machine (ELM) for classification. The method enhances fundus images through Contrast Limited Adaptive Histogram Equalization (CLAHE) to highlight lesions. The proposed framework achieved high accuracies of 91.78% and 97.27% on two datasets (Kaggle DR 2015 and APTOS 2019) and demonstrated stability across various dataset sizes. It outperformed existing models in classifier performance, model complexity, and prediction time, making it suitable for real-time medical applications. The study emphasizes the efficiency of the ELM in medical image analysis, particularly for multiclass classifications, and highlights the importance of recall in accurately identifying affected patients. Sacchini et al. [13] present a novel hybrid convolutional neural network (CNN) model for the automatic classification of diabetic retinopathy (DR) from fundus images. It combines two deep learning architectures, ResNet50 and InceptionV3, for feature extraction, achieving high performance metrics: accuracy of 96.85%, sensitivity of 99.28%, specificity of 98.92%, precision of 96.46%, and F1 score of 98.65%. The study emphasizes the importance of data quality and preprocessing techniques, utilizing a dataset of 44,119 high-resolution retinal images categorized into five classes of DR. The model was validated through 5-fold cross-validation, demonstrating consistent performance. Additionally, the research highlights the effectiveness of automated techniques in diagnosing DR, with results from both Japanese and American datasets showing promising sensitivity and specificity

Table 1. Standard architecture of the MobileNetV3.

Input	Operator	exp size	Out	SE	NL	Stride
224×224×3	Conv2D	-	16	-	HS	2
112×112×16	Bottleneck	16	16	-	RE	1
112×112×16	Bottleneck	64	24	-	RE	2
56×56×24	Bottleneck	72	24	-	RE	1
56×56×24	Bottleneck	72	40	Yes	RE	2
28×28×40	Bottleneck	120	40	Yes	RE	1
28×28×40	Bottleneck	120	40	Yes	RE	1
28×28×40	Bottleneck	240	80	-	HS	2
14×14×80	Bottleneck	200	80	-	HS	1
14×14×80	Bottleneck	184	80	-	HS	1
14×14×80	Bottleneck	184	80	-	HS	1
14×14×80	Bottleneck	480	112	Yes	HS	1
14×14×112	Bottleneck	672	112	Yes	HS	1
14×14×112	Bottleneck	672	160	Yes	HS	2
7×7×160	Bottleneck	960	160	Yes	HS	1
7×7×160	Bottleneck	960	160	Yes	HS	1
7×7×160	Conv2D 1x1	-	960	-	HS	1
7×7×960	AvgPool	-	960	-	-	-
1×1×960	FC	-	1280	-	HS	-
1×1×1280	FC	-	1000	-	-	-

rates.

3 Methodology

Using AI in healthcare brings transformative benefits, enhancing both patient care and operational efficiency [1]. AI-driven tools can analyze vast amounts of medical data swiftly, leading to faster and more accurate diagnoses for conditions such as brain tumors [14], cardiovascular diseases, and spine fractures [15]. Artificial intelligence has shown transformative potential in ophthalmology, particularly in the detection and classification of diabetic retinopathy (DR) [13, 23]. Automated DR screening systems powered by deep learning can analyze retinal fundus images with high sensitivity, identifying subtle lesions such as microaneurysms, hemorrhages, and neovascularization that may be missed during routine examination. By providing rapid and accurate grading across different severity levels, AI assists ophthalmologists in early diagnosis, reducing the risk of vision loss in diabetic patients. Furthermore, lightweight and efficient models, such as those based on MobileNet architectures, make it feasible to deploy DR detection tools in mobile devices and telemedicine platforms. This expands access to screening in remote or resource-limited settings, ensuring timely interventions and improving patient outcomes on a large scale.

3.1 MobileNetv3

The MobileNetv3 is a deep learning architecture that has been optimized for conducting image classification tasks on mobile and other devices with limited resources. MobileNet V3 is a Google innovation that advances MobileNet V1 [17] and V2 [16] by fusing elements of both MobileNet architectures along with the neural architecture search (NAS) technique to enhance speed and precision. New techniques brought in by MobileNet V3 include the use of squeeze-and-excitation modules that recalibrate the responses of the channels feature-wise on adaptive on a h-swish activation function, which is cheaper computationally than ReLU. These updates translate to improved accuracy in the use of MobileNetV3 as well as efficiency in computation and storage. It comes in two variations: MobileNetV3 Small and MobileNetV3 Large, which are specifically designed to meet the needs of speed and accuracy, and thus, they are ideal for real-time image processing applications such as smartphones and other IoT devices. The architecture of the MobileNetV3 is given in Table 1.

3.2 Convolution Block Attention Module (CBAM)

MobileNetV3, with the introduction of attention mechanisms, mainly Channel Attention Module (CAM) and Spatial Attention Module (SAM) to the Diabetic Retinopathy classification model, has noticeably set the bar. The CAM shifts the focus to certain key feature channels by imposing the schemes

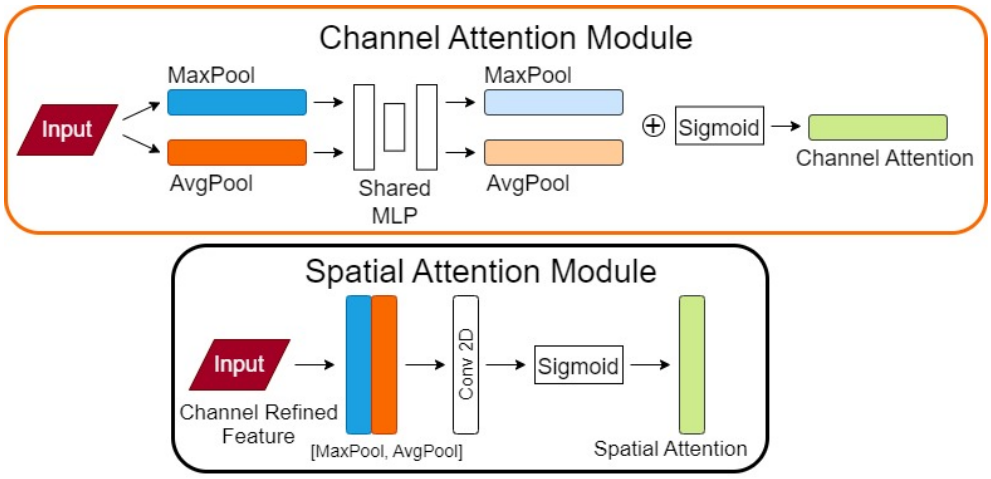

Figure 1. The standard architecture of the CBAM block.

of both global average pooling and max pooling, then throws fully connected layers to learn inter-channel dependencies, which highlight the important ones while subduing the less relevant ones. Both these mechanisms, together with other techniques, make the model better at focusing on important spatial and contextual features of the retina image, which is core to the precise detection and classification of DR. The course of action with the said modules in MobileNetV3 is to better scrutinize the feature extraction process, thus the model can tactfully capture the discrete changes in eye conditions and the severity of those, and in turn, boost the classification performance. The architecture of the CBAM block is shown in Figure 1.

3.3 Atrous Spatial Pyramid Pooling (ASPP)

The Atrous Spatial Pyramid Pooling (ASPP) [18] is a deep learning technique that can extract information at different scales; thus, inputs are transformed into a feature map, which a network can then analyze to predict the probability of classifying a pixel as a certain class more accurately. The so-called contextual information that had an impact on the results of the segmentation is proper to this methodology and is shown in Figure 2. The ASPP module first scales all the feature maps and then produces the multi-scale contextual features from all the feature maps that help in better performance.

3.4 Feature Pyramid Network (FPN)

The Feature Pyramid Network (FPN) is an architecture of a neural network constructed to enhance the detection of objects at various scales using the multiscale representation of features. In other words, it constructs a feature pyramid by combining the deep-layer-rich semantic contents with the early layers

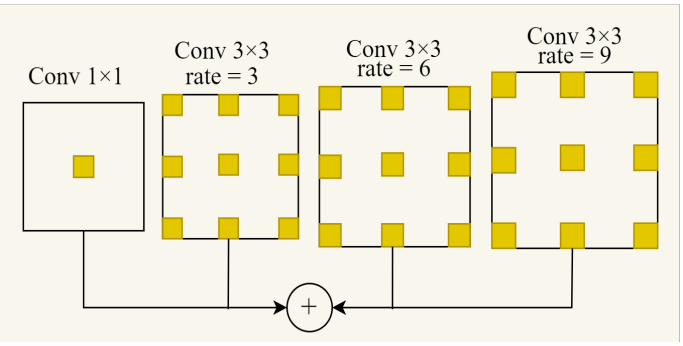

Figure 2. The standard architecture of the ASPP module.

of pronounced details in the network. The FPN extends the scope of the network with its ability to wear features of various aspects of the objects, as it does this chiefly by progressive upsampling and combining these features. Hence, this form is good in applications such as small object detection. The most prevalent use of FPN is for complex detection tasks such as Faster R-CNN and RetinaNet optimizers, which improve their performance over multiple scales with very little extra computational requirement. Figure 3 shows the architecture of the FPN module.

3.5 Proposed RetinoNet for Diabetic Retinopathy Detection

The proposed deep learning model for diabetic retinopathy detection builds upon the lightweight MobileNetV3 architecture and integrates advanced modules such as CBAM (Convolutional Block Attention Module), ASPP (Atrous Spatial Pyramid Pooling), and FPN (Feature Pyramid Network) to enhance its performance. MobileNetV3 provides an efficient foundation for feature extraction with its streamlined depthwise separable convolutions and attention mechanisms, making it suitable for

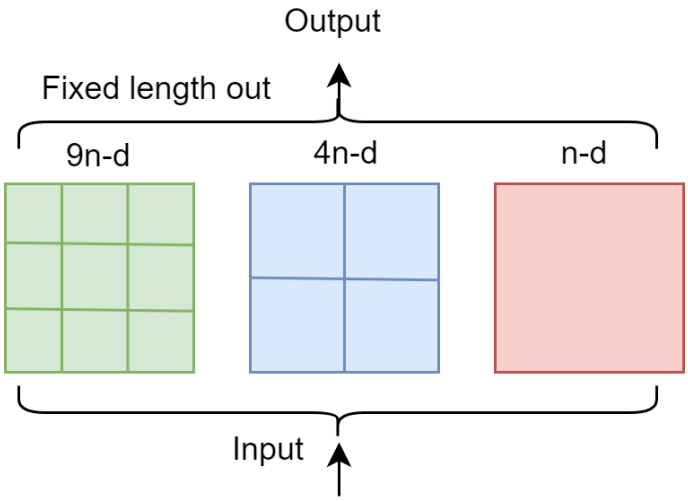

Figure 3. The standard architecture of the FPN module.

resource-constrained environments. The CBAM block is incorporated into specific bottleneck layers to refine feature representation by focusing on the most relevant spatial and channel information, enabling the model to prioritize subtle retinal abnormalities indicative of diabetic retinopathy. The ASPP module, added to deeper bottleneck layers, captures multi-scale context by applying atrous convolutions with varying dilation rates, which is critical for identifying lesions of different sizes and shapes in retina scans. Furthermore, the FPN is integrated to fuse hierarchical features from different layers, ensuring that both fine-grained details and high-level semantic information are utilized for robust classification. The architecture of the proposed model for diabetic retinopathy detection is shown in Figure 4.

4 Experimental Results and Discussion

4.1 Dataset Definition

The dataset comprises retinal images used to detect and classify diabetic retinopathy [19]. Originally sourced from the APTOS 2019 Blindness Detection dataset, the images have been resized to 224×224 pixels to ensure compatibility with various pre-trained deep learning models. The dataset is organized into five classes based on the severity or stage of diabetic retinopathy. The classes are No_DR, for healthy retinas with no signs of diabetic retinopathy, Mild for early-stage indicators, Moderate for noticeable symptoms requiring monitoring, Severe for advanced damage, and Proliferate_DR for cases with significant progression and high risk of vision loss. This structured format facilitates efficient training and evaluation of deep learning models for automated detection and severity classification of diabetic retinopathy. Figure 5 shows the sample images for detecting diabetic retinopathy.

4.2 Evaluation Metrics

4.2.1 Accuracy

Accuracy is defined as the ratio of correct predictions to the total number of predictions. This metric is calculated by dividing the number of correct predictions by the total number of predictions, then multiplying by 100 to express it as a percentage.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

4.2.2 Precision

Precision is a metric used to evaluate a model's ability to correctly identify positive instances from those it predicts as positive. It is calculated by dividing the number of true positive predictions by the sum of true positive and false positive predictions.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

4.2.3 Recall

Recall is a metric that measures a model's ability to identify all relevant positive instances within a dataset. It is calculated by dividing the number of true positive predictions by the sum of true positives and false negatives.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

4.2.4 F1 Score

The F1 score is a metric that combines precision and recall into a single value, providing a balanced assessment of a model, especially when dealing with imbalanced data. It is the harmonic mean of precision and recall.

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

4.3 Model Training Parameters

The dataset used in this research is divided into two subsets, with 80% allocated for training and 20% for testing. The model was trained using a learning rate of 0.001 with a step-based decay schedule to ensure stable convergence. The batch size was set to 32 to balance computational efficiency and

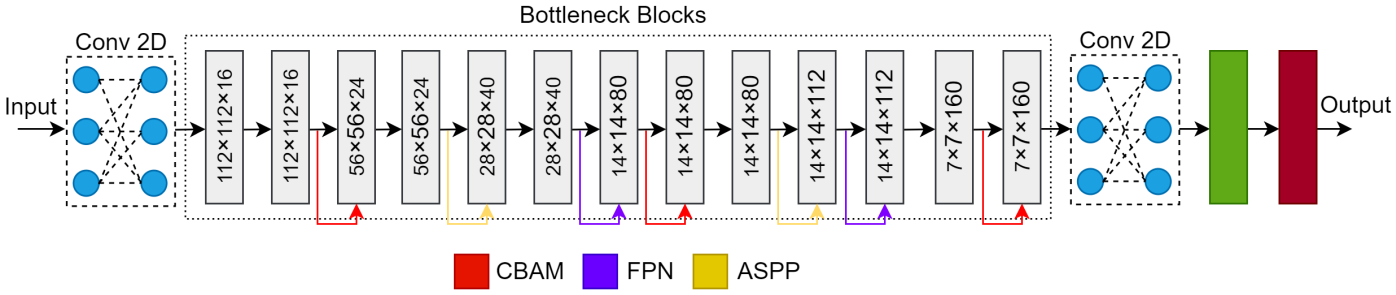

Figure 4. The architecture of the proposed model for diabetic retinopathy detection.

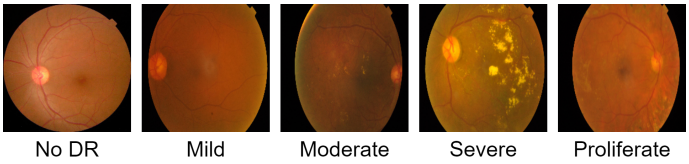

Figure 5. The sample images of the diabetic retinopathy detection dataset.

memory usage, while the number of epochs was chosen as 150 to allow sufficient learning without overfitting. Data augmentation techniques such as rotation, flipping, and brightness adjustment were applied to enhance generalization. The training was performed on an NVIDIA Tesla V100 GPU, leveraging its high computational power and memory capacity to handle the resized 224×224 retina images efficiently. The Adam optimizer was utilized for weight updates due to its adaptability and faster convergence, and categorical cross-entropy was used as the loss function to handle the multi-class classification of diabetic retinopathy stages. Early stopping and checkpointing mechanisms were also employed to prevent overfitting and save the best-performing model for validation and testing.

4.4 Results

The proposed model for the detection of diabetic retinopathy was evaluated in the APTOS 2019 blindness detection dataset at five severity levels (No_DR, Mild, Moderate, and Severe). The evaluation metrics include accuracy, precision, recall, and F1 score. Separate experiments were conducted to measure the impact of preprocessing and data augmentation on model performance.

4.4.1 Results of the RetinoNet Without Preprocessing or Data Augmentation

The model was first trained on the dataset without preprocessing or data augmentation. The results are summarized in Table 2.

Table 2. Results of the RetinoNet on raw image dataset.

Metric	No_DR	Mild	Moderate	Severe
Accuracy	0.85	0.78	0.80	0.83
Precision	0.86	0.75	0.82	0.85
Recall	0.84	0.76	0.81	0.82
F1 Score	0.85	0.75	0.81	0.83
			Total	82

4.4.2 Results of the RetinoNet with Preprocessing

The preprocessing steps, including resizing images to 224×224 pixels and normalizing pixel values, were applied. These steps significantly improved the performance of the model, as shown in Table 3.

Table 3. Results of the RetinoNet with preprocessing.

Metric	No_DR	Mild	Moderate	Severe
Accuracy	0.90	0.85	0.88	0.91
Precision	0.91	0.83	0.89	0.92
Recall	0.89	0.84	0.87	0.90
F1 Score	0.90	0.83	0.88	0.91
			Total	89

4.4.3 Results of the RetinoNet with Preprocessing and Data Augmentation

To further improve the model's generalization, data augmentation techniques such as random rotation, flipping, and brightness adjustment were applied. The results are shown in Table 4. The results

Table 4. Results of the RetinoNet with preprocessing and data augmentation.

Metric	No_DR	Mild	Moderate	Severe
Accuracy	0.94	0.89	0.92	0.94
Precision	0.95	0.88	0.93	0.95
Recall	0.93	0.89	0.91	0.93
F1 Score	0.94	0.88	0.92	0.94
			Total	92

demonstrate that preprocessing significantly improves model performance, particularly regarding recall

Table 5. Ablation study of the proposed RetinoNet model.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
RetinoNet	**92.0**	**95.0**	**93.0**	**94.0**
Without CBAM	89.5	91.8	90.2	91.0
Without ASPP	88.2	90.5	89.0	89.7
Without FPN	87.8	89.9	88.5	89.2

and F1 score. Data augmentation further improves generalization, achieving the highest precision and F1 scores in all stages of diabetic retinopathy. These findings emphasize the importance of preprocessing and data augmentation for robust and accurate diabetic retinopathy detection.

4.5 Ablation Study

To assess the contribution of each module to the proposed model, an ablation study was conducted by systematically removing key components such as CBAM, ASPP, and FPN. The impact of these modules on the accuracy, precision, recall, and F1 score of the model was evaluated on the APTOS 2019 Blindness Detection dataset.

The RetinoNet with all components achieves the highest performance across all metrics. Removing CBAM leads to a decline in precision and recall, as CBAM enhances spatial and channel attention. Eliminating ASPP reduces the model's ability to capture multi-scale features, leading to lower accuracy. Removing FPN results in a drop in F1-score, as the model loses hierarchical feature fusion across multiple network layers.

Table 5 presents the results of an ablation study conducted on the proposed MobileNetV3-based model to evaluate the contribution of key modules—CBAM (Convolutional Block Attention Module), ASPP (Atrous Spatial Pyramid Pooling), and FPN (Feature Pyramid Network). The full model incorporating all these components achieves the highest performance, with 92.0% accuracy, 95.0% precision, 93.0% recall, and 94.0% F1-score. Removing CBAM leads to a decline in all metrics, particularly precision and recall, indicating its role in enhancing spatial and channel attention for feature extraction. Excluding ASPP results in an even lower performance, as the model loses its ability to capture multi-scale contextual information. Similarly, removing FPN reduces hierarchical feature fusion, leading to the lowest accuracy (87.8%) and a significant drop in F1-score (89.2%). These results confirm that each module contributes significantly to the model's effectiveness in diabetic retinopathy classification, with CBAM, ASPP, and FPN collectively optimizing feature representation and classification accuracy.

Table 6. Comparison of the proposed model with state-of-the-art methods.

Reference	Model	Accuracy (%)
[20]	InceptionResnetV2	82.18
[11]	Hybrid	79.5
[10]	VGG16	84.31
[9]	InceptionV3	82
[21]	EfficientNet-B6	86.03
[22]	CNN	85
Ours	RetinoNet	92

4.6 Comparison of the proposed model with other deep learning models

The comparison table highlights the performance of the proposed Improved MobileNetV3 model against several state-of-the-art methods for diabetic retinopathy detection and classification. Each model is evaluated based on its accuracy, demonstrating its effectiveness in handling this medical imaging task. Table 6 shows the results of different deep learning models.

The InceptionResNetV2 achieves an accuracy of 82.18%. This result reflects its capacity for feature extraction through its hybrid architecture combining Inception and ResNet modules. However, while effective, it lags behind models with more optimized architectures for medical imaging. Similarly, Hybrid models, which often combine features from multiple architectures, show an accuracy of 79.5%, demonstrating moderate performance but lacking the specialized capabilities of more advanced networks.

Moving to VGG16, a widely used convolutional neural network, the accuracy improves to 84.31%, indicating its strength in feature representation for this dataset. Meanwhile, InceptionV3, a precursor to InceptionResNetV2, achieves a comparable accuracy of 82%, showcasing the consistency of the Inception family in diabetic retinopathy detection tasks.

The Table 6 also features EfficientNet-B6, a model

known for its efficiency and accuracy, achieving a relatively higher accuracy of 86.03%. This result highlights its ability to balance computational efficiency with performance, making it a competitive choice for medical imaging tasks. A general CNN model delivers an accuracy of 85%, showcasing its utility but also underscoring the need for specialized enhancements to achieve superior results.

The proposed improved MobileNetV3 model outperforms all the other models, achieving an accuracy of 92%. This significant improvement is attributed to the integration of advanced techniques such as the CBAM, ASPP, and FPN. These enhancements enable the model to focus on critical features, capture multi-scale context, and combine hierarchical representations, making it particularly effective for diabetic retinopathy classification. The results demonstrate the ability of the proposed model to address the challenges of varying the patterns and severity levels of the lesion in the retinal images, establishing it as the most effective solution among the methods compared.

5 Conclusion

This study presents an improved MobileNetV3-based deep learning model for the detection and classification of diabetic retinopathy across five severity stages. By integrating advanced modules such as the Convolutional Block Attention Module (CBAM), Atrous Spatial Pyramid Pooling (ASPP), and Feature Pyramid Network (FPN), the proposed model effectively captures critical spatial, channel, and multi-scale contextual information from retinal images. These enhancements, combined with preprocessing and data augmentation, significantly improve the model's accuracy and generalization capabilities, achieving state-of-the-art results on the APTOS 2019 Blindness Detection dataset. The results demonstrate the importance of lightweight architectures for resource-constrained environments, highlighting MobileNetV3's efficient feature extraction and computational scalability. The integration of CBAM ensures attention to subtle retinal abnormalities, while ASPP captures multi-scale lesion features, and FPN fuses hierarchical representations for robust classification. The proposed methodology addresses the computational challenges associated with deep learning in medical imaging, making it suitable for deployment in both clinical and remote healthcare settings.

Future work will focus on testing the model on additional datasets to further validate its robustness and exploring transfer learning techniques to enhance its performance in real-world applications. The proposed approach offers a promising direction for automated diabetic retinopathy detection, contributing to early diagnosis and improved patient outcomes.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

This study uses the anonymized, public APTOS 2019 dataset (CC BY-NC-SA 3.0 license). No human subjects, identifiable data, or interactions were involved. Ethical approval and consent are not required under guidelines for secondary data analyses (e.g., Helsinki Declaration).

References

[1] Cushley, L. N., Csincsik, L., Virgili, G., Curran, K., Silvestri, G., Galway, N., & Peto, T. (2024). The NaviSight study: Investigating how diabetic retinopathy and retinitis pigmentosa affect navigating the built environment. *Disabilities, 4*(3), 507-524. [CrossRef]

[2] Cross, N., van Steen, C., Zegaoui, Y., Satherley, A., & Angelillo, L. (2022). Retinitis pigmentosa: burden of disease and current unmet needs. *Clinical Ophthalmology*, 1993-2010. [CrossRef]

[3] Khalifa, M., & Albadawy, M. (2024). Artificial intelligence for diabetes: Enhancing prevention, diagnosis, and effective management. *Computer methods and programs in biomedicine update, 5*, 100141. [CrossRef]

[4] Rana, M., & Bhushan, M. (2023). Machine learning and deep learning approach for medical image analysis: diagnosis to detection. *Multimedia Tools and Applications, 82*(17), 26731-26769. [CrossRef]

[5] Latif, J., Xiao, C., Imran, A., & Tu, S. (2019, January). Medical imaging using machine learning and deep learning algorithms: a review. In *2019 2nd International conference on computing, mathematics and engineering technologies* (*iCoMET*) (pp. 1-5). IEEE. [CrossRef]

[6] Dastgir, A., Wang, B., Saeed, M. U., Sheng, J., & Saleem, S. (2025). MAFMv3: An automated multi-scale attention-based feature fusion MobileNetv3 for spine lesion classification. *Image and Vision Computing*, 155, 105440. [CrossRef]

[7] Grauslund, J. (2022). Diabetic retinopathy screening in the emerging era of artificial intelligence. *Diabetologia, 65*(9), 1415-1423. [CrossRef]

[8] Guefrachi, S., Echtioui, A., & Hamam, H. (2025). Diabetic retinopathy detection using deep learning multistage training method. *Arabian Journal for Science and Engineering, 50*(2), 1079-1096. [CrossRef]

[9] Kurup, G., Jothi, J. A. A., & Kanadath, A. (2021). Diabetic retinopathy detection and classification using pretrained inception-v3. In *2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)* (pp. 1–6). IEEE. [CrossRef]

[10] Bodapati, J. D., Shaik, N. S., & Naralasetti, V. (2021). Deep convolution feature aggregation: An application to diabetic retinopathy severity level prediction. *Signal, Image and Video Processing*, 15, 923–930. [CrossRef]

[11] Mohanty, C., Mahapatra, S., Acharya, B., Kokkoras, F., Gerogiannis, V. C., Karamitsos, I., & Kanavos, A. (2023). Using deep learning architectures for detection and classification of diabetic retinopathy. *Sensors, 23*(12), 5726. [CrossRef]

[12] Nahiduzzaman, M., Islam, M. R., Goni, M. O. F., Anower, M. S., Ahsan, M., Haider, J., & Kowalski, M. (2023). Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier. *Expert Systems with Applications, 217*, 119557. [CrossRef]

[13] Sacchini, F., Mancin, S., Cangelosi, G., Palomares, S. M., Caggianelli, G., Gravante, F., & Petrelli, F. (2025). The role of artificial intelligence in diabetic retinopathy screening in type 1 diabetes: A systematic review. *Journal of Diabetes and its Complications*, 109139. [CrossRef]

[14] Khalighi, S., Reddy, K., Midya, A., Pandav, K. B., Madabhushi, A., & Abedalthagafi, M. (2024). Artificial intelligence in neuro-oncology: advances and challenges in brain tumor diagnosis, prognosis, and precision treatment. *NPJ precision oncology, 8*(1), 80. [CrossRef]

[15] Xiao, S., Zhou, Y., Wu, Q., Wang, X., Hu, Y., Pan, Q., ... & Pan, D. (2022). Prevalence of cardiovascular diseases in relation to total bone mineral density and prevalent fractures: a population-based cross-sectional study. *Nutrition, Metabolism and Cardiovascular Diseases, 32*(1), 134-141. [CrossRef]

[16] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018, June). MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4510-4520). IEEE. [CrossRef]

[17] Kadam, K., Ahirrao, S., Kotecha, K., & Sahu, S. (2021). Detection and localization of multiple image splicing using MobileNet V1. *IEEE Access, 9*, 162499–162519. [CrossRef]

[18] Lian, X., Pang, Y., Han, J., & Pan, J. (2021). Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation. *Pattern Recognition, 110*, 107622. [CrossRef]

[19] Diabetic retinopathy detection. (n.d.). Kaggle: Your Machine Learning and Data Science Community. Retrieved from https://www.kaggle.com/competitions/diabetic-retinopathy-detection

[20] Gangwar, A. K., & Ravi, V. (2020). Diabetic retinopathy detection using transfer learning and deep learning. In *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020), Volume 1* (pp. 679-689). Singapore: Springer Singapore. [CrossRef]

[21] Maqsood, Z., & Gupta, M. K. (2022). Automatic detection of diabetic retinopathy on the edge. In *Cyber Security, Privacy and Networking: Proceedings of ICSPN 2021* (pp. 129-139). Singapore: Springer Nature Singapore. [CrossRef]

[22] Thomas, N. M., & Albert Jerome, S. (2021). Grading and classification of retinal images for detecting diabetic retinopathy using convolutional neural network. In *International Conference on Advances in Electrical and Computer Technologies* (pp. 607–614). Springer. [CrossRef]

[23] Ali, G., Dastgir, A., Iqbal, M. W., Anwar, M., & Faheem, M. (2023). A hybrid convolutional neural network model for automatic diabetic retinopathy classification from fundus images. *IEEE Journal of Translational Engineering in Health and Medicine, 11*, 341–350. [CrossRef]

Muhammad Usman Saeed received his Ph.D. in Computer Science and Engineering (2021-2025) from Central South University, China. He completed a B.S. degree in Information Technology from the University of Education, Lahore, Pakistan, in 2019 and an MS degree in Computer Science from the University of Okara, Pakistan, in 2021. From 2021 to 2022, he was a Lecturer at the University of Okara, Pakistan. His research interests include Medical Image Analysis, Computer Vision, Object Detection, Semantic Segmentation, and Medical Image Processing.

Aqsa Dastgir is currently pursuing a PhD. degree in Computer Science and Engineering from Central South University, China. She completed a B.S. degree in Information Technology from the University of Education, Lahore, Pakistan, in 2018 and an MS degree in Computer Science from the University of Okara, Pakistan, in 2021. Her research interests include Computer Vision, Bioinformatics, and Medical Image Analysis.