



HAF-Net: Hybrid Attention Fusion Network for Diabetic Retinopathy Classification

Sirichai Triamlumlerd¹, Pauline Kongsuwan² and Anuruk Prommakhot³

ABSTRACT

Diabetic retinopathy (DR) is a serious diabetes complication that can cause blindness if not treated promptly. Accurate DR classification plays an important role in supporting diagnosis. To improve classification performance, this research proposed a hybrid attention fusion network (HAF-Net) for DR classification. The HAF-Net integrates a convolutional operator with a convolutional block attention module (CBAM) and a bottleneck attention module (BAM) due to their advantages in enhancing feature representation, focusing on both local and global contextual information. These strengths allow the network to improve its ability to distinguish between subtle differences and help gather important features in medical images. A residual layer is also added to compensate for feature loss during training. A hybrid fusion of multi-layer perceptron (MLP), long short-term memory (LSTM), and bidirectional LSTM (BiLSTM) features are employed to achieve DR classification. The HAF-Net model is tested on the a retinal fundus image dataset published on Kaggle for DR classification research, covering 11 disease types. The proposed HAF-Net achieved a training accuracy of 99.35% with a training loss of 0.0129. The hybrid model-MLP achieved the highest validation accuracy of 76.41% and test accuracy of 80.17%, demonstrating the best generalization performance. The results indicate that the HAF-Net achieves competitive classification performance compared with previous studies and represents a promising approach for automated DR classification.

Article information:

Keywords: Deep Learning, Convolutional Block Attention Module, Bottleneck Attention Module, Diabetic Retinopathy

Article history:

Received: December 4, 2025

Revised: April 22, 2026

Accepted: July 30, 2026

Published: August 22, 2026

(Online)

DOI: 10.37936/ecti-cit.2026204.265089

1. INTRODUCTION

Diabetic retinopathy (DR) is a complication of diabetes in which the retinal capillaries malfunction, potentially causing swelling of the blood vessels at the back of the eye. It generally appears in people who cannot control their blood sugar, resulting in visual impairment and potentially leading to permanent vision loss which is one of the leading causes of blindness globally [1]-[2]. DR has two stages, non-proliferative and proliferative [3]-[4]. Non-proliferative DR is defined by degeneration of blood vessel walls to a lesser or greater extent, while proliferative DR comes with abnormal growth of blood vessels in the retina. Due to its progressive and often asymptomatic nature, early detection is crucial to prevent irreversible vision loss, yet clinical diagno-

sis remains time-consuming and resource-intensive as it relies on expert ophthalmologists. Hence the recent works have been motivated in developing suitable automated analysis techniques, especially for abnormality detections like microaneurysms, hemorrhages, fluid exudates, capillary degeneration, pigment epithelium detachment and other anomalies in retinal images [5]-[6]. Predominantly driven by deep learning (DL) algorithms [7]-[9], these automated methods are utilized as highly accurate and efficient clinical decision support systems, aiming to reduce the diagnostic burden on ophthalmologists.

In recent times, DL has recently gained importance when tackling medical image analysis, a prime example being retinal classification for diagnosing DR [9]-[10]. Convolutional neural networks (CNN) have be-

^{1,2,3}The authors are with the Department of Computer Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand, Email: sirichai.t@en.rmutt.ac.th, pauline.k@en.rmutt.ac.th and anuruk.p@en.rmutt.ac.th

³Corresponding author: anuruk.p@en.rmutt.ac.th

come a dominant approach in this domain due to their ability to automatically extract hierarchical features. CNN have become a leading DL technique because of their ability to automatically extract unique features and simplify the decision-making process. Existing studies have explored various strategies, including segmentation-based methods (e.g., CNN with U-Net), transfer learning using pretrained models such as VGG-16 and MobileNet-V2, and hybrid architectures integrating attention mechanisms and feature optimization techniques. For instance, a study used DL to solve the problem of data imbalance by choosing CNN integrated with U-Net for lesion segmentation. To improve the discrimination of the lesions, Min-max normalization and CLAHE in combination with resizing as preprocessing techniques to process MESSIDOR dataset [11]. Similarly, an alternative method used pretrained CNN such as VGG-16 and MobileNet-V2 for classification of DR [12]. More advanced approaches incorporate hybrid CNN architectures, attention mechanisms, and efficient networks (e.g., ResNet, Inception, EfficientNet) to enhance feature extraction and classification performance across datasets such as Messidor-2 and IDRiD [13]-[17]. Several other studies [18]-[19] have reported that the use of DL methods can significantly improve DR classification, detection, and recognition, thereby facilitating rapid DR diagnosis.

However, despite these advancements, DL-based methods for DR classification still face several limitations. These include dependence on large-scale annotated datasets, sensitivity to image quality variations, and challenges in capturing both fine-grained local features and global contextual information simultaneously. In addition, variations in disease characteristics, the complexity of deep neural networks, and limitations in large-scale and multi-scale feature learning further affect model performance. Moreover, many approaches rely on single attention mechanisms, which may not sufficiently model complex retinal structures, leading to suboptimal feature representation [9]-[14], [16], [18]. To address these challenges, this study proposes a HAF-Net, which integrates convolutional operators with CBAM and BAM to enhance feature representation. CBAM focuses on the extraction of essential spatial features, whereas BAM enhances the detection of deep features inside the convolutional operators. The integration of these complementary mechanisms enables more robust and discriminative feature learning. The model also employs a skip connection through a residual layer that allows the network to re-learn data from previous layers while making the learning process more stable and facilitating the extraction of complex features. In addition, hybrid classifiers including MLP, LSTM, and BiLSTM are incorporated to further improve classification performance. This research used the Retinal Fundus Images (RFI) dataset, published on Kag-

gle for DR classification research, covering 11 disease types. Experimental results were evaluated using accuracy, precision, recall, and F1-score, along with confusion matrix analysis to assess classification performance. The proposed framework aims to improve feature representation and generalization for more reliable DR classification.

2. METHOD

2.1 Convolutional neural network

The convolutional operator is the core process in CNN [11], [13]-[17], [20] that performs feature extraction from data. A small filter or kernel is used to perform convolution across the input data to compute the output at each location on the feature map, as follows:

$$Y(i, j) = \sum_{m, n}^{m, n} X(i + m, j + n) * K(m, n) \quad (1)$$

Where $Y(i, j)$ is the output at position (i, j) , $X(i + m, j + n)$ is the value of the input data, $K(m, n)$ is the kernel size and $*$ is multiplication followed by summation in the convolution operation.

2.2 Residual layer

Residual layer (RL) [20]-[21] is a technique used to address the vanishing gradient problem in deep networks by introducing shortcut connections that allow information to flow more efficiently through the network. Instead of learning the target function $H(x)$ directly, the network learns a residual function $F(x) = H(x) - x$, and then reconstructs the output as $H(x) = F(x) + x$. This formulation enables smoother training, even in very deep models. Residual connections also help preserve critical information from earlier layers, leading to more stable training and better convergence, and an improved ability to capture complex characteristics of the data, as follows:

$$H(x) = F(x; W_i) + x \quad (2)$$

Where $H(x)$ is the output of the residual block, $F(x; W_i)$ is the function learned through convolutional layers with parameters W_i and x is the input connected via a shortcut connection.

2.3 Convolutional block attention module

The CBAM [22] is an attention module that enhances CNN learning by focusing on important features in both channel and spatial modes through sequential operation. In the channel attention stage, a feature map $F \in \mathbb{R}^{H \times W \times C}$ from the previous convolutional layer is used, where H , W and C denote height, width, and channels. Channel importance is

computed using a global average pooling (GAP) and global max pooling (GMP), as follows:

$$F_{avg}^c = GAF(F), F_{avg}^c = GAP(F) \quad (3)$$

Next, the results from both pooling processes are passed through a MLP to calculate the weights of each channel, as follows:

$$M_C(F) = \sigma(MLP(F_{avg}^c) + MLP(F_{max}^c)) \quad (4)$$

Where σ is the sigmoid function, which is used to generate channel mask M_C that highlights the important output features, and the mask is then combined with the original feature map to obtain an output that emphasizes only the important channels.

$$F' = M_C(F) * F \quad (5)$$

Next, CBAM uses spatial attention to locate important regions by applying max-pooling (MaxP) and average-pooling (AvgP) along spatial dimensions, generating two spatial maps as follows:

$$F_{avg}^c = Average(F', axis = C) \quad (6)$$

$$F_{max}^c = MAX(F', axis = C) \quad (7)$$

Next, the two results are then concatenated and processed using a convolution kernel to generate a spatial attention map, as follows:

$$M_S(F') = \sigma(ConV_{7 \times 7}([F_{avg}^S, F_{max}^S])) \quad (8)$$

Where σ is the sigmoid function, it generates an attention map highlighting important spatial regions. The spatial attention map is then multiplied with the input feature map to produce the final output, emphasizing key features in both channel and spatial dimensions, as follows:

$$F'' = M_S(F') * F' \quad (9)$$

Where F'' is the feature map that is passed to the next stage in the sequence.

2.4 Bottleneck attention mechanism

The BAM [23] is an attention module integrated into CNNs to enhance feature representation through parallel channel and spatial attention branches, whose outputs are combined to generate a final attention map, as follows:

$$F_{avg}^c = GAP(F) \quad (10)$$

Where $F \in \mathbb{R}^{H \times W \times C}$ is spatially averaged across each channel to extract information representing overall channel importance, the resulting values are passed through a two-layer MLP with a ReLU activation in the hidden layer and a sigmoid activation

at the output. This transforms the values into the range $[0,1]$, as follows:

$$M_{ca} = \sigma(MLP(F_{avg}^c)) \quad (11)$$

Next, BAM assesses spatial attention importance by applying multiple dilated convolutions to the feature map F , expanding the receptive field to cover a larger area. A convolution size is then used for dimensionality reduction, followed by a sigmoid activation to construct the spatial attention map, as follows:

$$M_{ca} = \sigma(ConV_{1 \times 1}(DilatedConV(F))) \quad (12)$$

Finally, BAM combines the channel attention map and the spatial attention map elementwise using broadcasting, scales them, and multiplies them to generate the final attention map, as follows:

$$M = \sigma(M_{ca} * M_{ca}) \quad (13)$$

This attention map is multiplied with the original feature map via a residual attention connection, as follows:

$$\hat{F}' = (1 + M) * F \quad (14)$$

Where the addition of $1 + M$ in the residual attention mechanism helps retain the original information while amplifying important regions. The resulting $\hat{F}$ is then passed to the next layer of the model to support more accurate learning in medical imaging.

2.5 Classification techniques

This study applied MLP [24], LSTM [25], and BiLSTM [26] for the network structure above in processing medical sequence data more efficiently. On the one hand, MLP can extract deep features for structured data; on the other hand, LSTM is able to capture long-term dependencies in sequential data by learning temporal patterns and thereby avoid the vanishing gradient problem. Furthermore, BiLSTM can learn contextual information from both sides within a sequence, thereby providing bidirectional context-based learning and improving model interpretation or prediction in time series data.

3. PROPOSED METHOD

In Fig. 1, the HAF-Net is designed for DR classification. The network operates in Block A and Block B. In Block A, the process begins by receiving input images from the dataset, which are then processed through a convolutional layer, as detailed in Equation (1), the network designs a 3-layer convolutional filter with sizes of 32, 64, and 128 filters (3×3 kernel size), respectively, with a 2×2 MaxP to retain only important features and ReLU to extract non-linear features, working together with SAME padding to pre-

serve spatial information during convolution, while MaxPooling reduces the feature map size. Then, the obtained features are passed to CBAM and BAM to enhance emphasis on important positions and channels in the image. CBAM is an attention mechanism consisting of two main parts: channel attention and spatial attention. In the channel attention part, it starts by using GAP and GMP to extract the statistical characteristics of each channel, as detailed in Equation (3). Then, the result of the pooling process is passed through a MLP implemented using 1×1 convolution layers with reduction ratio ($r = 16$), as detailed in Equation (4), to learn the importance of each channel using the ReLU and sigmoid functions, where the sigmoid function sets weights in the range $[0, 1]$ to create a channel mask that emphasizes only important channels. This mask is then combined with the original feature map elementwise, as detailed in Equation (5). Next, spatial attention operates using AvgP and MaxP, as detailed in Equation (6) and (7), applied along the channel axis and concatenated, to emphasize spatially important features. And they are combined and passed through a 7×7 convolutional layer (7×7 kernel size), which works with a sigmoid function to construct a spatial attention map, as detailed in Equation (8). Finally, the attention map is combined with to extract the important features, as detailed in Equation (9), and is passed to the next layer to be merged with the BAM features. BAM is an attention module that works with a convolutional layer to highlight salient features from both channel and spatial perspectives. It consists of two parallel branches: channel attention and spatial at-

tention. For channel attention, it begins by applying GAP to the feature map to extract the overall importance of each channel, as detailed in Equation (10). The pooled features are passed through a two-layer fully connected layers with reduction ratio ($r = 16$), using ReLU and sigmoid activations to generate channel attention weights, as detailed in Equation (11). In the spatial branch, a sequence of convolutional layers is applied, including a 1×1 convolution followed by a 3×3 dilated convolution (*dilation rate: 4*), and a final 1×1 convolution with sigmoid activation to generate the spatial attention map, as described in Equation (12). After obtaining both channel and spatial attention, BAM combines the two outputs elementwise using broadcasting and multiplication to generate the final attention map, as detailed in Equation (13). This attention map is then multiplied with the original feature map using residual connections, as described in Equation (2), to preserve original information and improve the accuracy of highlighting important features, as described in Equation (14). BAM helps the model focus more effectively on features, important in terms of both position and channel. In this process, the features from CBAM and BAM are concatenated and passed to Block B, which has the same workflow as Block A, with additional residual connections and sequential feature modeling to further improve feature representation and learning stability. The resulting features are then forwarded to the classification techniques.

After feature fusion, the resulting feature maps are reshaped into sequences ($H \times W \times C$) and passed to classification layers. The features are processed using

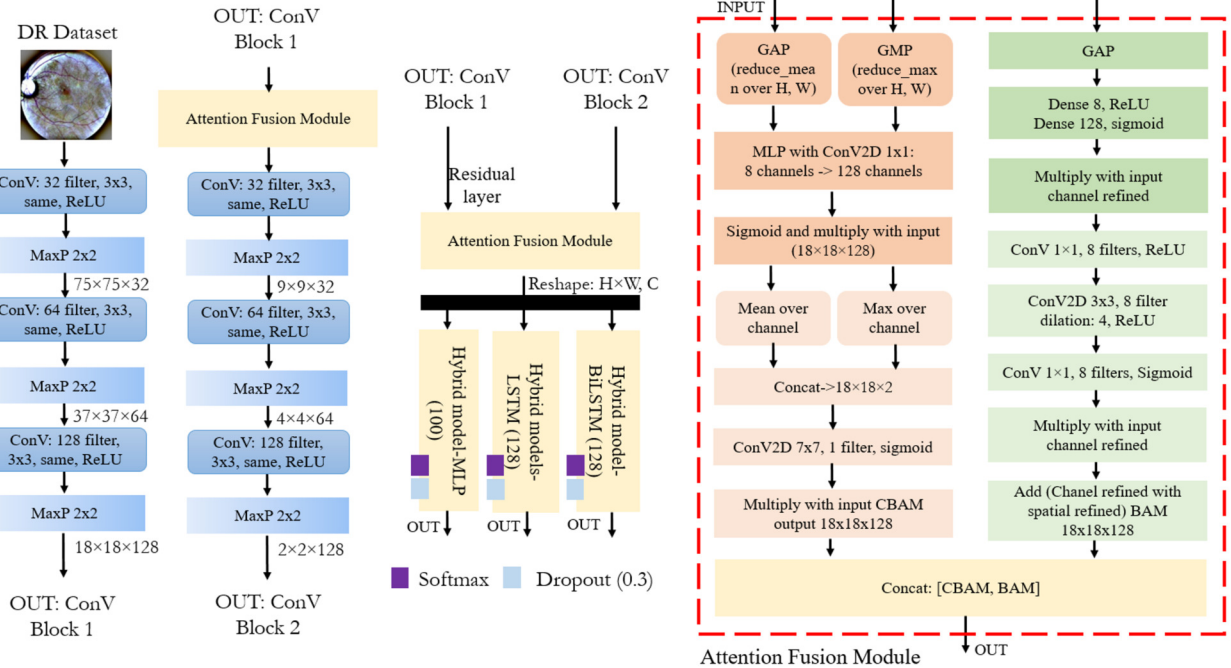


Fig.1: Proposed method.

an MLP with 256 nodes and 0.3 dropout, as well as, LSTM and BiLSTM layers with 128 hidden units and 0.3 dropout, followed by a softmax layer for DR classification. Accordingly, the HAF-Net is divided into three networks: hybrid model-MLP, hybrid model-LSTM and hybrid model-BiLSTM. The experimental and classification results are presented in Fig. 3 to 6 and Table 2.

4. EXPERIMENTS AND RESULTS

4.1 Experimental setup

The HAF-Net was trained on a windows system with an Intel Core i5-12400F LGA 1700 CPU, 32 GB of RAM at 5600 MHz and an NVIDIA RTX 4070 GPU including 12 GB of VRAM and 5888 CUDA cores. The methods and algorithms were implemented in NumPy and TensorFlow. The network was trained with $150 \times 150 \times 3$ input image size, and a learning rate of 10^{-4} to moderate how fast the weights were adjusted in training. The batch size of 32 and 100 epochs were used to train the models, ensuring stable and efficient learning. Stratified 5-fold cross-validation was applied to the combined training and validation data to evaluate the robustness and stability of the proposed model. The experimental parameters are summarized in Table 1.

Table 1: Training hyperparameters.

Parameter	Value
Image Input	$150 \times 150 \times 3$
Epoch	100
Loss Function	Categorical cross entropy
Learning rate	10^{-4}
Batch size	32

4.2 Dataset

This study employed the retinal fundus images (RFI) dataset [27], a public Kaggle resource widely used in DR research [28]-[29]. The dataset contains 21,746 retinal images across 11 disease classes, including Moderate DR (5,288), Dry AMD (1,360), Wet AMD (587), Proliferative DR (1,416), Cataract (1,505), Normal Fundus (2,874), Hypertensive Retinopathy (1,344), Mild DR (2,438), Glaucoma (1,878), Pathological Myopia (1,265), and Severe DR (1,791). The dataset was split into 80% for training (60% training, 20% validation) and 20% for testing. Sample images are shown in Fig. 2, and class distribution is presented in Fig. 7.

4.3 Evaluation

The performance metric used to evaluate models for training is accuracy, as detailed in Equation (15), which represents the percentage of total data that the model is capable of correctly predicting. Conversely,

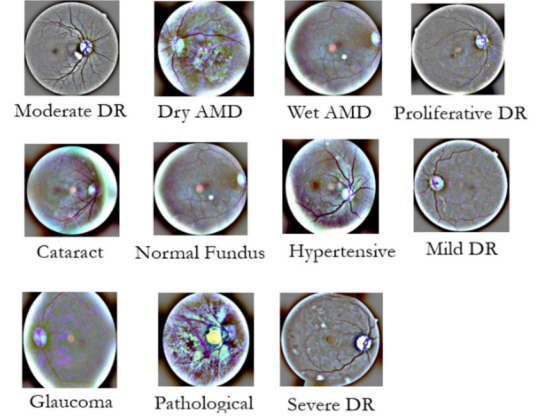


Fig.2: Sample retinal fundus images.

classification performance in terms of Precision, Recall, F1-score and AUC as described by Equations (16) to (19) and confusion matrix used standardized.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$F1 = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (18)$$

$$AUC = \int_0^1 TPR(FPR)d(FPR) \quad (19)$$

In DR classification, a True Negative (TN) occurs when a sample that does not belong to the target class is correctly identified as not belonging to that class. True Positive (TP) occurs when a sample belonging to the target class is correctly classified into its corresponding class. False Negative (FN) occurs when a sample belonging to the target class is incorrectly classified as another class. False Positive (FP) occurs when a sample from another class is incorrectly classified as the target class. Since diabetic retinopathy classification is formulated as a multi-class classification problem, TP, TN, FP, and FN are computed using a one-versus-rest strategy for each class, and the reported Precision, Recall, and F1-score are calculated using macro averaging across all classes.

4.4 Training results

In Fig. 3, the training accuracy of the three hybrid models over 100 epochs is presented. The hybrid model-BiLSTM achieved the highest training accuracy of 99.35%, followed by the hybrid model-MLP with 98.88%, and the hybrid model-LSTM with 98.63%. The results of this experiment indicate stable convergence throughout the training process, demon-

strating effective optimization and progressive learning of representative features from the DR dataset.

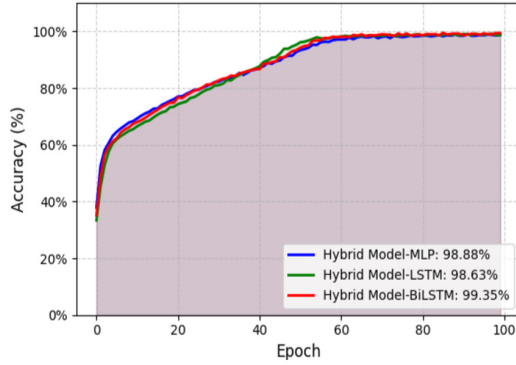


Fig.4: Training accuracy during the training process.

In Fig. 4, the loss values for each model across 100 epochs during training are presented. The hybrid model-BiLSTM achieved the lowest training loss of 0.0129, indicating higher learning efficacy than the hybrid model-MLP and hybrid model-LSTM, which

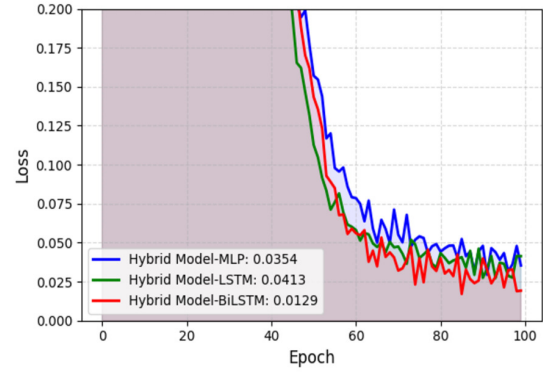
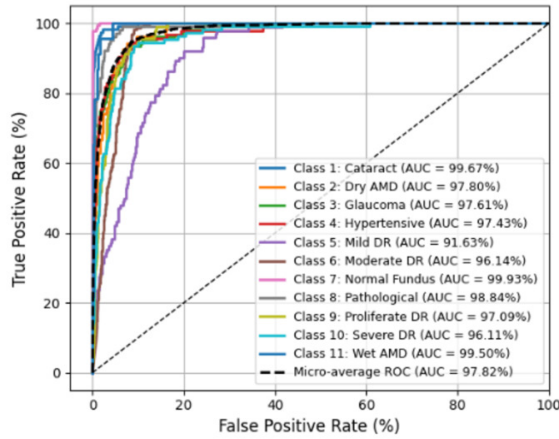
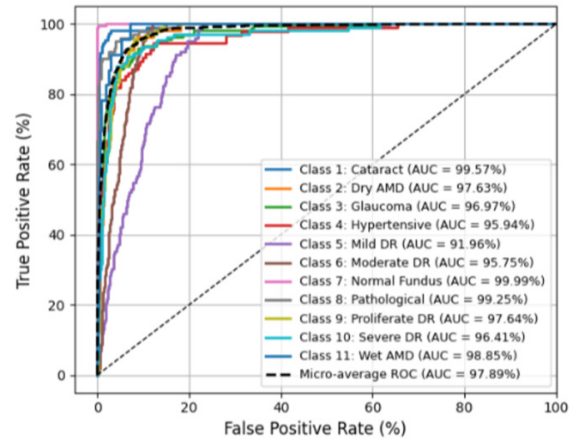


Fig.5: Training loss during the training process.

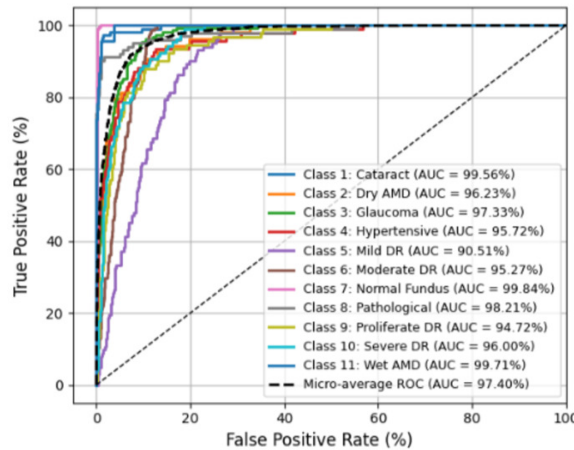
obtained training losses of 0.0354 and 0.0413, respectively. The loss for all models dropped to their respective minimum thresholds early in the training before stabilizing, indicating stable convergence throughout the training process. These results describe the optimization behavior during training only, while the validation and test performances are presented separately to assess the models generalization capability.



(a) Hybrid model-MLP



(b) Hybrid model-LSTM



(c) Hybrid model-BiLSTM

Fig.3: AUC curve for DR classification.

The validation and test results provide a more reliable evaluation of the models' generalization performance. The hybrid model-MLP achieved the highest validation accuracy of 76.41% and test accuracy of 80.17%, followed by the hybrid model-BiLSTM with 74.46% and 79.71%, respectively, and the hybrid model-LSTM with 74.42% and 79.51%. These results indicate that the MLP-based hybrid model provides the best generalization performance.

4.5 Classification performance

In Fig. 5, the classification task accuracies are presented, comprised of results from both hybrid model-MLP and hybrid model-LSTM-based models in addition to hybrid Model-BiLSTM, giving an evaluation score considering AUC across 11 disease classes. The rate ranged from approximately 91% to 99% for all three models, which suggests strong overall classification. But every model has its strengths and weaknesses in various classes. The average of the hybrid model-MLP was the highest with 97.82%, it reached

Normal Fundus of 99.93% and Cataract of 99.67%. However, the MLP had limited capability for distinguishing Mild DR with a score of 91.63%, suggesting that it was not versatile enough to detect subtle characteristics of the disease. The hybrid model-LSTM, like MLP, and with an average of 97.89%, worked as well (better in Pathological Myopia and Normal Fundus), a symptom likely characteristic of such a network for learning features related to chronic diseases. Additionally, hybrid model-BiLSTM having the lowest average of 97.40%. Nevertheless, the model performed the best in classifying complex diseases like Wet AMD compared to model others. These results support the use of AUC as a performance metric and highlight the appropriateness of each model according to target class complexity.

In Fig. 6, confusion matrices of the hybrid model-MLP, hybrid model-LSTM, and hybrid model-BiLSTM, respectively, on 11 DR disease classes illustrate the classification performances. In recognising Normal Fundus, the hybrid model-MLP achieved the

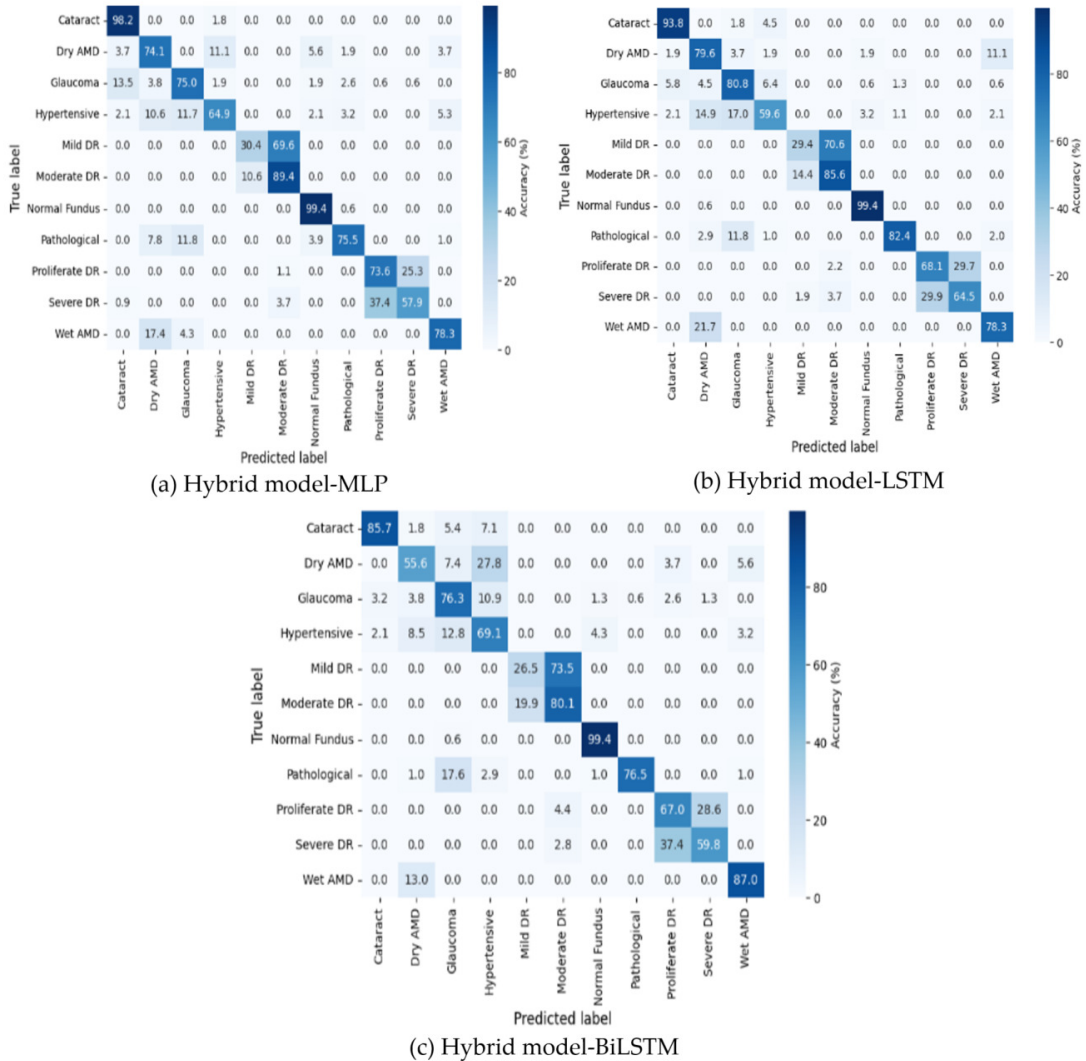


Fig.6: Confusion matrix for DR classification.

Table 2: *Classification performance.*

DR Disease	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
	hybrid model-MLP			hybrid model-LSTM			hybrid model-BiLSTM		
Cataract	80.88%	98.21%	88.71%	87.60%	94.64%	90.99%	93.20%	85.71%	89.30%
Dry AMD	58.82%	74.07%	65.57%	49.23%	59.26%	53.78%	60.00%	55.56%	57.69%
Glaucoma	82.98%	75.00%	78.79%	70.35%	77.56%	73.78%	74.38%	76.28%	75.32%
Hypertensive	84.72%	64.89%	73.49%	78.69%	51.06%	61.94%	60.19%	69.15%	64.36%
Mild DR	57.41%	30.39%	39.74%	42.19%	26.47%	32.53%	38.57%	26.47%	31.40%
Moderate DR	71.75%	89.35%	79.59%	69.77%	83.33%	75.95%	67.84%	80.09%	73.46%
Normal Fundus	93.68%	99.44%	96.48%	97.78%	98.32%	98.05%	96.22%	99.44%	97.80%
Pathological	89.53%	75.49%	81.91%	86.14%	85.29%	85.71%	98.73%	76.47%	86.19%
Proliferate DR	62.04%	73.63%	67.34%	66.29%	64.84%	65.56%	57.01%	67.03%	61.62%
Severe DR	72.09%	57.94%	64.25%	69.23%	67.29%	68.25%	69.57%	59.81%	64.32%
Wet AMD	69.23%	78.26%	73.47%	61.90%	56.52%	59.09%	74.07%	86.96%	80.00%

highest accuracy of 99.4%, followed by Cataract at 98.2% and Moderate DR at 89.4%. It showed some significant misclassification errors on the Glaucoma class, as well as Pathological, Mild DR, Wet AMD, and a few more. The hybrid model-LSTM and the hybrid model-BiLSTM also performed similarly to each other, showing confusion on classes that have similarity in visual appearance. The complexity of the data and the presence of overlapping features between some diseases might be the reason for the confusion, such that there is no clear class separation among different diseases, thus leading to an increase in classification errors.

In Table 2, the results of the hybrid model for 11 DR diseases are presented in terms of precision, recall, and f1-score. Hybrid model-MLP, it achieved impressive performance in classifying Cataract and Normal Fundus, however, it showed particularly ineffective results for Mild DR (Precision 57.41%, Recall 30.39%, F1-score 39.74%), suggesting its incapability of catching subtle disease characteristics. Hybrid model-LSTM was proved to be a superior model through-depth analysis under several cases Contract, Normal Fundus and Pathological, when compared with MLP model. However, recognizing the complex characteristic of disease was a significant challenge for the model in showing that it struggled to classify Mild DR, Severe DR, Glaucoma and Wet AMD. Furthermore, although the hybrid model-BiLSTM achieved a competitive F1-score in several primary classes such as Normal Fundus, Wet AMD, Cataract, and Pathological, the recall for multiple other classes was lower, thereby diminishing the overall classification performance. These results further confirm that each model has its strengths in identifying certain disease categories but none of them have a wide range of proficiency performance over all classes, especially for those cases with overlap features, or hard-to-observe symptom-progress diseases like Mild DR, Proliferate DR and Severe DR, respectively, which indicates a significant challenge exists while building diabetes retinopathy disease models that are comprehensive.

4.6 Performance comparison

This study compares the experimental results of the HAF-Net with those of previously published studies [30]-[33] to ensure consistency, focusing specifically on studies that utilize the RFI dataset. The comparison also includes model parameters and GFLOPs to evaluate both accuracy and efficiency. As the compared studies differ in experimental settings, the comparison is intended as an indicative reference rather than a direct benchmark. The performance comparison is shown in Table 3.

In Table 3, a comparison is shown of research on DR performance using various architectures such as SMOTE, Weinmed, VGG-19, MobileNetV2, ResNet, GoogleNet, EfficientNet-B5, Vision Transformer (ViT), and image processing techniques including CLAHE, U-Net, and BV segmentation, as well as the application of TL and FedCNN to address the problem of limited data and improve accuracy. The results from these studies show accuracy ranging from approximately 72.5 to 99%. The proposed method demonstrates performance comparable to previous studies while reducing model complexity with fewer parameters of 0.3731M–0.7016M and GFLOPs of 0.506, making it suitable for deployment in systems with limited computing resources. In addition, stratified 5-fold cross-validation achieved average accuracies of $80.28\% \pm 4.77\%$, $77.89\% \pm 4.13\%$, and $79.66\% \pm 4.48\%$ for the hybrid model-MLP, hybrid model-LSTM, and hybrid model-BiLSTM, respectively, indicating stable and reliable performance across different folds. The maximum training accuracy of 99.35% further demonstrates the model’s learning capability during training.

5. DISCUSSION

The HAF-Net effectively combines convolutional operations with dual attention mechanisms, CBAM and BAM to improve feature representation in DR classification. It shares conceptual similarities with dual-model or fusion-based approaches reported in prior studies [14], [20], [22], [34]-[35], but its distinctiveness lies in adopting a hybrid design that si-

Table 3: Performance comparison of baseline and proposed models.

Paper/Year	Research Focus:	Accuracy (Training)	Testing performance	Parameter (M)	GFLOPs
[10]/2025	- Integrating SMOTE, Weinmed, with VGG-19, and Federated Learning (FedCNN) for the DR classification.	98%	98.5%	-	-
[11]/2025	- Applies min-max normalization and CLAHE, integrates U-Net for red lesion segmentation, and employs CNN for DR classification.	-	99.84%	-	-
[12]/2025	- Pre-trained VGG-16 and MobileNet-V2 to the system achieved accuracy rates of 90% and 92%, respectively, for estimating the severity of DR from retinal images.	90%, 92%	-	-	-
[13]/2021	- A two-stage transfer learning approach with ROIs and pre-trained CNNs to enhance DR classification.	98.60%	99.7, 99.8%	-	-
[14]/2023	- An approach to detect DR using a CNN model, incorporating two different DL architectures, ResNet50 and InceptionV3, concatenated into the CNN for DR classification.	96.85%	90.24%	-	-
[15]/2023	- The MobileNetV2 method, based on transfer learning with fine-tuning to DR classification.	72.47%	69.6%	-	-
[16]/2020	- Combining the Contrast Limited Adaptive Histogram Equalization (CLAHE) method with the EfficientNet-B5 architecture to improve DR classification.	- 92,	94.5% (AUC)	-	-
[19]/2024	- The method combines a cascaded dense block (CDB) and 2X-FEM to create a new architecture, called CDB-2X-FEM, to improve DR.	87.22, 89.10, 92.22%	89.10, 87.22, 92.22%	3.16	2.785
[30]/2023	- A ViT pipeline using transfer learning to train a large model on a small dataset.	-	80.4%	307	-
[31]/2024	- Integrates GoogleNet and ResNet to develop a lesion-based DR detection model, enhancing both feature extraction for classification performance.	-	94%	-	-
[32]/2024	- Uses color-based BV and CLAH-TH segmentation with transfer-learned VGG19 and InceptionV3 to improve DR grading accuracy.	96.7%	98.1, 97.1%	21.07	-
[33]/2023	- A TL approach with optimized features for DR using pre-trained ShuffleNet and ResNet-18 network to enhance DR.	-	82, 96, 75%	-	-
Proposed	- A CNN with (32, 64, 128, 256) filters, a 3×3 kernel, and Dense layers (256, 128) is used to compare the performance of the proposed model.	99.27%	79.69%	1.276	2.259
	- A CBAM-CNN is used to compare the performance of the proposed model.	98.34%	77.99%	1.287	2.287
	- A BAM-CNN is used to compare the performance of the proposed model.	99.35%	78.72%	1.296	2.281
	- A hybrid model (HAF-Net) integrated MLP, LSTM, and BiLSTM to enhance diabetic retinopathy classification.	98.88, 99.35, 98.63%	80.17, 79.51, 79.71%	0.3731 0.5031 0.7016	0.506

Note: “-” denotes that the parameters and GFLOPs were not reported in the referenced study.

multaneously integrates CBAM and BAM, enabling adaptive manipulation of image features for enhanced discriminative power. By employing three convolutional blocks with increasing filter sizes of 32, 64, and 128, coupled with ReLU activation and 2×2 max-pooling, the model can extract robust low-to-high-level features while maintaining spatial resolution through SAME padding. Furthermore, the inclusion of CBAM enables the model to flexibly emphasize the most relevant channels and spatial locations via a sequential attention strategy. Specifically, channel attention learns the importance of each channel using global pooling and MLP operations, while spatial attention highlights informative regions through convolutional processing. BAM enhances this performance by introducing a parallel attention structure with extended dilation, which expands the receptive

field and enables the model to capture long-range dependencies and subtle differences in lesion patterns, essential for early DR classification.

The validation and test results provide a more reliable assessment of the models generalization capability. The hybrid model-MLP achieved the highest validation accuracy of 76.41% and test accuracy of 80.17%, indicating the best generalization performance. During training, the proposed HAF-Net achieved the highest training accuracy of 99.35% with the lowest training loss of 0.0129, indicating stable optimization. Furthermore, when considering the classification results of all models, as shown in Fig. 5, the model demonstrates high average AUC values across each class of 99.67, 97.80, 97.61, 97.43, 91.96, 96.14, 99.99, 99.25, 97.64, 96.41, and 99.71% for Cataract, Dry AMD, Glaucoma, Hypertensive, Mild DR, Mod-

erate DR, Normal Fundus, Pathological, Proliferative DR, Severe DR and Wet AMD, respectively. These results indicate that the proposed HAF-Net effectively distinguishes most retinal disease classes. In particular, the model performed well in Cataract, Glaucoma, Moderate DR, Normal Fundus, Pathological and Wet AMD. However, the lower AUC of Mild DR at 91.96% and the lower Precision, Recall, and F1-score for some classes indicate that early-stage DR remains challenging to classify. Further improvements and validation on larger clinical datasets are therefore required before broader clinical application.

6. CONCLUSION

This study proposes a HAF-Net that integrates the convolutional operators and CBAM with BAM for efficient DR feature extraction, coupled with residual connections and a hybrid fusion of MLP, LSTM, and BiLSTM to improve feature learning and DR classification. The model is evaluated on the RFI dataset comprising 11 disease classes and demonstrates high accuracy, precision, recall, F1-score, and AUC, achieving performance comparable to or exceeding existing methods. These results indicate that the proposed HAF-Net is a promising research model for automated DR classification, although further clinical validation is required.

The validation and test results confirmed the model's generalization performance, with the hybrid model-MLP achieving the highest validation accuracy of 76.41% and test accuracy of 80.17%. During training, the proposed HAF-Net achieved a training accuracy of 99.35% with a training loss of 0.0129, indicating stable optimization. The model also showed high performance in classification tasks with AUC value of 99.67, 97.80, 97.61, 97.43, 91.96, 96.14, 99.99, 99.25, 97.64, 96.41, and 99.71% for Cataract, Dry AMD, Glaucoma, Hypertensive, Mild DR, Moderate DR, Normal Fundus, Pathological, Proliferative DR, Severe DR, and Wet AMD, respectively, indicating that the proposed HAF-Net is effective in distinguishing retinal diseases, although the performance varies across disease classes.

Future work will focus on developing a computationally efficient HAF-Net for deployment on mobile and edge devices. Additionally, we aim to further optimize the network architecture and validate the model on larger and more diverse clinical datasets to improve the classification of challenging disease classes.

AUTHOR CONTRIBUTIONS

Conceptualization, A.P. and S.T.; methodology, A.P. and S.T.; software, A.P. and S.T.; validation, S.T., P.K. and A.P.; formal analysis, S.T., P.K. and A.P.; investigation, S.T. and A.P.; data curation, A.P. and S.T.; writing—original draft preparation,

A.P. and S.T.; writing—review and editing, S.T., P.K. and A.P.; visualization, A.P. and S.T.; supervision, A.P.; All authors have read and agreed to the published version of the manuscript.

References

- [1] M. Duggal *et al.*, “Real-world evaluation of AI-driven diabetic retinopathy screening in public health settings: Validation and implementation study,” *JMIR Medical Informatics*, vol. 13, 2025.
- [2] K. Lee, G. H. Lee, S. E. Lee, J. M. Yang and K. Bae, “Glycemic control and retinal microvascular changes in type 2 diabetes mellitus patients without clinical retinopathy,” *Diabetes & Metabolism Journal*, vol. 48, no. 5, pp. 983–992, 2024.
- [3] Y. Zhang *et al.*, “Serum disease-specific IgG Fc glycosylation as potential biomarkers for nonproliferative and proliferative diabetic retinopathy using mass spectrometry,” *Molecular & Cellular Proteomics*, no. 109555, 2023.
- [4] X. Ding *et al.*, “Expanded field OCT angiography biomarkers for predicting clinically significant outcomes in non-proliferative diabetic retinopathy,” *American Journal of Ophthalmology*, vol. 270, pp. 216–226, 2025.
- [5] D. S. Pushparani, J. Varalakshmi, K. Roobini, P. Hamshapriya and A. Livitha, “Diabetic retinopathy—A review,” *Current Diabetes Reviews*, vol. 21, no. 7, pp. 43–55, 2025.
- [6] P. Ansari, N. Tabasumma, N. N. Snigdha, N. H. Siam, R. V. N. R. S. Panduru, S. Azam, J. M. A. Hannan and Y. H. A. Abdel-Wahab, “Diabetic retinopathy: An overview on mechanisms, pathophysiology and pharmacotherapy,” *Diabetology*, vol. 3, no. 1, pp. 159–175, 2022.
- [7] C. Sen, S. Doganay, and G. Ozcan, “New accurate deep learning model for diabetic retinopathy detection utilizing sequential pre-processing and transfer learning,” *Biomed. Signal Process. Control*, vol. 109, no. 108060, 2025.
- [8] P. Hariobulesu and F. Shaik, “Enhanced multi-grade diabetic retinopathy detection and classification via ensemble deep learning model from retinal fundus images,” *Expert Systems with Applications*, vol. 285, no. 128116, 2025.
- [9] E.-M. Chakour, Z. Sadok, R. Kachouri, A. Mansouri, I. B. Andaloussi, M. Akil and A. Ahaitouf, “Mobile-based deep learning system for early detection of diabetic retinopathy,” *Intelligence-Based Medicine*, vol. 12, no. 100259, 2025.
- [10] S. Alanazi and R. Alanazi, “Enhancing diabetic retinopathy detection through federated convolutional neural networks: Exploring different stages of progression,” *Alexandria Engineering Journal*, vol. 120, pp. 215–228, 2025.
- [11] G. M. Ramadan, M. M. Adnan, Z. L. Naser, B. V. Kumar and K. Parthiban, “Diabetic

- Retinopathy Segmentation and Classification Using U-Net with Convolutional Neural Network,” *2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, Tumkur, India, pp. 1-5, 2023.
- [12] A. M. A and S. S. S. Priya, “Detection and Classification of Diabetic Retinopathy Using Pre-trained Deep Neural Networks,” *2023 International Conference on Innovations in Engineering and Technology (ICIET)*, Muvattupuzha, India, pp. 1-7, 2023.
- [13] F. Saeed, M. Hussain and H. A. Aboalsamh, “Automatic diabetic retinopathy diagnosis using adaptive fine-tuned convolutional neural network,” in *IEEE Access*, vol. 9, pp. 41344–41359, 2021.
- [14] G. Ali, A. Dastgir, M. W. Iqbal, M. Anwar and M. Faheem, “A Hybrid Convolutional Neural Network Model for Automatic Diabetic Retinopathy Classification From Fundus Images,” in *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 11, pp. 341-350, 2023.
- [15] S. Y. Prasetyo, A. Wulandari, Z. N. Izdihar, G. Z. Nabillah and H. A. Saputri, “Comparing convolutional neural network models with MobileNet-based in diabetic retinopathy classification,” in *Proceedings International Workshop on Artificial Intelligence and Image Processing (IWAIIIP)*, Yogyakarta, Indonesia, pp. 1–5, 2023.
- [16] A. M. Pour, H. Seyedarabi, S. H. A. Jahromi and A. Javadzadeh, “Automatic detection and monitoring of diabetic retinopathy using efficient convolutional neural networks and contrast limited adaptive histogram equalization,” in *IEEE Access*, vol. 8, pp. 136668–136673, 2020.
- [17] X. Luo *et al.*, “A Lesion-Fusion Neural Network for Multi-View Diabetic Retinopathy Grading,” in *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 5, pp. 3184-3193, May 2025.
- [18] Y. Niu, L. Gu, Y. Zhao and F. Lu, “Explainable Diabetic Retinopathy Detection and Retinal Image Generation,” in *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 44-55, Jan. 2022.
- [19] S. Madarapu, S. Ari and K. Mahapatra, “C2x-FNet: Cascaded Dense Block With Twofold Cross-Feature Enhancement Module for Diabetic Retinopathy Grading,” in *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1-10, 2025.
- [20] H. Mustafa, S. F. Ali, M. Bilal and M. S. Hanif, “Multi-stream deep neural network for diabetic retinopathy severity classification under a boosting framework,” in *IEEE Access*, vol. 10, pp. 113172–113183, 2022.
- [21] R. Rashid, W. Aslam, A. Mehmood, D. L. R. Vargas, I. D. L. T. Diez and I. Ashraf, “A Detectability Analysis of Retinitis Pigmentosa Using Novel SE-ResNet Based Deep Learning Model and Color Fundus Images,” in *IEEE Access*, vol. 12, pp. 28297-28309, 2024.
- [22] M. M. Farag, M. Fouad and A. T. Abdel-Hamid, “Automatic Severity Classification of Diabetic Retinopathy Based on DenseNet and Convolutional Block Attention Module,” in *IEEE Access*, vol. 10, pp. 38299-38308, 2022.
- [23] M. U. Yaseen, M. M. Nasralla, F. Aslam, S. S. Ali and S. B. A. Khattak, “A Novel Approach Based on Multi-Level Bottleneck Attention Modules Using Self-Guided Dropblock for Person Re-Identification,” in *IEEE Access*, vol. 10, pp. 123160-123176, 2022.
- [24] K. Bharatheedasan, T. Maity, L. A. Kumaraswamidhas and M. Durairaj, “Enhanced fault diagnosis and remaining useful life prediction of rolling bearings using a hybrid multilayer perceptron and LSTM network model,” *Alexandria Engineering Journal*, vol. 115, pp. 355–369, 2025.
- [25] A. R. Zangene, O. Williams Samuel, A. Abbasi, K. Nazarpour, A. A. McEwan and G. Li, “An Attention-based Bidirectional LSTM Model for Continuous Cross-Subject Estimation of Knee Joint Angle during Running from sEMG Signals,” *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Sydney, Australia, pp. 1-4, 2023.
- [26] J. Huang, M. Lin, J. Li and G. Liu, “Based on MLP and BiLSTM battery capacity forecast method research,” *2024 4th International Conference on Intelligent Power and Systems (ICIPS)*, Yichang, China, pp. 407-411, 2024.
- [27] K. S. Sanjay Nithish, “Retinal fundus images,” Kaggle, 2023. [Online]. Available: <https://www.kaggle.com/datasets/kssanjaynithish03/retinal-fundus-images>. [Accessed: Jul. 2, 2025].
- [28] C. Wan, J. Zhao, X. Hong, W. Yang, and S. Zhang, “HCSP-Net: A novel model of age-related macular degeneration classification based on color fundus photography,” *Computers, Materials and Continua*, vol. 79, no. 1, pp. 391–407, 2024.
- [29] M. M. Hemal and S. Saha, “Explainable deep learning-based meta-classifier approach for multi-label classification of retinal diseases,” *Array*, vol. 26, Art. no. 199402, 2025.
- [30] W. Nazih, A. O. Aseeri, O. Y. Atallah, and S. El-Sappagh, “Vision transformer model for predicting the severity of diabetic retinopathy in fundus photography-based retina images,” in *IEEE Access*, vol. 11, pp. 117546–117561, 2023.
- [31] A. Jabbar, H. B. Liaqat, A. Akram, M. U. Sana,

- I. D. Azpíroz, and I. De La Torre Diez, “A lesion-based diabetic retinopathy detection through hybrid deep learning model,” in *IEEE Access*, vol. 12, pp. 40019–40036, 2024.
- [32] A. J. Kadhim, H. Seyedarabi, R. Afrouzian, and F. S. Hasan, “Diabetic retinopathy classification using hybrid color-based CLAHE and blood vessel in deep convolution neural network,” in *IEEE Access*, vol. 12, pp. 194750–194761, 2024.
- [33] W. K. Wong, F. H. Juwono, and C. Apri-ono, “Diabetic retinopathy detection and grading: A transfer learning approach using simultaneous parameter optimization and feature-weighted ECOC ensemble,” in *IEEE Access*, vol. 11, pp. 83004–83016, 2023.
- [34] A. Raza, S. Bin Musa, A. S. Bin Khalid, M. M. Alam, M. M. Su’ud, and F. Noor, “Enhancing medical image classification through PSO-optimized dual deterministic approach and robust transfer learning,” in *IEEE Access*, vol. 12, pp. 177144–177159, 2024.
- [35] M. Shafiq, Q. Fan, F. H. Alghamedy, and W. J. Obidallah, “DualEye-FeatureNet: A dual-stream feature transfer framework for multi-modal ophthalmic image classification,” in *IEEE Access*, vol. 12, pp. 143985–144008, 2024.



Pauline Kongsuwan received the B.Eng. degree in Computer Engineering from Kasetsart University, Thailand, in 2004. She received the M.Eng. degree and the D.Eng. degree in Information Systems from Inje University, South Korea, in 2009 and 2015, respectively. She is currently an Assistant Professor with Faculty of Engineering, Rajamangala University of Technology Thanyaburi (RMUTT), Pathum Thani, Thailand.

Her current research focuses on software engineering, information systems, optimization algorithms, and artificial intelligence. She can be contacted at pauline.k@en.rmUTT.ac.th



Sirichai Triamlumlerd received the B.Eng. degree in Computer Engineering from Rajamangala University of Technology Thanyaburi (RMUTT), Thailand, in 1998, the M.Eng. degree in Computer Engineering from King Monkut’s University of Technology Thonburi (KMUTT), Thailand, and the Ph.D. degree from the University of Northumbria, Newcastle, U.K., in 2013. He is currently a lecturer and researcher

with RMUTT. His current research interests include image processing and artificial intelligence. He can be contacted at sirichai.t@en.rmUTT.ac.th



Anuruk Prommakhot received a B.Eng. degree in Computer Engineering from Rajamangala University of Technology Isan Sakonnakhon Campus, in 2015, and the M.Eng. degrees in Electronic and Communication Engineering from Rajamangala University of Technology Thanyaburi (RMUTT), in 2018. He is currently pursuing the Ph.D. degree in engineering from RMUTT, Since 2019. respectively. His research inter-

ests deep learning, optimization algorithms, machine learning, vision AI, RAG system, and agentic AI. He can be contacted at anuruk.p@en.rmUTT.ac.th