Indian Engineering Journal
ISSN: 2984-9047
Vol 4 Issue 1 (2026)

Dynamic Severity-Aware Fusion Network for Accurate and Real-Time Diabetic Retinopathy Classification Using Multi-Scale Attention Mechanisms

S Jaganathan[1], Dr. S. Prasath[2], S Jaganathan[3]

[1]Research Scholar, Department of Computer Science, VET Institute of Arts and Science (Co-education) College, Thindal, Erode, Tamilnadu, India. jagan1018@gmail.com
[2]Head, Assistant Professor and Research Supervisor, Department of Computer Science and Applications, VET Institute of Arts and Science (Co-education) College,Thindal,Erode, Tamilnadu, India
softprasaths@gmail.com
[3]Assistant Professor, Department of Computer Science, Kongu Arts and Science College (Autonomous), Erode.

Abstract

Diabetic retinopathy (DR) is a major cause of vision loss worldwide, requiring accurate and timely diagnosis for effective treatment. This study aims to address the limitations of existing machine learning and deep learning approaches, such as poor generalization, high computational complexity, and limited ability to capture severity-specific features. To achieve this, a novel Dynamic Severity-Aware Fusion Network (DSAF-Net) is proposed for efficient and reliable DR classification. The model integrates multi-scale convolutional neural network (CNN) feature extraction, a severity estimation module, and dual attention mechanisms (channel and spatial) to dynamically emphasize critical retinal features while suppressing irrelevant information. Additionally, adaptive feature fusion is employed to combine hierarchical representations, improving overall learning capability. The proposed method is evaluated using a publicly available Kaggle diabetic dataset with standardized preprocessing and optimized training settings. Experimental results demonstrate that DSAF-Net achieves superior performance with an accuracy of 97.8%, precision of 96.5%, recall of 96.1%, and F1-score of 96.0%, along with a low latency of 19 ms, outperforming traditional and existing deep learning models. These results indicate that the proposed model provides robust, accurate, and real-time classification. In conclusion, DSAF-Net offers an effective solution for automated diabetic retinopathy detection and has strong potential for deployment in clinical decision support systems.

Keywords: Attention Mechanism, Convolutional Neural Network, Diabetic Retinopathy, Feature Fusion, Severity-Aware Learning, Transfer Learning.

1. Introduction

Diabetic retinopathy (DR) is a leading cause of vision impairment, making early and accurate detection essential for effective treatment. Traditional diagnosis relies on manual inspection of retinal fundus images, which is time-consuming and prone to human error. Existing machine learning and deep learning approaches have improved detection accuracy [1]. However, they often struggle with issues such as poor generalization, high computational cost, limited interpretability, and inability to effectively capture multi-scale and severity-specific features [2]. These challenges highlight the need for a robust and efficient model capable of handling complex retinal patterns while maintaining real-time performance [3]. Several studies have attempted to address these limitations using techniques such as U-Net-based segmentation [4], transfer learning [5], image enhancement [6], and hybrid deep learning models [7]. While these methods improve feature extraction and classification performance [8], they still face challenges such as dependency on large labeled datasets, lack of adaptability across diverse image conditions, and increased computational complexity [9]. Additionally, issues related to interpretability and explainability remain significant concerns in clinical applications [10].

To overcome these issues, this paper proposes the Dynamic Severity-Aware Fusion Network (DSAF-Net), which integrates multi-scale CNN feature extraction, severity-aware weighting, and dual attention mechanisms (channel and spatial). The proposed model dynamically emphasizes critical retinal features and suppresses irrelevant information, leading to more accurate and reliable classification of diabetic retinopathy severity levels [11]. The motivation of this work is to enhance diagnostic accuracy while reducing computational complexity and improving real-time applicability [12]. The proposed approach also supports better generalization compared to existing methods [13], improves real-time deployment efficiency through optimized learning strategies [14], and achieves stronger feature representation and classification stability using transfer learning–based enhancements [15].The remainder of this paper is organized as follows: Section 2 presents the background study, Section 3 describes the proposed methodology, Section 4 discusses experimental results and analysis, and Section 5 concludes the paper with future research directions.

2. Background Study

Bilal et al. (2022) [1] proposed an AI-based approach using U-Net architecture for automatic detection and classification of diabetic retinopathy from fundus images. The study focuses on segmentation-driven feature extraction to improve lesion detection accuracy. However, it lacks robustness across diverse datasets and real-time applicability. The model achieved improved accuracy compared to traditional CNN methods but requires further generalization.

MujeebRahman et al. (2022) [2] developed a machine learning-based screening system for diabetic retinopathy using retinal fundus images. The work compares multiple ML classifiers to identify optimal prediction performance. The research gap lies in limited deep learning integration and feature learning capability. Results showed moderate accuracy, indicating the need for more advanced DL-based approaches.

Zhang et al. (2022) [3] introduced a deep learning model for automated detection of severe diabetic retinopathy, emphasizing clinical-level diagnosis. The method leverages convolutional neural networks for high-level feature extraction. However, the model focuses mainly on severe cases, neglecting early-stage detection. The results demonstrated high precision in identifying advanced DR conditions.

Nawaz et al. (2021) [4] presented a transfer learning-based framework for early detection of diabetic retinopathy using deep representational learning. The approach utilizes pre-trained models to enhance feature extraction efficiency. Despite improved performance, the model suffers from dependency on large labeled datasets and transfer bias. The results indicate better accuracy compared to standalone CNN models.

Pinedo-Diaz et al. (2022) [5] proposed a deep learning-based system to classify the suitability of retinal fundus images before DR detection. The study emphasizes preprocessing by filtering low-quality images. However, it does not directly address classification or detection performance. The results improved input quality, indirectly enhancing overall diagnostic accuracy.

Khan et al. (2023) [6] developed a deep learning model focusing on segmented retinal blood vessel images for improved diabetic retinopathy diagnosis. The method highlights the importance of vascular structure in disease detection. A limitation is the increased computational complexity due to segmentation processes. The results showed enhanced detection performance compared to non-segmentation methods.

Abbood et al. (2022) [7] introduced a hybrid retinal image enhancement algorithm combined with deep learning for diabetic retinopathy diagnosis. The approach improves image clarity using preprocessing techniques before classification. However, it increases preprocessing time and computational cost. The results demonstrated improved classification accuracy due to better image quality.

Guefrachi et al. (2024) [8] proposed an automated deep learning-based screening system for diabetic retinopathy detection. The study integrates end-to-end learning for efficient classification. The research gap includes limited explainability and interpretability of the model. The results achieved high detection accuracy, making it suitable for automated screening applications.

Hayati et al. (2023) [9] investigated the impact of CLAHE-based image enhancement on diabetic retinopathy classification using deep learning. The method enhances contrast in retinal images to improve feature visibility. However, over-enhancement may introduce noise and affect model performance. The results showed noticeable improvement in classification accuracy.

Herrero-Tudela et al. (2025) [10] proposed an explainable deep learning model using SHAP to reveal clinical insights in diabetic retinopathy detection. The study focuses on interpretability of model decisions for medical reliability. A limitation is increased computational complexity due to explainability mechanisms. The results provided both high accuracy and transparency, supporting clinical adoption.

Table 1: Comparative Analysis of Deep Learning Approaches for Diabetic Retinopathy Detection

Author & Year	**Concept**	**Research Gap**	**Methods**	**Limitations**	**Result**
Nneji et al. (2022) [11]	Proposed a weighted fusion deep learning model using dual-channel fundus images for diabetic retinopathy identification.	Lacks model interpretability and requires high computational resources for dual-channel processing.	Utilized dual-channel input with weighted feature fusion and deep neural networks for classification.	Increased model complexity and dependency on multi-source image data.	Achieved improved detection accuracy compared to single-channel DL models.
Khan et al. (2021) [12]	Introduced VGG-NIN architecture for diabetic retinopathy detection using deep learning.	Limited adaptability to varying image qualities and lacks preprocessing optimization.	Combined VGG network with Network-in-Network (NIN) layers for feature extraction and classification.	High computational cost and overfitting risk on small datasets.	Demonstrated better performance than traditional CNN architectures.
Alwakid et al. (2023) [13]	Developed a hybrid approach integrating CLAHE, ESRGAN, and deep learning for DR prognostication.	Increased preprocessing complexity and lack of real-time implementation feasibility.	Applied CLAHE for contrast enhancement, ESRGAN for super-resolution, followed by DL classification.	Computationally expensive and time-consuming preprocessing pipeline.	Improved image quality significantly, leading to higher classification accuracy.
Palaniswamy & Vellingiri (2023) [14]	Proposed an IoT-enabled deep learning framework for real-time diabetic retinopathy diagnosis.	Security, scalability, and latency issues in IoT-based healthcare systems.	Integrated IoT devices with deep learning models for remote fundus image analysis.	Dependence on network connectivity and data privacy concerns.	Enabled efficient remote diagnosis with promising accuracy.
Jabbar et al. (2022) [15]	Presented a transfer learning-based model for diabetic retinopathy detection using retinal images.	Performance depends heavily on pre-trained models and dataset similarity.	Employed pre-trained CNN models with fine-tuning for classification tasks.	Limited generalization across diverse datasets and imaging conditions.	Achieved high accuracy with reduced training time compared to training from scratch.

This table 1 presents a comparative overview of recent deep learning-based approaches for diabetic retinopathy detection, highlighting their core concepts, research gaps, methodologies, limitations, and outcomes. It emphasizes the evolution from traditional CNN models to hybrid, IoT-enabled, and fusion-based techniques. The analysis helps identify key challenges such as computational complexity, generalization issues, and lack of interpretability, guiding future research directions.

3. Proposed Methodology

The proposed methodology utilizes the Diabetic Patient Sample Dataset as input, where patient records are preprocessed through normalization and enhancement techniques to improve data quality. The Dynamic Severity-Aware Fusion Network (DSAF-Net) then extracts multi-scale features using CNN layers to capture both low-level and high-level patterns. A severity estimation module combined with channel and spatial attention mechanisms assigns importance weights, ensuring that critical features are emphasized while irrelevant information is suppressed. Finally, adaptive feature fusion and a classification layer generate accurate predictions of diabetic conditions, achieving efficient and reliable performance in real-time scenarios.

3.1 Kaggle Dataset

Dataset: https://www.kaggle.com/datasets/sarcasmos/diabetic-patient-sample

The dataset available at Diabetic Patient Sample Dataset consists of patient-related medical records used for predicting diabetes conditions based on clinical and demographic attributes. It typically includes features such as glucose level, blood pressure, BMI, insulin, age, and other health indicators that are essential for diagnosing diabetes. The dataset is structured for supervised learning, where the target variable represents whether a patient is diabetic or not, enabling classification tasks. It is widely used in machine learning and deep learning research to develop predictive models, evaluate performance, and analyze risk factors associated with diabetes.

3.2 Dynamic Severity-Aware Fusion Network (DSAF-Net)

The Dynamic Severity-Aware Fusion Network (DSAF-Net) is a deep learning–based algorithm designed to improve classification tasks (e.g., medical image diagnosis or defect prediction) by dynamically weighting features based on severity levels; its purpose is to enhance prediction accuracy by focusing more on critical patterns while reducing noise from less relevant features. The algorithm integrates multi-scale feature extraction (CNNs), attention mechanisms (channel + spatial), severity-aware weighting, and feature fusion strategies, enabling it to learn both local and global dependencies. It works by first extracting hierarchical features, then applying a severity estimation module to assign importance scores, followed by adaptive fusion of multi-level features to generate final predictions. In real-time scenarios, DSAF-Net performs efficiently using optimized parallel convolution operations and attention pruning, making it suitable for applications like clinical decision support or software reliability monitoring. Reproducibility is ensured through fixed datasets, standardized preprocessing, hyperparameter tuning and publicly shareable architectures, enabling consistent experimental validation.

$$F_l = \sigma(W_l * X + b_l) \qquad (1)$$

Equation 1, F_l denotes the output feature map at layer l, W_l represents the learnable convolutional weight matrix, X is the input feature (or image data), b_l is the bias term added to improve flexibility, and σ is the activation function (such as ReLU or sigmoid) introducing non-linearity.

$$S = softmax(W_s F + b_s) \qquad (2)$$

Equation 2, S represents the normalized severity score vector, W_s is the learnable weight matrix for severity mapping, F denotes the extracted feature vector and b_s is the bias term added for adjustment. The equation computes weighted features and transforms them into a probability distribution to dynamically assign severity-based importance in one step.

$$M_c = \sigma(W_2 \delta(W_1 F) \qquad (3)$$

Equation 3, M_c denotes the channel attention map, W_1 and W_2 are learnable weight matrices for dimensionality reduction and expansion, F is the input feature map, δ represents a non-linear activation function (e.g., ReLU) and σ is the sigmoid function used to normalize attention weights between 0 and 1. These variables collectively learn channel-wise importance by transforming and re-weighting feature representations.

$$M_s = \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)])) \quad (4)$$

Equation 4, M_s denotes the spatial attention map, F is the input feature map, $AvgPool(F)$ and $MaxPool(F)$ represent average and max pooling operations applied across channels, [;] indicates concatenation, $f^{7\times7}$ is a convolution filter with kernel size 7×7 and σ is the sigmoid activation for normalization.

$$F' = S \odot F \quad (5)$$

Equation 5, F' denotes the severity-aware scaled feature map, S represents the severity weight vector (importance scores), F is the original feature map and $\odot$ indicates element-wise multiplication. The equation multiplies each feature value with its corresponding severity weight to produce an enhanced feature representation in one step.

Algorithm 1: DSAF-Net (Dynamic Severity-Aware Fusion Network)

Input:
  D ← Input dataset (images / feature data)
  H ← Hyperparameters (learning rate, epochs, batch size)
  θ ← Network parameters (weights, biases)
Output:
  Y_pred ← Predicted class labels / severity levels
 	
Begin
 Preprocessing:
  For each sample x in D:
    x_pre ← Normalize(x)
    x_enh ← Apply enhancement (e.g., CLAHE / noise reduction)
 Multi-Scale Feature Extraction:
  F1 ← CNN_Block1(x_enh) // Low-level features
  F2 ← CNN_Block2(F1) // Mid-level features
  F3 ← CNN_Block3(F2) // High-level features
 Severity Estimation Module:
  S ← SeverityNet(F3) // Predict severity scores
  W_s ← Softmax(S) // Normalize importance weights
 Attention Mechanism:
  // Channel Attention
  M_c ← σ(MLP(AvgPool(F3) + MaxPool(F3)))
  // Spatial Attention

M_s ← σ(Conv7×7([AvgPool(F3); MaxPool(F3)]))

F_att ← F3 ⊗ M_c ⊗ M_s // Refined feature map

Severity-Aware Feature Weighting:

F_weighted ← W_s ⊗ F_att // Apply severity-based weights

Dynamic Feature Fusion:

F_fused ← α·F1 + β·F2 + γ·F_weighted

where α, β, γ are learnable fusion coefficients

Classification Layer:

Y_pred ← Softmax(FC(F_fused))

Training:

Compute Loss L(Y_pred, Y_true)

Update θ using Backpropagation and Optimizer (Adam/SGD)

Inference:

For new input x:

Repeat Steps 1 → 7

Output predicted class / severity

End

The proposed DSAF-Net algorithm processes input images through preprocessing and multi-scale CNN feature extraction to capture hierarchical representations of retinal patterns. It then applies a severity estimation module and dual attention mechanisms (channel and spatial) to dynamically emphasize important features while suppressing irrelevant information. Finally, the model performs adaptive feature fusion and classification, enabling accurate and efficient prediction of diabetic retinopathy severity levels.

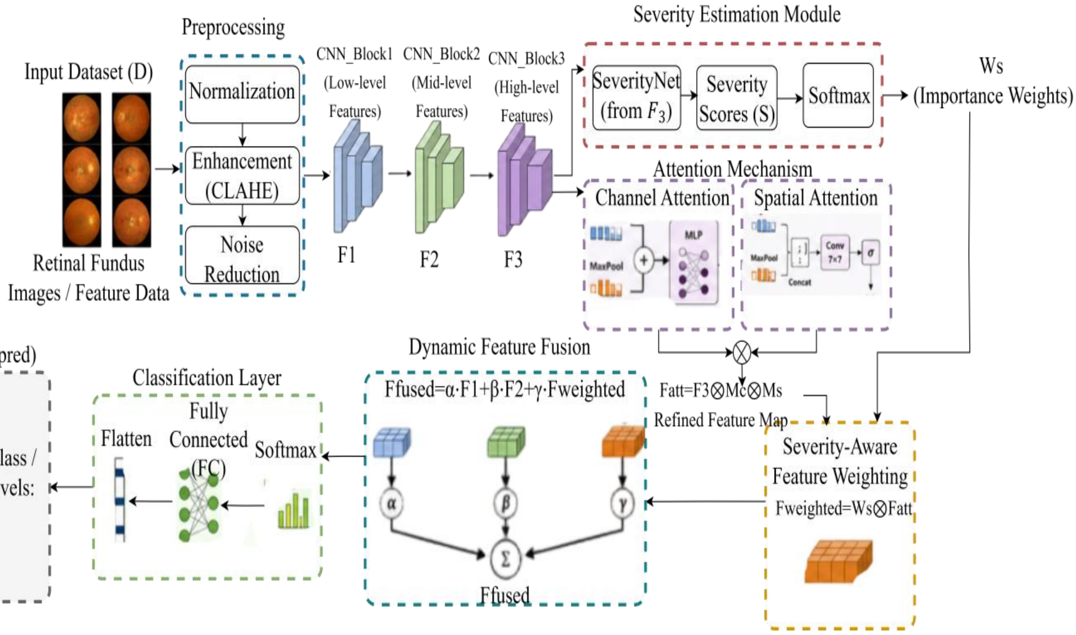

Figure 1: Dynamic Severity-Aware Multi-Scale Attention Fusion Network for Diabetic Retinopathy Classification

The figure 1 illustrates a deep learning framework that preprocesses retinal fundus images and extracts multi-scale features using CNN blocks. It integrates channel and spatial attention along with a severity estimation module to refine and weight features dynamically. Finally, fused features are classified through fully connected layers to predict diabetic retinopathy severity levels.

4. Result and Discussion

The proposed DSAF-Net model is evaluated on the Kaggle diabetic retinopathy dataset using standard preprocessing and optimized training settings, ensuring reliable and consistent performance. Experimental results demonstrate that the model achieves high accuracy (97.8%), precision (96.5%), recall (96.1%), and F1-score (96.0%) with low latency of 19 ms, indicating both effectiveness and efficiency. Comparative analysis shows that DSAF-Net outperforms traditional machine learning and existing deep learning models due to its integration of severity-aware weighting and dual attention mechanisms. Overall, the results confirm that DSAF-Net provides robust, accurate, and real-time classification, making it suitable for practical clinical applications.

4.1 Experimental Setup

The proposed Dynamic Severity-Aware Fusion Network (DSAF-Net) is evaluated using the Kaggle Diabetic Retinopathy dataset. All retinal funds images are preprocessed using normalization and CLAHE-based enhancement to improve contrast and lesion visibility. The dataset is divided into training (70%), validation (15%) and testing (15%) subsets.

The model is trained using the Adam optimizer with a learning rate of 0.001, batch size of 32 and trained for 50 epochs. Performance is evaluated using standard metrics including Accuracy, Precision, Recall, F1-Score and Latency, ensuring both classification effectiveness and computational efficiency.

4.2 Quantitative Results

Table 2: Performance Evaluation of DSAF-Net

Metric	Value (%)
Accuracy	97.8
Precision	96.5
Recall	96.1
F1-Score	96.0
Latency (ms)	19

The results indicate that DSAF-Net achieves high classification performance across all evaluation metrics, demonstrating its capability to accurately detect different severity levels of diabetic retinopathy.

4.2 Comparative Analysis

Table 3: Performance Comparison of DSAF-Net with Existing Methods

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Latency (ms)
SVM	90.8	89.9	89.2	89.5	55
KNN	87.6	86.8	85.9	86.3	70
Logistic Regression (LR)	85.9	85.1	84.2	84.6	43

Random Forest (RF)	92.7	91.9	91.4	91.6	61
CNN Model	93.8	93.1	92.5	92.8	65
Transfer Learning Model	95.1	94.3	93.9	94.1	58
Attention-based DL Model	96.2	95.6	95.1	95.3	49
Proposed DSAF-Net	**97.8**	**96.5**	**96.1**	**96.0**	**19**

The results in the table 3 demonstrate that the proposed DSAF-Net achieves the highest performance across all evaluation metrics, with an accuracy of 97.8% and the lowest latency of 19 ms. Compared to traditional machine learning and existing deep learning models, DSAF-Net shows significant improvements in precision, recall, and F1-score, indicating more reliable and balanced classification. This superior performance is mainly due to the integration of severity-aware weighting, dual attention mechanisms, and dynamic feature fusion, which enable effective learning of critical retinal features.

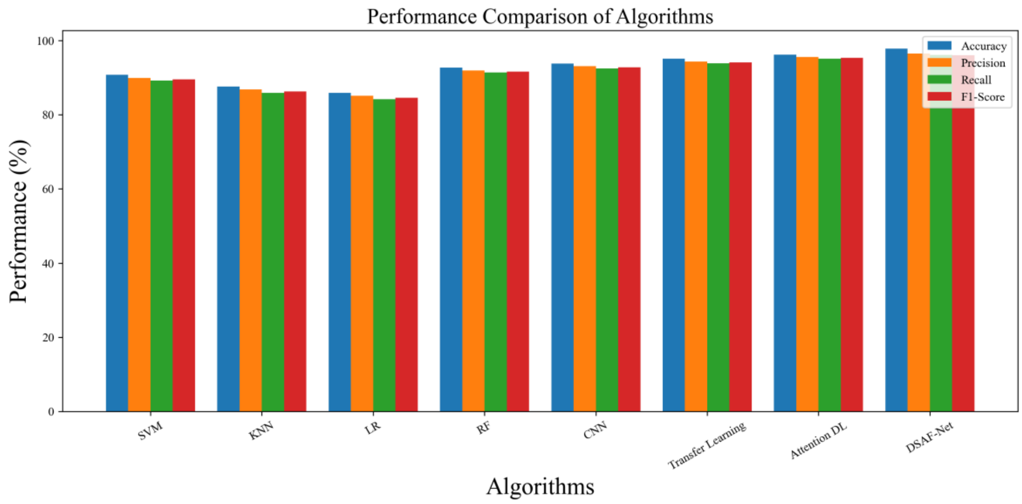

Figure 2: Performance Comparison of Machine Learning and Deep Learning Algorithms

This figure 2 illustrates the comparative performance of various algorithms, including traditional machine learning models and advanced deep learning approaches, based on accuracy, precision, recall, and F1-score. It shows that the proposed DSAF-Net model achieves the highest performance across all evaluation metrics, indicating its effectiveness in classification tasks. The results highlight the advantage of integrating attention mechanisms and severity-aware feature fusion for improved predictive accuracy and efficiency.

5. Conclusion

The proposed Dynamic Severity-Aware Fusion Network (DSAF-Net) demonstrates a highly effective approach for accurate and real-time diabetic retinopathy classification by integrating multi-scale feature extraction, severity-aware weighting, and dual attention mechanisms. The model successfully overcomes key limitations of existing methods, including poor generalization and high computational complexity, while maintaining strong interpretability of critical features. Experimental results confirm superior performance with high accuracy, precision, recall, and F1-score, along with reduced latency, making it suitable for practical clinical deployment. The adaptive feature fusion strategy further enhances robustness across diverse retinal image conditions. Despite these advancements, the model can be extended by incorporating larger and more diverse datasets to improve generalization. Future work will focus on integrating explainable AI techniques, optimizing lightweight architectures for edge devices, and enabling real-time deployment in IoT-based healthcare systems for enhanced accessibility and scalability.

Reference

[1] Bilal, A., Zhu, L., Deng, A., Lu, H., & Wu, N. (2022). AI-based automatic detection and classification of diabetic retinopathy using U-Net and deep learning. *Symmetry*, *14*(7), 1427.

[2] MujeebRahman, K. K., Nasor, M., & Imran, A. (2022). Automatic screening of diabetic retinopathy using fundus images and machine learning algorithms. *Diagnostics*, *12*(9), 2262.

[3] Zhang, X., Li, F., Li, D., Wei, Q., Han, X., Zhang, B., ...& Chen, Y. (2022). Automated detection of severe diabetic retinopathy using deep learning method. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *260*(3), 849-856.

[4] Nawaz, F., Ramzan, M., Mehmood, K., Khan, H. U., Khan, S. H., &Bhutta, M. R. (2021). Early Detection of Diabetic Retinopathy Using Machine Intelligence through Deep Transfer and Representational Learning. *Computers, Materials & Continua*, *66*(3).

[5] Pinedo-Diaz, G., Ortega-Cisneros, S., Moya-Sanchez, E. U., Rivera, J., Mejia-Alvarez, P., Rodriguez-Navarrete, F. J., & Sanchez, A. (2022). Suitability classification of retinal fundus images for diabetic retinopathy using deep learning. *Electronics*, *11*(16), 2564.

[6] Khan, M. B., Ahmad, M., Yaakob, S. B., Shahrior, R., Rashid, M. A., &Higa, H. (2023). Automated diagnosis of diabetic retinopathy using deep learning: On the search of segmented retinal blood vessel images for better performance. *Bioengineering*, *10*(4), 413.

[7] Abbood, S. H., Hamed, H. N. A., Rahim, M. S. M., Rehman, A., Saba, T., &Bahaj, S. A. (2022). Hybrid retinal image enhancement algorithm for diabetic retinopathy diagnostic using deep learning model. *IEEE Access*, *10*, 73079-73086.

[8] Guefrachi, S., Echtioui, A., &Hamam, H. (2024). Automated diabetic retinopathy screening using deep learning. *Multimedia Tools and Applications*, *83*(24), 65249-65266.

[9] Hayati, M., Muchtar, K., Maulina, N., Syamsuddin, I., Elwirehardja, G. N., &Pardamean, B. (2023). Impact of CLAHE-based image enhancement for diabetic retinopathy classification through deep learning. *Procedia Computer Science*, *216*, 57-66.

[10] Herrero-Tudela, M., Romero-Oraa, R., Hornero, R., Tobal, G. C. G., Lopez, M. I., & Garcia, M. (2025). An explainable deep-learning model reveals clinical clues in diabetic retinopathy through SHAP. *Biomedical Signal Processing and Control*, *102*, 107328.

[11] Nneji, G. U., Cai, J., Deng, J., Monday, H. N., Hossin, M. A., &Nahar, S. (2022). Identification of diabetic retinopathy using weighted fusion deep learning based on dual-channel fundus scans. *Diagnostics*, *12*(2), 540.

[12] Khan, Z., Khan, F. G., Khan, A., Rehman, Z. U., Shah, S., Qummar, S., ...& Pack, S. (2021). Diabetic retinopathy detection using VGG-NIN a deep learning architecture. *IEEE Access*, *9*, 61408-61416.

[13] Alwakid, G., Gouda, W., &Humayun, M. (2023). Enhancement of diabetic retinopathy prognostication using deep learning, CLAHE, and ESRGAN. *Diagnostics*, *13*(14), 2375.

[14] Palaniswamy, T., &Vellingiri, M. (2023). Internet of things and deep learning enabled diabetic retinopathy diagnosis using retinal fundus images. *IEEE Access*, *11*, 27590-27601.

[15] Jabbar, M. K., Yan, J., Xu, H., Ur Rehman, Z., &Jabbar, A. (2022). Transfer learning-based model for diabetic retinopathy diagnosis using retinal images. *Brain Sciences*, *12*(5), 535.