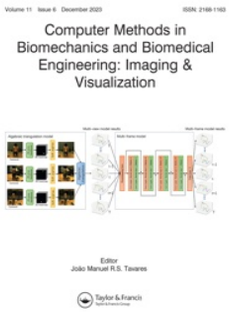


Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization

ISSN: 2168-1163 (Print) 2168-1171 (Online) Journal homepage: www.tandfonline.com/journals/tciv20

Enhancing diabetic retinopathy classification accuracy through dual-attention mechanism in deep learning

Abdul Hannan , Zahid Mahmood , Rizwan Qureshi & Hazrat Ali

To cite this article: Abdul Hannan , Zahid Mahmood , Rizwan Qureshi & Hazrat Ali (2025) Enhancing diabetic retinopathy classification accuracy through dual-attention mechanism in deep learning, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 13:1, 2539079, DOI: 10.1080/21681163.2025.2539079

To link to this article: https://doi.org/10.1080/21681163.2025.2539079

Published online: 28 Jul 2025.

COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING: IMAGING & VISUALIZATION
2025, VOL. 13, NO. 1, 2539079
https://doi.org/10.1080/21681163.2025.2539079

Enhancing diabetic retinopathy classification accuracy through dual-attention mechanism in deep learning

Abdul Hannan[a], Zahid Mahmood[a], Rizwan Qureshi[b] and Hazrat Ali[c]

[a]Department of Electrical and Computer Engineering, COMSATS, University Islamabad, Abbottabad, Pakistan; [b]Center for Research in Computer Vision, The University of Central, Orlando, FL, USA; [c]Division of Computing Science and Mathematics, University of Stirling, Stirling, UK

ABSTRACT

Automatic classification of Diabetic Retinopathy (DR) can assist ophthalmologists in devising personalised treatment. However, imbalanced data distribution in the dataset becomes a bottleneck in the generalisation of deep learning models trained for DR classification. In this work, we combine global attention block (GAB) and category attention block (CAB) into the deep learning model, thus effectively overcoming the imbalanced data distribution problem in DR classification. Our proposed approach is based on an attention-based deep learning model employing three pre-trained networks, namely, MobileNetV3-small, Efficientnet-b0, and DenseNet-169 as the backbone architecture. We evaluate the proposed method on two publicly available datasets of retinal fundoscopy images for DR. Experimental results show that on the APTOS dataset, the DenseNet-169 yielded 83.20% mean accuracy, followed by MobileNetV3-small and EfficientNet-b0, which yielded 82% and 80% accuracies, respectively. On the EYEPACS dataset, the EfficientNet-b0 yielded a mean accuracy of 80%, while the DenseNet-169 and MobileNetV3-small yielded 75.43% and 76.68% accuracies, respectively. In addition, we also compute an F1-score of 82.0%, a precision of 82.1%, a sensitivity of 83.0%, a specificity of 95.5%, and a kappa score of 88.2% for the experiments. The proposed approach achieves competitive performance that is at par with recently reported works on DR classification.

ARTICLE HISTORY
Received 8 April 2025
Accepted 13 July 2025

KEYWORDS
Attention mechanism; deep learning; diabetic retinopathy; image classification; medical imaging

1. Introduction

Diabetes Mellitus (DM) is a long-lasting sickness in which blood sugar levels increase due to the inability of the pancreas to secrete sufficient blood insulin Atwany et al. (2022); Chen et al. (2023). The International Diabetes Federation (IDF) reports that over 537 million individuals between the ages of 20 and 79 worldwide have DM Federation (2021). Over the past two decades, the prevalence of diabetes among adults has increased over threefold, and approximately 230 million people with diabetes, which is half of all cases, remain undiagnosed. An alarming estimate is that the global DM population is expected to increase to 643 million by the year 2030 and further to 783 million by 2045 Federation (2021).

Diabetic retinopathy (DR) is a complication of diabetes that affects small blood vessels in the retina, causes blockage and bleeding in the retina, and ultimately leads to vision loss. Moreover, DR can also affect human organs like the liver, heart, kidneys, and eyes. Therefore, early detection and treatment of the DR are crucial to prevent permanent damage. The primary abnormal characteristics observed in the DR consist of (i) microaneurysms, which is an early stage of the DR in which tiny red dots appear on the retina (Yu et al. (2022)). These dots are defined by sharp margins with a size of at most 125 μm. (ii) Hemorrhages, which are indicated by large spots on the retina surface with irregular margin sizes above 125 μmm. (iii) Hard exudates, which appear as a result of plasma leakage, become visible as yellow spots on the retina surface and span the outer retina layers with sharp margins. (iv) Soft exudates, which appear as a result of swelling of nerve fibers, have an oval-like appearance on the retina surface. Based on the nature and quantity of abnormalities observed in fundus images, the DR can be categorised into five distinct phases, which include the absence of (DR0), minor (DR1), moderate (DR2), severe (DR3), and proliferative (DR4) Federation (2021).

CONTACT Hazrat Ali hazrat.ali@live.com Division of Computing Science and Mathematics, University of Stirling, Stirling FK94LA, UK

Deep learning methods have shown great promise in automatically classifying the distinct phases of diabetic retinopathy (DR), particularly through tailored CNN architectures (Yang et al. (2024)). In recent years, Convolutional Neural Networks (CNNs) have made significant strides and found extensive applications in computer vision tasks such as image classification, object detection, and semantic segmentation Li et al. (2020); He et al. (2016); Ren et al. (2015); He et al. (2017); Iqbal et al. (2021). These networks seamlessly combine feature extraction and classification into a unified process, leading to remarkable advancements in these domains. Moreover, CNNs have seen extensive utilisation in the domain of retinal image analysis, thus playing an important role in tasks such as blood vessel segmentation Yang et al. (2024), identifying optic disc boundaries, and glaucoma screening Fu et al. (2018), Raghavendra et al. (2018); Fu et al. (2018). While numerous deep learning methods have been introduced for diabetic retinopathy classification on balanced datasets, however, they often struggle to address the challenge of imbalanced datasets. Improving accuracy on imbalanced datasets poses a significant challenge for algorithms that demonstrate impressive accuracy on a balanced dataset and are over-parameterised.

Recently, a two-stage CNN architecture to detect lesions and grade DR in fundus images has shown promising results (Yang et al. (2017)). This method addresses the DR grading process through a two-stage framework; however, the pipeline entails greater intricacy than a single-stage strategy. While CNN-based methods for grading DR have demonstrated favourable outcomes, their implementation in clinical practice remains challenging due to the inherent complexity of the task. One major difficulty arises from the striking similarities in colour and texture among the five DR grades, making it prone to confusion during the grading process. This significantly hampers the performance of the model in making a distinction between the different classes. Furthermore, certain lesions within fundus images are minute, comprising only a few pixels. Hence, there is a need to explore newer methods that can achieve the task of DR classification and also effectively address the challenge of an unbalanced dataset. In this connection, we introduce an attention-based mechanism that uses pre-trained models as a backbone to perform the tasks of DR classification in retinal fundus images. Our main contributions are listed below.

- We introduce a CNN-based approach while reducing the number of parameters, which automatically classifies the DR. The proposed method is trained and tested on two well-known publicly available APTOS and EYEPACS datasets. We focus on two crucial challenges. Initially, we tackle the challenge of imbalanced datasets, ensuring that appropriate accuracy is achieved. Later, we optimise the number of parameters utilised by the CNN algorithms, thereby improving their efficiency without compromising accuracy.
- Our proposed DR classification approach incorporates the Global Attention Block (GAB) and Category Attention Block (CAB) into backbone networks, such as MobileNetV3-small, DenseNet-169, and EfficientNet-b0. The proposed approach exhibits three key characteristics for GAB and CAB. Firstly, GAB and CAB are distinct attention blocks with different functionalities. GAB encompasses channel attention and spatial attention, while CAB emphasises category attention. By combining these two blocks, we enhance the performance of DR classification. Secondly, our model employs the CAB to produce class-specific attention feature maps in a category-oriented manner. Consequently, the model effectively captures more complex GAB and CAB features. This approach proves highly advantageous when dealing with imbalanced datasets in the context of DR grading tasks.
- Our model employs the GAB (a global attention block) within a single-branch architecture specifically designed for Diabetic Retinopathy (DR) grading. GAB captures rich global contextual features, which are subsequently refined by a Category Attention Block (CAB) that emphasises discriminative, category-specific information. Together, these modules form a dual attention mechanism that processes spatial and channel features in parallel – unlike conventional sequential models such as CBAM and SE-ResNet. This parallel design preserves essential feature interactions and adapts dynamically to varying image contexts through a learnable unification strategy. Additionally, the proposed architecture effectively mitigates class imbalance and achieves competitive accuracy with significantly fewer parameters, as demonstrated in Section 4.6, Table 3.

The manuscript is organised as follows: Section 2 briefly reviews the recent literature on DR classification. Section 3 presents our proposed methodology. Section 4 presents detailed results and discussions. Finally, conclusion is provided in Section 5. To facilitate the readers, Table 1 shows the key acronyms used in the text.

Table 1. Key abbreviations used in the text.

Acronym	Meaning
CAB	Category Attention Block
CNN	Convolutional Neural Network
DNN	Deep Neural Network
GAB	Global Attention Block
FC	Fully Connected
DR	Diabetic Retinopathy
GAP	Globel Average Pooling
EyePACS	Eye Picture Archive Communication System
APTOS	Asia Pacific Tele-Ophthalmology Society
NPDR	Non-Proliferative DR
PDR	Proliferative DR
Acc	Accuracy
MLP	Multilayer Perception
CBAM	Convolutional Block Attention Module
BNN	Bayesian Neural Network
PCA	Principal Component Analysis

2. Related work

Earlier works on DR classification were mostly based on hand-crafted features [Reddy et al. (2020); Bhatia et al. (2016). These approaches typically analysed anatomical features such as blood vessels and the optic disc to identify abnormalities such as microaneurysms, haemorrhages, and exudates. However, these methods were labour-intensive and time-consuming. With the rise of Machine Learning (ML) and Deep Learning (DL), there has been a shift towards automatic feature extraction and classification. Several Jiang et al. (2019); Qomariah et al. (2019); Qummar et al. (2019) have leveraged deep convolutional neural networks (DCNNs) for DR detection and classification. Similarly, methods proposed in Yan et al. (2018); Wu et al. (2020) focused on segmenting retinal vessels, while others (Zhao et al. (2018); Iqbal and Ali (2018); Ahmad et al. (2022)) used generative adversarial networks (GANs) to synthesise realistic retinal images for improved training.

Dai et al. (2018) introduced a multi-sieving CNN with an 'image-to-text mapping' mechanism to locate microaneurysms in retinal images in real-time. Jain et al. (Jain et al. (2019)) assessed the effectiveness of transfer learning using VGG19, VGG16, and Inception-v3 for both binary and multi-class classification. Zeng et al. (Zeng et al. (2019)) designed a Siamese CNN architecture with transfer learning to classify fundus images into two categories. The study in Kassani et al. (2019) utilised a multilayer perceptron (MLP) alongside an enhanced Xception network, which fused multiple feature maps for better accuracy. Another approach by Gao et al. (2019) developed four separate Inception models, each responsible for grading a quadrant of the fundus image, enabling more granular classification.

Multi-model frameworks have also been explored to improve classification performance. For instance, Yan et al. (2018) used a two-stage attention mechanism with ResNet50 for detecting DR and diabetic macular oedema (DME), training on both Messidor and IDRID datasets. A three-stage approach was proposed in Elswah et al. (2020), involving preprocessing, feature extraction using ResNet50, and classification using support vector machines (SVM) and neural networks on the ISBI'2018 IDRID dataset Balagurunathan et al. (2021) Similarly, Mustafa et al. (2022a) presented a multi-stream deep neural network combining ResNet50, DenseNet121, PCA for dimensionality reduction, and ensemble classification using boosting techniques.

It is well established that increasing CNN depth without sufficient training data can lead to overfitting Zhou et al. (2019); Marmanis et al. (2016). To address this, Farag et al. (2022) proposed a DenseNet-169 model enhanced with a Convolutional Block Attention Module (CBAM), where features were extracted, refined, and averaged using global average pooling before classification. In Tan et al. (2017), a two-stage detector was used to identify specific DR lesions such as haemorrhages and exudates, while Guo et al. Guo (2022) focused on segmentation-based classification.

A comparative study by Dutta et al. (2018) evaluated CNN, DNN, and BNN models using a dataset of 2000 retinal images for 5-stage DR classification. The DNN model achieved the highest accuracy of 86.3%. Another work by Esfahani et al. (2018) employed ResNet-34 with preprocessing techniques on the Kaggle dataset, achieving 85% accuracy. Haq et al. (2024) developed a hybrid YOLOv2 and ResNet-based framework

achieving over 96% accuracy in detecting and counting colorectal cancer cells, demonstrating the strength of integrated localisation – classification pipelines in biomedical image analysis.

Recent advancements have focused on secure and collaborative learning strategies. Bhulakshmi and Rajput proposed the FedDEO and FedDL models Bhulakshmi and Rajput (2024b, 2024a), leveraging federated learning to ensure privacy-aware, distributed DR classification. These approaches align with our current study's objective to develop generalised, interpretable, and privacy-preserving DR classification models.

3. Proposed method

In this section, we present our proposed architectural design, training configurations, and the evaluation criteria employed in our study. We also explain the exploration of data augmentation, balancing techniques, and subsequent detailed analysis of our approach. Figure 1 shows the overall workflow of our approach.

3.1. Data preprocessing

To enhance the accuracy of the DR classification, we have augmented the dataset using various techniques. Data preprocessing is done earlier than data manipulation to fit the data for the network that is employed in later stages. We performed the data preprocessing in the following manner.

Image Rescaling: The APTOS and EYEPACS datasets contain images of different sizes. Therefore, we rescale each image to a resolution of 512×512 pixels.

Image Augmentation: Figure 2 shows the outcomes for the data augmentation techniques. To enhance the robustness and generalisation capability of the model, we applied data augmentation techniques to both the APTOS and EyePACS datasets. Specifically, the images were augmented using rotation (90°, 180°, and 270°) and horizontal flipping. These transformations increased the diversity of the training data and approximately doubled the dataset size by generating new, distinct variations of the original images. This

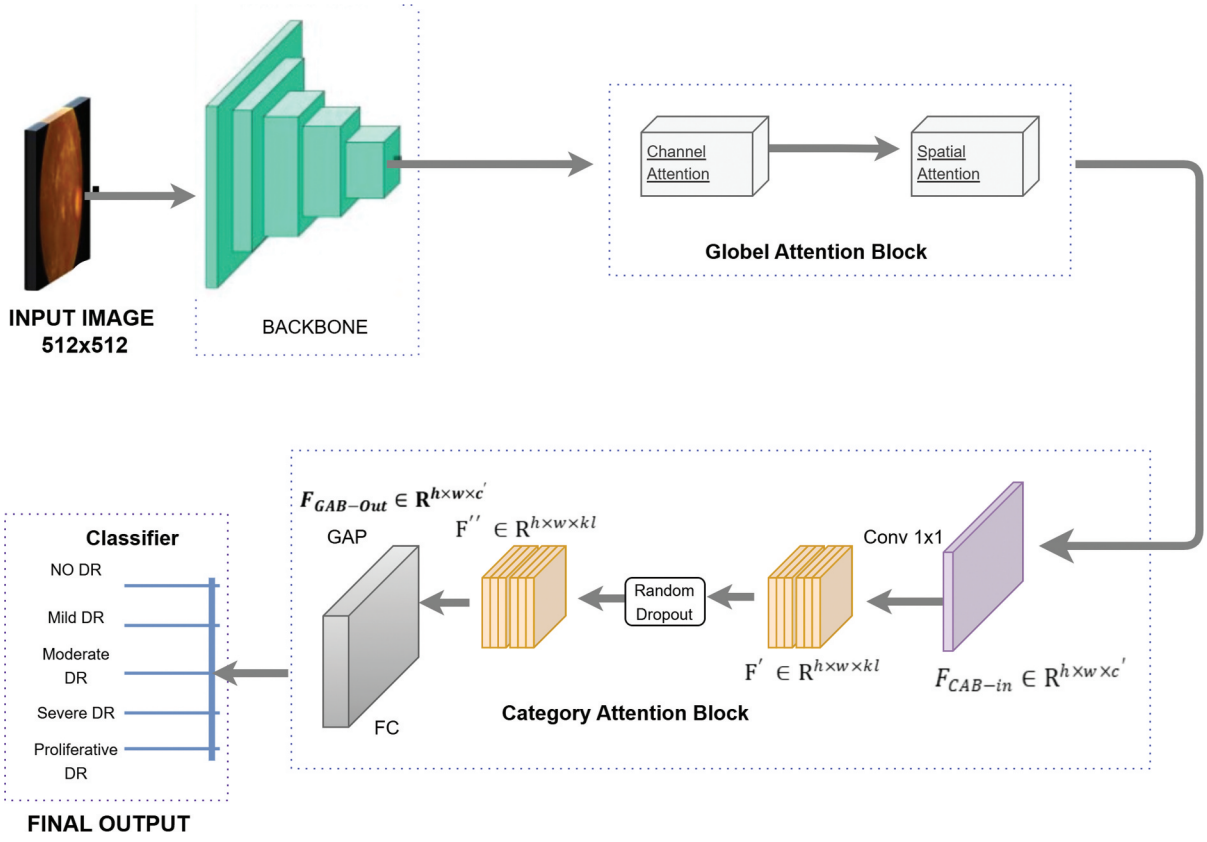


Figure 1. Overall workflow of the proposed approach for diabetic retinopathy classification.

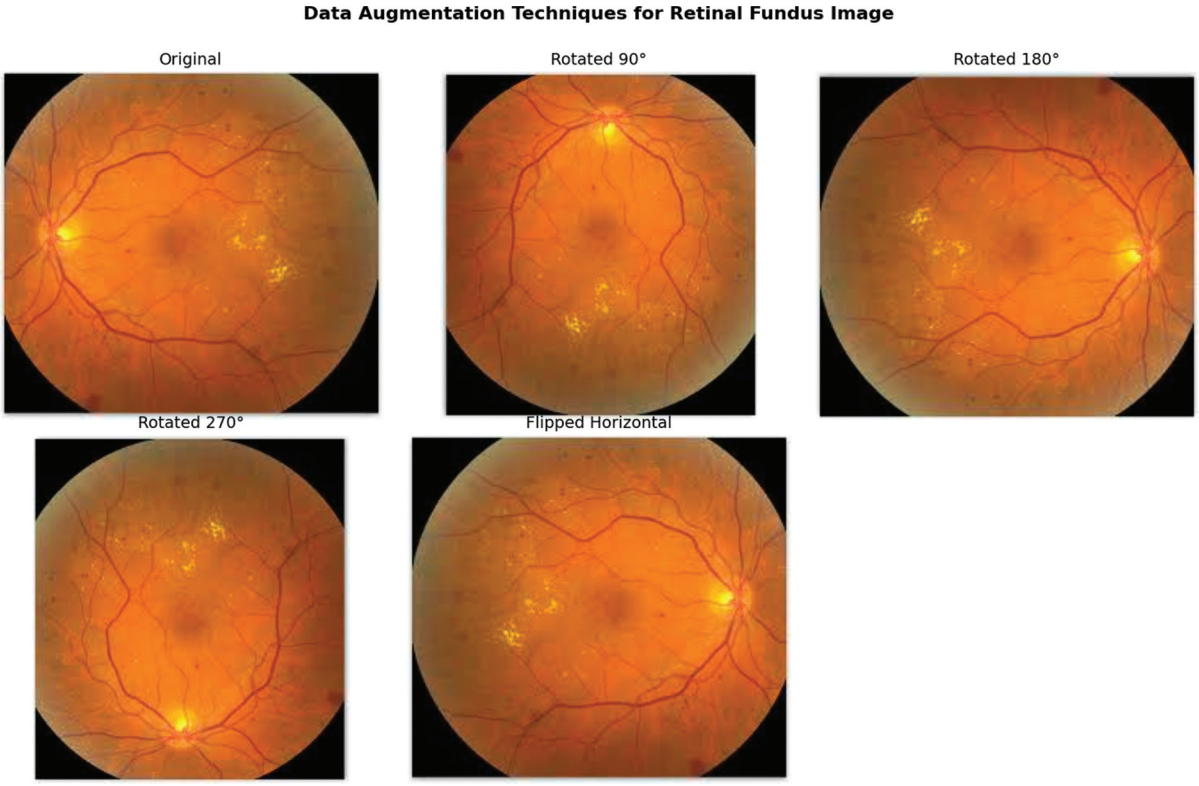


Figure 2. Augmented image: this figure illustrates the data augmentation techniques applied to retinal fundus images, including rotations (90°, 180°, 270°) and horizontal flipping. These augmentations enhance dataset diversity and improve model generalisation.

augmentation strategy helps reduce overfitting and improves the model's performance in recognising retinal abnormalities under different orientations.

Image Labelling: With the assistance of professional ophthalmologists, we manually label the dataset as DR0, DR1, DR2, DR3 and DR4 for DR categorised into five distinct phases. These labels refer to the absence of DR (DR0), minor DR (DR1), moderate DR (DR2), severe DR (DR3), and proliferative DR (DR4).

3.2. Architecture

As indicated in Figure 1, our proposed pipeline incorporates GAB and CAB into the three backbone networks, namely MobileNetV3-small, DenseNet-169, and EfficientNet-b0. It exhibits three key characteristics for the GAB and CAB. Firstly, the GAB and CAB are distinct attention blocks with different functionalities. GAB encompasses channel attention and spatial attention, while CAB emphasises category attention. Integrating these two blocks, we can enhance the performance of the DR classification algorithm. Secondly, our model employs the CAB to produce class-specific attention feature maps in a category-oriented manner. Consequently, the model can effectively capture more complex GAB and CAB features. We will analyse in the next section that this approach proves highly advantageous when dealing with imbalanced datasets in the context of DR grading tasks.

Thirdly, our model incorporates a single-branch GAB, which is specifically tailored for the DR classification task. The CAB is subsequently applied utilising the attention feature maps, which are generated by the GAB. By leveraging the enhanced global attention feature maps that are independent of specific categories, the CAB contributes significantly to the improvement of the DR grading. Below, we briefly describe the GAB and the CAB modules.

GAB: The GAB module contains channel attention and spatial attention blocks as indicated in Figure 3. GAB takes the features $F_{reduce} \in R^{h \times w \times c'}$ as input and learns global attention feature maps. Initially, we compute the channel attention feature maps using the formula specified in Equation 1:

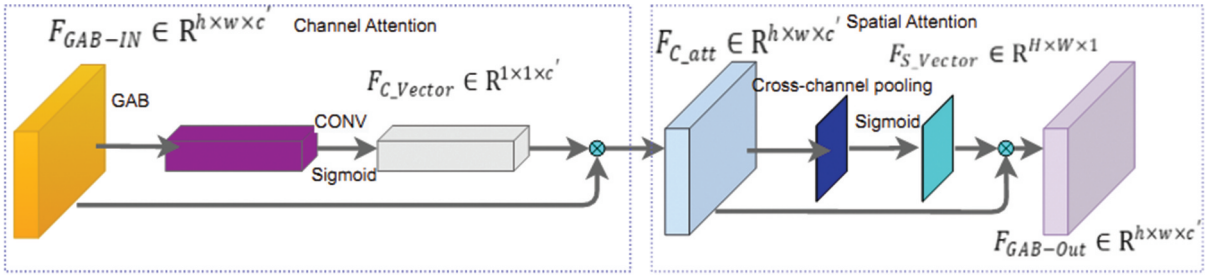


Figure 3. The GAB structure.

$$F_{ch\_att} = (\sigma(Conv2(GAP(F_{GAB-in})))) \otimes F_{GAB-in} \tag{1}$$

Where $F_{ch\_att} \in R^{h\times w\times c'}$, $Conv2$ shows two 1×1 convolutional layers, σ denotes the sigmoid function, GAP denotes the pooling layer, $F_{GAB-in} = F_{reduce}$, and $\otimes$ denotes element-wise multiplication.

After this, we calculate the output of GAB. The spatial attention feature maps ($F_{GAB-out}$) are computed using Eq. 2:

$$F_{GAB-out} = F_{ch\_att} \otimes (\sigma(C\_GAP(F_{ch\_att}))) \tag{2}$$

where C_GAP denotes the average pooling of the cross channel and $F_{GAB-out}$ is the output produced by the category attention feature maps He et al. (2021); Zafar et al. (2024).

Figure 3 also shows the Channel Attention and the Spatial Attention blocks. The former serves as a feature selector at the channel level, learning channel-wise attention weights to determine the significance of each feature channel and suppress less informative channels. The latter highlights the importance of individual spatial positions by learning spatial attention weights. This aspect complements the channel attention mechanism. As indicated in Figure 1, the GAB is placed before the CAB so that we can achieve a sequential arrangement. This arrangement aims to extract comprehensive lesion information globally and preserve smaller lesion regions, thereby minimising the information loss. The CAB focuses more on discriminative regions and enhances the features generated by the GAB. Our study indicates that reversing the order of these two blocks causes the CAB to lose fine-grained details, resulting in negative consequences for the final outcomes.

CAB: The CAB block is constructed to learn discriminative regions from fundus images, enhancing the fine-grained task of classification of DR with L numbers of classes. Considering an incoming feature map denoted as $F_{CAB-in} \in R^{h\times w\times c}$, it is initially processed using a 1×1 convolutional layer, which yields feature maps represented as $F \in R^{h\times w\times kL}$.

The k represents the number of channels needed to identify distinctive regions for individual class levels. To guarantee that each of the k feature maps within a class captures unique discriminative regions, we introduce a dropout mechanism while the model is being trained. Specifically, we randomly eliminate half of the features, setting their values to zero, yielding $F \in R^{h\times w\times kL}$, which preserves half of the elements from every feature map during the dropout process. Nonetheless, during the inference phase, the dropout operation is omitted, and all elements in the k feature maps are employed for prediction. Consequently, the scores $S_i = \{S1, S2, \ldots, SL\}$ for each class can be computed using Equation 3.

$$S_i = \frac{1}{k}\sum_{j=1}^{k} GMP(f'_{ij}), i \in 1, 2.., L \tag{3}$$

To obtain the feature map for each class, a category-wise cross-channel average pooling operation is performed on F'. This operation calculates the average of the feature maps across channels separately for each class as indicated by Equation 4.

$$F'_{i\_avg} = \frac{1}{k}\sum_{j=1}^{k}(f'_{ij}), i \in 1, 2.., L \tag{4}$$

Where $f'_{i,j}$ denotes the representation of the ith class pertaining to the jth feature map extracted from F'. Moreover the term $F'_{i\_avg} \in R^{h\times w\times 1}$ are semantic feature maps specific to the ith class category attention $ATT_{CAB} \in R^{h\times w\times 1}$ that can be obtained by:

$$ATT_{CAB} = \frac{1}{L}\sum_{i=1}^{L} S_i \tag{5}$$

Where the Category Attention Block (ATT_{CAB}) highlights the informative regions for DR grading. Finally, the feature maps input of category attention block F_{CAB-in} can be converted to the feature maps output of category attention block $F_{CAB-out}$ by the ATT_{CAB} as indicated by Equation 6.

$$F_{CAB-out} = F_{CAB-in} \otimes ATT_{CAB} \tag{6}$$

Where $F_{CAB-out}$ represents the resulting feature maps from CAB, which enhances the discerning areas within F_{CAB-in} to determine the severity of the DR. The proposed CAB addresses the issue of imbalanced dataset distribution in DR grading datasets. In conventional CNNs, all feature maps are combined without considering the distinct categories. This may result in mixed information among different categories with less focus on categories with fewer samples of feature maps. To overcome this, the CAB allocates a specific number of feature channels to each DR category, confirming that each DR grade receives an equal number of feature channels. This approach helps prevent channel prejudice and increases the differentiation between different DR categories. As a result, CAB effectively mitigates the problem of imbalanced data distribution commonly encountered in DR grading datasets. CAB efficiently produces attention features using a limited number of parameters that lead to reduced computational cost. It consolidates the distinctive regions related to each category into a single feature map, which shares the same dimensions as the original feature maps. As a result, integrating CAB with the GAB module becomes feasible, thus allowing for a unified combination with significant performance. Algorithm 1 shows the pseudo-code of our approach. From lines (2) to (8), data preprocessing is done in which each image is manually labelled and rescaled into a resolution of 512×512 pixels. Later, augmentation and data manipulations are also performed, as indicated in lines (7) and (8), respectively.

Lines (9) to (18) in Algorithm 1 indicate the selection and application of backbones and several training strategies. Line (10) indicates the loading of three networks, MobileNetV3, Efficientnet-b0, and DenseNet-169, as a backbone model. Meanwhile, pre-trained layers are frozen, trained, and then unfrozen before adding the GAB and the CAB module. Later, Eq (1) and (2) are used to select and highlight the features through the GAB module. To address the class imbalances, Eq 6 is used through the CAB module. Lines (19) through (26) describe the transfer learning and other network parameters. In this step, the number of epochs and the batch size are set to 40 and 16, respectively. Meanwhile, for better CNN training, two different learning rates are set, as depicted in lines (23) and (24) of Algorithm 1, along with the selection of five channels.

Algorithm 1. Pseudocode of proposed DR classification method

1: Input: Obtain colored DR images from Kaggle
2: do:
3: Process collected data obtained in line (1).
4: Perform pre-processing operations:
5: Manually label and rescale each image to a resolution of 512×512 pixels.
6: Perform image augmentation and rotate images at 90°, 180°, and 270°.
7: Perform data manipulation.▸ Isolate into validation, train, and test data set
8: end
9: begin
10: Load MobileNetV3-small OR Efficientnet-b0 OR DenseNet-169, as a backbone model.
11: Select backbone model
12: Freeze the pre-trained layers and train upper layers
13: Unfreeze the pre-trained layers.
14: Add Conv 1x1 Layer
15: Add the GAB and the CAB modules
16: Use Equation1 and Equation2 to highlight and select the features by the GAB.
17: Use Eq. (6) addressing the issue of imbalanced dataset by CAB.
18: end
19: begin

(*Continued*)

Algorithm 1. Pseudocode of proposed DR classification method

20: Transfer learning of the CNN
21: Apply fine tuning and set the parameters as:
22: Epochs = 40 and Batch size = 16,
23: Set learning rate 1 = 5×10^{-3}.
24: Set learning rate 2 = 8×10^{-5}
25: No of channels K = 5
26: end
27: Apply FC and the GAP classifier:
28: Output: Final classification result:
29: DR{Absent, Mild, Moderate, Severe, or Proliferative}

4. Results

In this section, we briefly describe the system descriptions, training settings, datasets used, and evaluation parameters used in this study. We also analyse the performance of our approach on two publicly available datasets and present the discussions and comparisons.

4.1. System specifications and training settings

We ran the experiment on a Google Colab GPU T4, 25 GB of RAM. Our method employs a CNN-based model and an attention module. As described above, we also used other architectures, for instance, MobileNetV3-small, DenseNet-169, and EfficientNet-b0, as a backbone model to see the outcomes yielded by our method. These models have fewer parameters. We also fine-tuned the network to deliver efficient features. Our dataset was divided into training, validation, and testing sets, with portions of 50%, 30%, and 20%, respectively. The network's input resolution is 512×512 pixels. We started with an initial learning rate of 5×10^{-3}. For the non-satisfactory model's improvement on the validation set for three consecutive epochs, we reduced the learning rate by a factor of 0.8. Training was continued for 40 epochs using the Adam optimiser and the cross-entropy loss function. The batch size was set to 32 images per batch.

4.2. Datasets

We used APTOS and EYEPACS benchmark datasets in our experiments. Below, we briefly describe the datasets.

APTOS Dataset: It was launched in 2019 through the Kaggle competition Aptos (2019). This dataset included a fundus image dataset of numerous DR severity levels. Overall, the APTOS dataset consists of 3,669 images. The primary objective of using the fundus imaging dataset was to assess the severity of the disease by generating a probability score that an image belonged to one of the five categories. The usual method to grade DR includes ranking its difficulty into five different classes, which are (i) No DR, (ii) Minor (mild) DR, (iii) Moderate DR, (iv) Severe DR, and (v) Proliferative DR.

EYEPACS Dataset: It consists of 35,108 colour fundus images for DR grading EyePACS (2015). The images were captured under various conditions through several devices, which were provided by EYEPACS. These were placed at multiple primary care sites throughout California and other locations in the USA. We observe a class imbalance, with 73.3% of images belonging to the No DR class, 6.9% to the Minor DR class, 15.2% to the moderate DR class, 2.6% to the severe DR class, and 2% to the Proliferative DR class. The class imbalance issue, as described above, indicates that these datasets are crucial and could be handy to investigate the performance of deep learning algorithms.

4.3. Evaluation metrics

In our work, the performance was assessed using four commonly used metrics as described briefly below from Equation 7 through Equation 10.

$$Acc = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{7}$$

$$Se = \frac{(TP)}{(TP + FN)} \tag{8}$$

$$Sp = \frac{(TN)}{(TN + FP)} \tag{9}$$

$$F1 = \frac{TP}{TP + \frac{1}{2}(FN + FP)} \tag{10}$$

Where *Acc, Se, Sp* indicate the Accuracy, Sensitivity, and the Specificity, respectively. The TP/TN indicates True Positives/Negatives and FP/FN indicates False Positives/Negatives, respectively.

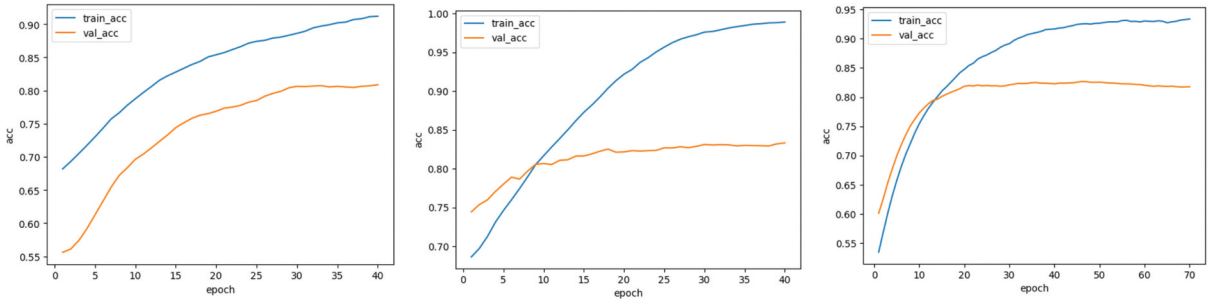
The quadratic weighted kappa (QWK) score for a 5-level classification of DR measures the agreement between raters. It is calculated using the formula:

$$Kappa - score = K = 1 - \frac{\sum_{i,j} W_{i,j} O_{i,j}}{\sum_{i,j} W_{i,j} E_{i,j}} \tag{11}$$

Where, $W_{i,j}$ is the pre-defined weight for agreement between raters i and j, $O_{i,j}$ is the observed agreement, and $E_{i,j}$ is the expected agreement by chance. The resulting K value ranges from [−1,+1], with +1 indicating excellent agreement, 0 indicating chance agreement, and negative values indicate better than random chance agreement.

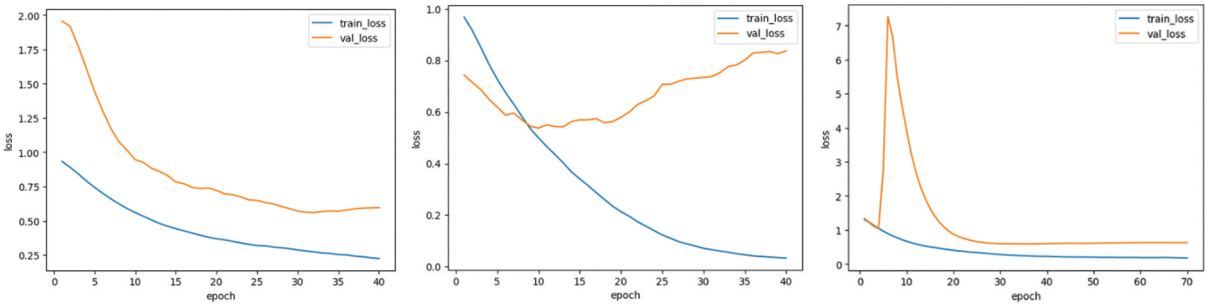
4.4. Classification analysis

We employed supervised learning to classify the five different classes of DR. Figure 4 shows the training curves (accuracy versus epochs) on the APTOS dataset for MobileNetV3-small, DenseNet-169, and EfficientNet-b0 architectures, respectively. These backbones were selected for their proven effectiveness in medical image analysis. Their lightweight architectures and feature extraction capabilities make them well-suited for retinal disease classification tasks such as diabetic retinopathy detection. For MobileNetV3-small, as shown in Figure 4a, both the training and validation accuracy increase with the number of epochs. However, a slight decline in validation accuracy was observed at the 38th epoch. Overall, the training accuracy reaches up to 93% and the validation accuracy reaches 82%, following a monotonic rising pattern. Similarly, Figure 4b depicts that for the DenseNet-169 network, both the training and validation accuracies generally increase by up to 10th epoch. Here, the training accuracy increases by up to 98.5% for 39th epoch, while the validation accuracy stays nearly constant at 84% up to the 40th epoch. For the EfficientNet-b0 architecture, as shown in Figure 3c, it is evident that both the training and validation accuracies rise up to the 17th epoch; however, after that, the validation accuracy curve stays flat at 82%, while the training accuracy approaches nearly 93.5%. In this case, the epochs were set to 70 to observe the detailed trends in training



(a). MobileNetV3-small (b). DenseNet-169 (c). EfficientNet-b0

Figure 4. Training and validation curves for APTOS dataset, accuracy versus number of epochs.



(a). MobileNetV3-small (b). DenseNet-169 (c). EfficientNet-b0

Figure 5. Loss curves for APTOS dataset showing training and validation losses versus number of epochs.

and validation accuracies. The analysis of the training versus accuracy curves indicates that these models converge up to 40 epochs.

Figure 5 shows the loss curves for the APTOS dataset for MobileNetV3-small, DenseNet-169, and EfficientNet-b0 architectures, respectively. As shown in Figure 4a, for MobileNetV3-small, both the training and validation losses decrease when the epochs are set to 40. Overall, it is observed that till the 40th epoch, the training and validation losses are found to be 0.20% and 0.60%, respectively. Moreover, Figure 5b depicts that both the training and validation losses decrease up to 10th epoch. For the 40 epochs that are set for this arrangement, we observe that the training loss converges to nearly zero. For this scenario, the validation loss was found to be 0.80. In Figure 5c, the training and validation losses for EfficientNet-b0 indicate that both the training and validation losses are near zero. The analysis of the loss versus epoch curves presented above on the APTOS dataset suggests that all models can converge within 40 epochs. Based on this analysis, the optimal model is selected by identifying the minimum validation set loss. EfficientNet-b0 was trained for 70 epochs, resulting in an 82% validation accuracy. Additionally, a peak validation accuracy of 84% was attained using DenseNet-169 on the APTOS dataset.

Figure 6 illustrates the key performance metrics of the proposed model evaluated on the APTOS dataset. The model achieved an overall accuracy of 83.6%, along with an F1 score of 82.0%, a precision of 82.1%, a sensitivity of 83.0%, a specificity of 95.5%, and a kappa score of 88.2%. These results highlight the model's strong and balanced ability to detect diabetic retinopathy across different severity levels. The use of distinct hatch patterns enhances visual clarity and facilitates comparison between the evaluation metrics.

The analysis presented above on the APTOS dataset gives good insight into the performance of three different architectures.

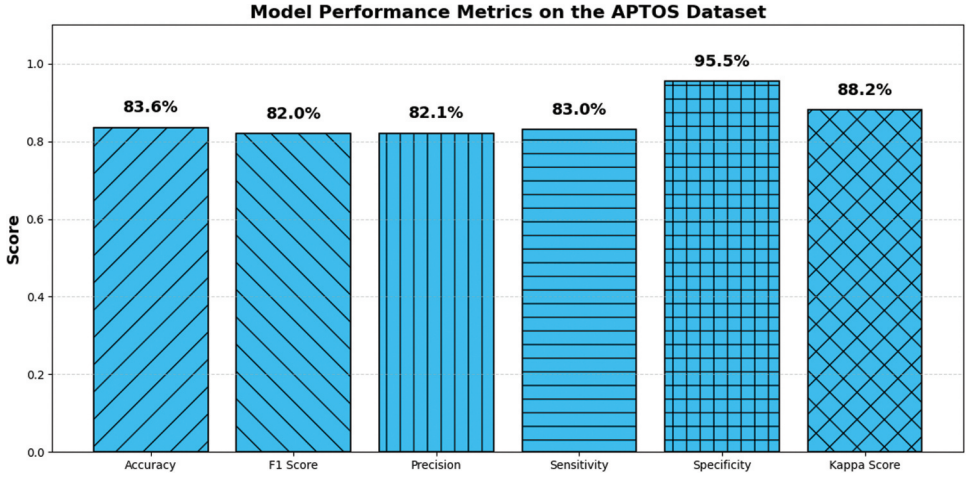


Figure 6. Model performance evaluated on the APTOS dataset. The figure illustrates key metrics including accuracy (83.6%), F1 score, precision, sensitivity, specificity, and kappa score, using distinct hatch patterns for visual clarity. These parameters reflect the effectiveness and robustness of the proposed classification model.

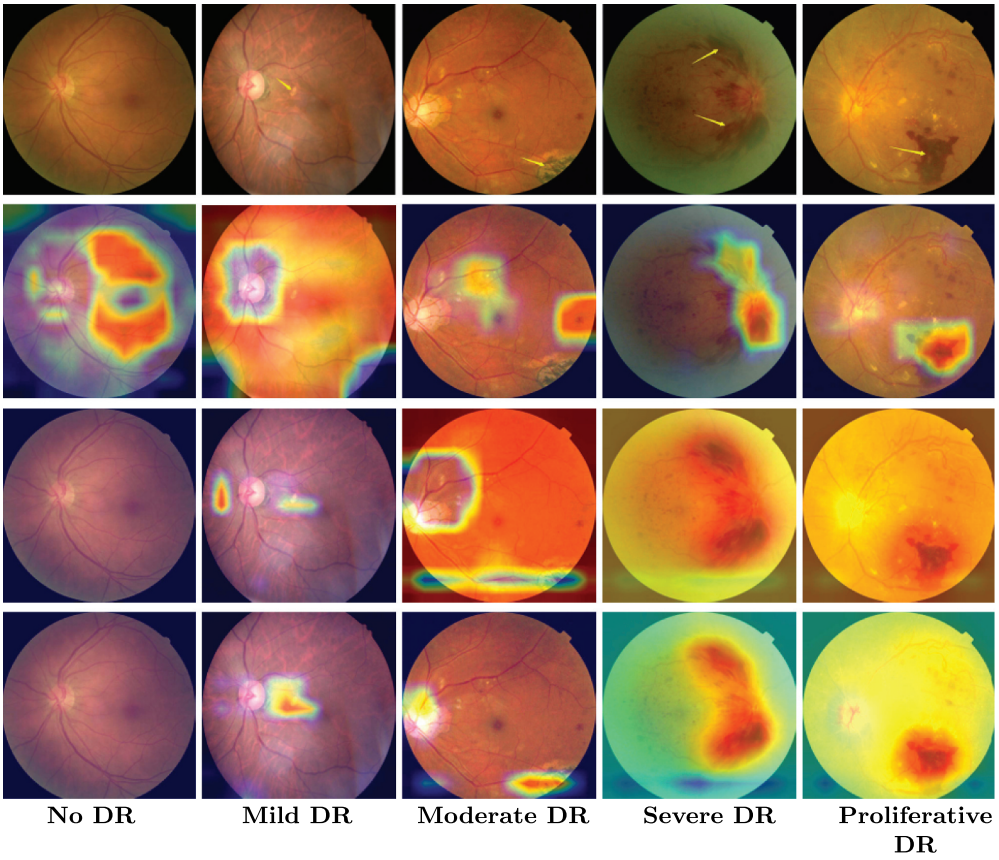


Figure 7. Grad-CAM visualisation results of GAB and CAB on aptos dataset. We show five DR grade levels (0–4 from left to right, i.e. No DR, mild DR, moderate DR, severe DR and Proliferative DR, respectively). The top row provides original images where yellow arrows indicate the lesion regions. The second row provides the heatmaps without attention, the third row provides the heatmaps of GAB, and the bottom row shows the heatmaps refined by CAB.

To evaluate the explainability of the proposed model and understand the influence of CAB, we employed Grad-CAM visualisations. As illustrated in Figure 7, the top row presents five retinal images from the APTOS dataset, representing diabetic retinopathy levels from 0 to 4 (No DR to Proliferative DR). The second row displays attention maps generated without applying attention mechanisms, which tend to highlight irrelevant areas or miss key lesion sites – especially in DR 0, DR 1, DR 3, and DR 4. The third row shows the outputs of the GAB, which improves feature localisation compared to the non-attentive model, but occasionally focuses on non-discriminative regions due to its global perspective. Finally, the fourth row includes results refined by the CAB, demonstrating clearer localisation of disease-specific features and more precise coverage of lesion regions. This improvement enhances the transparency of classification decisions and supports potential clinical adoption. Notably, even subtle lesions – such as those in DR 1 are accurately highlighted by the model, indicating its robustness in detecting early-stage diabetic changes. To enhance clinical trust and support decision-making by ophthalmologists, our model incorporates attention-based interpretability techniques such as Grad-CAM, which visually highlight regions in the retinal fundus image contributing most to the model's prediction. By presenting these attention heatmaps, clinicians can cross-reference the model's focus areas with known pathological features such as microaneurysms, haemorrhages, and exudates. This visual justification not only aligns the model's decisions with clinical reasoning but also increases transparency in the diagnostic process. Furthermore, by showing the evolution of attention from generic to refined lesion-focused regions through the GAB and CAB modules, the model demonstrates a human-like reasoning pattern. Such interpretable outputs are crucial in high-stakes environments like diabetic retinopathy screening, where false negatives or overlooked regions can severely affect patient outcomes.

Table 2. Ablation study of the model adopting DenseNet-169 as the baseline on APTOS dataset. Accuracy, F1 score, and number of parameters (# parameters) are reported.

Method	Accuracy (%)	F1 Score (%)	# Parameters (M)
Baseline (DenseNet-169)	79.0	80.0	16.8
+ GAB only	82.2	81.5	16.8
+ CAB only	82.1	81.2	16.8
+ GAB + CAB (Ours)	**83.6**	**82.0**	17.0

Table 3. Comparison with recent works on DR classification.

	Backbone	APTOS				EyePACS			
Ref		Acc %	F1-score	Kappa-score	#Para	Acc %	F1-score	Kappa-score	#Para
Farag et al. (2022)	DenseNet-169	82	0.68	0.88	8.50	–	–	–	–
He et al. (2021)	DenseNet-121	–	–	–	–	86%	–	0.86	8.12 M
Mustafa et al. (2022b)	DenseNet-121	72	–	–	14 M	85%	–	–	14 M
Hu et al. (2022)	–	83.50	–	–	–	–	–	–	–
Patel and Chaware (2020)	–	80	–	–	2.20 M	–	–	–	–
Khan et al. (2021)	–	–	0.53	–	45 M	–	–	–	–
Proposed Method	MobileNetV3-small	82	0.82	0.88	1.6 M	76.68	0.72	0.52	0.90 M
	DenseNet-169	**83.60**	**0.82**	**0.88**	**17 M**	**75.43**	**0.73**	**0.59**	**18 M**
	EfficientNet-b0	80	0.79	0.87	7.1 M	80	0.73	–	7.3 M

4.5. Ablation studies on APTOS dataset

To evaluate the effectiveness of the proposed dual-attention mechanism, we conducted an ablation study using DenseNet-169 as the baseline model on the APTOS dataset. Table 2 presents the comparison of different configurations in terms of accuracy, F1 score, and number of parameters. The baseline model without any attention modules achieved an accuracy of 79.0% and an F1 score of 80.0%. Incorporating the Global Attention Block (GAB) alone improved the performance to 82.2% accuracy and 81.5% F1 score, highlighting its effectiveness in capturing global contextual features. The use of only the Channel Attention Block (CAB) yielded a similar improvement, reaching 82.1% accuracy and 81.2% F1 score, demonstrating its capability in enhancing relevant channel-wise information. The best results were achieved when both GAB and CAB were integrated, resulting in an accuracy of 83.6% and F1 score of 82.0%. Importantly, the total number of parameters increased only slightly from 16.8 M to 17.0 M, confirming the lightweight nature of the model. This study confirms that each attention module contributes positively to performance, and their combination provides complementary advantages for diabetic retinopathy classification. The individual contributions of the GAB and CAB modules are quantified in Table 2. When added separately to the baseline model, GAB improved the accuracy by 3.2% and the F1 score by 1.5%, while CAB alone improved the accuracy by 3.1% and the F1 score by 1.2%. The combination of both modules led to the best performance, with a total gain of 4.6% in accuracy and 2.0% in F1 score. These results clearly demonstrate that each attention block contributes uniquely to performance improvements, and their synergy provides complementary enhancements in diabetic retinopathy classification. The dual-attention mechanism – comprising Global Attention Block (GAB) and Channel Attention Block (CAB) – was chosen intentionally to capture both spatial and channel-wise contextual information, which are crucial for accurate diabetic retinopathy classification. GAB focuses on the global spatial structure of the lesion areas, while CAB highlights the most relevant feature channels. Although this setup adds modest complexity, the ablation study in Table 2 clearly demonstrates that each module contributes uniquely and significantly to performance improvement. Compared to models with only one attention mechanism, our dual-attention approach achieves better accuracy and F1 scores with minimal increase in parameters. This performance gain justifies the slightly higher architectural complexity, especially given the lightweight nature of our model, which remains suitable for real-time and low-resource applications.

5. Discussion

Advancements in fundoscopy devices and deep learning algorithms have sparked significant interest in automatic DR screening. While deep learning techniques have demonstrated promising results in classifying the DR, there still remains a notable disparity when it comes to their practical application in clinical settings.

The improved classification accuracy of the proposed model carries significant clinical relevance. Enhanced precision – particularly the reduction of false negatives – enables earlier detection of diabetic retinopathy (DR), which is critical in preventing irreversible vision loss and ensuring timely treatment. The model's reliability supports its use as a decision-support tool for ophthalmologists, acting as a second reader and reducing diagnostic workload in large-scale screening programmes. Notably, the model is lightweight, with a reduced number of parameters compared to conventional architectures, making it highly efficient and practical for real-world deployment. Its low computational cost allows seamless integration into real-time DR screening pipelines, including on-edge devices and mobile health platforms. This makes it especially suitable for use in low-resource or rural healthcare settings, where access to specialised diagnostic tools is limited. In summary, the model not only advances diagnostic accuracy but also enhances scalability, accessibility, and practical utility in diverse clinical environments. Due to its compact architecture and reduced parameter count, the model achieves faster inference speeds with minimal hardware requirements. This makes it ideal for deployment in mobile screening units, edge devices, or real-time clinical workflows without compromising diagnostic accuracy.

• This study presents the CABNet, which is a novel method specifically developed to tackle the issue of imbalanced data distribution during DR detection. To further enhance the model's interpretability, we additionally generate location maps that indicate potentially abnormal lesions in fundus images. This feature aids ophthalmologists in making more accurate judgements based on the model's outputs. The experiments demonstrate the flexibility of our approach and further show its effectiveness across various backbones and its ability to achieve superior performance on two publicly available datasets. The success of our approach in the DR grading can be attributed to two key elements, as discussed in the preceding experimental section: the Category Attention Block and CABNet, which is a combination of CAB and GAB.

• The dual attention-based model proposed in this study holds promising clinical potential for real-world diabetic retinopathy (DR) screening. Early diagnosis is vital in preventing permanent vision loss, especially in remote or underserved areas where access to ophthalmologists is limited. By effectively identifying underrepresented stages such as mild and moderate DR, the model can assist in timely intercession that may otherwise be missed by standard methods. Its lightweight architecture and reduced parameter count make it suitable for mobile and teleophthalmology applications, supporting scalable and cost-effective screening initiatives. Moreover, the attention mechanisms enhance the model's transparency by accenting key retinal abnormalities such as haemorrhages and microaneurysms, making the outputs easier for clinicians to validate. These features enable the model to function as a practical assistive tool in clinical workflows, reducing diagnostic burden while improving detection coverage and patient outcomes.

• The RL-MODE framework Singh et al. (2024) integrates reinforcement learning with multi-objective decision-making to optimise Quality of Service metrics in resource-constrained IoT networks. This hybrid approach dynamically balances critical parameters including power efficiency, data transmission delays, and network throughput. Such adaptive optimisation techniques show promising potential for managing multifaceted health data streams in gestational diabetes care, where they could enable more precise monitoring and tailored treatment strategies for pregnancy-related metabolic complications. Recent research by Singh et al. (2024) demonstrates how artificial intelligence-enhanced optimisation techniques can strengthen security systems in consumer devices. The proposed methodology employs bio-inspired search algorithms to dynamically improve threat identification while maintaining system efficiency. These advanced computational strategies show significant potential for adaptation to healthcare diagnostics, particularly in developing more sensitive screening tools for diabetes-related reproductive health disorders through pattern recognition in complex clinical data. The integration of AI-driven metaheuristic algorithms enhances classification accuracy, facilitating earlier and more precise detection of diabetic complications in gynaecological health. This advancement contributes to timely interventions and reduces the likelihood of missed diagnoses, thereby improving patient outcomes (Singh et al. (2025)). The AI-powered metaheuristic framework presented in this study is designed for efficient integration into real-time screening systems. Its lightweight and adaptable nature makes it suitable for deployment in low-resource healthcare settings, enhancing the timely detection of diabetic complications in gynaecological health Kumar et al. (2024). The survey systematically classifies facial recognition techniques into three primary categories: static image analysis, dynamic video processing, and three-dimensional modelling approaches. (Mahmood et al. (2017)). Through comparative evaluation using benchmark datasets, the analysis highlights the distinct advantages and

constraints of each methodology while providing practical recommendations for algorithm selection tailored to different implementation scenarios (Mahmood et al. (2017)).

• Despite its commendable performance, our proposed method still has areas for enhancement. To illustrate practical applicability, imagine a healthcare setting where a technician captures a fundus image using a portable retinal camera. Our proposed model processes the image and identifies moderate DR, prompting timely referral to a specialist. This scenario emphasises the model's potential for supporting DR screening in under-resourced environments. Firstly, the entire network is trained solely with image-level supervision, which makes it highly challenging to pinpoint smaller lesion regions accurately. Secondly, in terms of clinical applications, our model can furnish a grading score and an approximation of the lesion regions. However, it does not address localisation and detection tasks. Furthermore, our work does not specify the type of DR lesions, such as soft exudate, hard exudate, microaneurysm, or haemorrhage. This information is crucial for the DR screening and should be a focal point for future research endeavours.

5.1. Comparison

Table 3 compares our method with a few recent methods. Below, we list important observations while comparing our methods.

• On the APTOS dataset, our proposed method, DenseNet-169 achieved the highest accuracy of 83.60% and thereby outperforms all the six compared methods. Moreover, our proposed approach with MobileNetV3-small yields at par accuracy of 82% with the work developed in Farag et al. (2022). The same is the case with Patel and Chaware (2020),where our proposed EfficientNet-b0 has a similar accuracy. Overall, on the APTOS dataset, the work of Mustafa et al. (2022b) yields the least accuracy than the compared methods. The mean accuracy of our proposed method by combining the three architectures is 81.86%. Furthermore, our approach yields the highest F1-score and Kappa-score on the APTOS dataset. Additionally, our proposed model, MobileNetV3-small requires 1.6 million parameters for the APTOS dataset and 0.9 million for the EYEPACS dataset. On this dataset, the work of Khan et al. (2021)requires the highest number of parameters, which is 40 million.

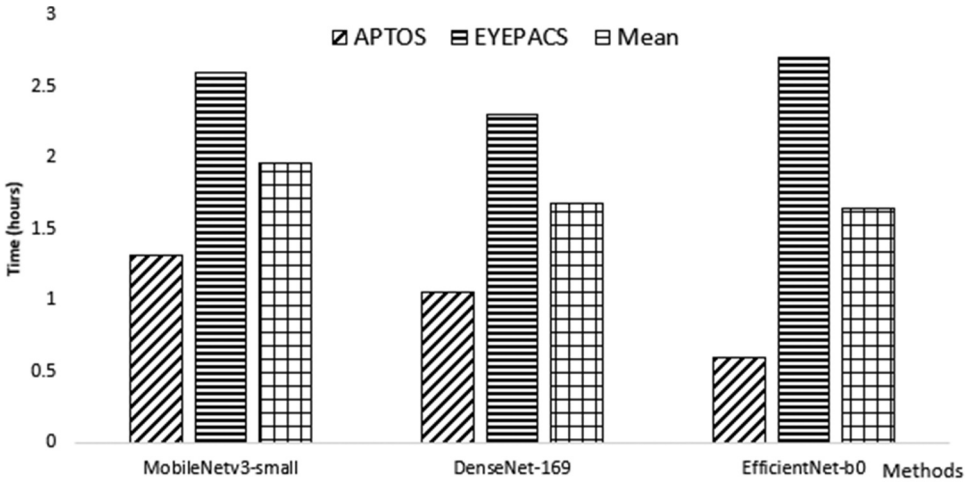
• On the EYEPACS dataset, our proposed methods, MobileNetV3-small, DenseNet-169, and EfficientNet-b0 yield 76.68%, 75.43%, and 80% accuracies, respectively. Moreover, our proposed MobileNetV3-small, DenseNet-169, and EfficientNet-b0 yield 76.68%, 75.43%, and 80% accuracy, respectively. Although here, our method ranks 3rd in terms of accuracy, but it achieved an F1-score of 0.72. Further, our proposed model, MobileNetV3-small and EfficientNet-b0 requires 0.9 million and 7.3 million parameters, respectively, on the EYEPACS dataset. These parameters are considerably less than the methods reported in He et al. (2021) and Mustafa et al. (2022b).

• As indicated in Table 3, most of the methods either require high computations or have limited validation on datasets. Our work is the newest addition to the domain, and it uses a fine-tuned version of the GAB. We anticipate relatively higher DR classification accuracy with much less computational complexity. Similarly, a few of the above-described methods do not handle the class imbalance problem.

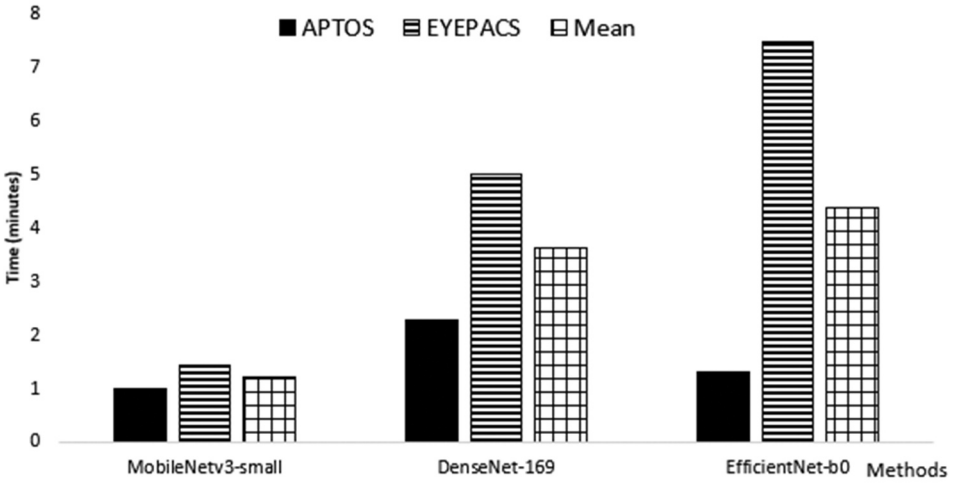
Our proposed method handles the class imbalance issue with the aim of achieving accurate and reliable DR classification. In this research, multiple architectures are implemented and analysed. Therefore, we are optimistic that our developed algorithm will be robust against the variations that are provided in the standard datasets, such as APTOS and EYEPACS. Table 3 shows that on the APTOS dataset, the proposed algorithm outperforms MobliNetV3-small (backbone) in terms of parameters and accuracy. The bottom line of this study is that on the APTOS dataset, the MobileNetV3-small architecture in our proposed algorithm requires 1.6 million parameters and outperforms Farag et al. (2022), Mustafa et al. (2022b), Patel and Chaware (2020), and Khan et al. (2021). On the other hand, on the APTOS dataset, our proposed algorithm outperforms Farag et al. (2022), He et al. (2021), Mustafa et al. (2022b), Hu et al. (2022), Patel and Chaware (2020), and Khan et al. (2021) by yielding 83.6% accuracy using the DenseNet-169 backbone.

5.2. Computational complexity

Figure 8a illustrates the computational complexity of our approach during the training phase on both the APTOS and EYEPACS datasets. The x-axis in Figure 8a represents the backbone networks, which are



(a). Time for training computations.



(b). Time for test computations.

Figure 8. Computational complexity in terms of computational time for training and test phases.

MobileNetV3-small, Densenet-169, and EfficientNet-b0. On the y-axis, the time consumed is depicted in hours. During the training stage, the resolution of the images was set to 512×512 pixels. On the APTOS dataset, the EfficientNet-b0 consumed 0.60 hours, followed by the DenseNet-169 and MobileNetV3-small, which consumed 1.06 and 1.32 hours, respectively. Hence, on the APTOS dataset, the EfficientNet-b0 requires the least time, while the MobileNetV3-small requires the highest time. On the EYEPACS dataset, which contains over 35,000 images, the DenseNet-169 consumed 2.30 hours during the training phase, while the MobileNetV3-small and EfficientNet-b0 consumed 2.60 and 2.30 hours, respectively. As indicated in Figure 8a, on the EYEPACS dataset, DenseNet-169 consumed the least time during the training phase. Moreover, on average, EfficientNet-b0 requires 1.65 hours during the training phase on both datasets, while the DenseNet-169 and MobileNetV3-small, on average, consumed 1.68 and 1.96 hours, respectively. Therefore, during the training phase, our study indicates that the EfficientNet-b0 is computationally the most effective, followed by the DenseNet-169 and the MobileNetV3-small.

Figure 8b shows the test time of the dataset in minutes. For the 512×512 pixel image resolution, on the APTOS dataset, the MobileNetV3-small consumed nearly 1 minute to test the whole dataset, while the DenseNet-169 and EfficientNet-b0 consumed 2.30 and 1.30 minutes, respectively. So, on the APTOS dataset, the MobileNetV3-small consumed the least time. On the EYEPACS dataset, EfficientNet-b0 consumed the highest 7.50 minutes to test this dataset, followed by the DenseNet-169 and MobileNetV3-small, which

consumed 5 minutes and 1.45 minutes, respectively. Moreover, as indicated in Figure 8b, the MobileNetV3-small consumed 1.22 minutes during the test phase on both datasets. Furthermore, the DenseNet-169 and EfficientNet-b0, on average, consumed 3.65 and 4.40 minutes, respectively. Therefore, during the test phase, our study indicates that the MobileNetV3-small is computationally most efficient, followed by DenseNet-169 and EfficientNet-b0, respectively.

6. Conclusion

In this paper, we presented a deep learning pipeline for DR grading. The proposed pipeline incorporates CABNet, a novel approach that merges the CAB and the GAB modules. We trained CABNet holistically for DR grading, leveraging an attention module to learn distinctive features. The proposed approach classifies retinal images into five grades of DR. The approach employs a CNN-based model and an attention module. We have also explored MobileNetV3-small, DenseNet-169, and EfficientNet-b0 as the backbone model. We evaluated the proposed method on the APTOS and the EYEPACS datasets. Experimental results revealed that on the APTOS dataset, the DenseNet-169 achieved a mean accuracy of 83.20%, followed by the MobileNetV3-small and EfficientNet-b0, which achieved 82% and 80%, respectively. On the EYEPACS dataset, the EfficientNet-b0 yielded a mean accuracy of 80%. On this dataset, the DenseNet-169 and MobileNetV3-small yielded 75.43% and 76.68% accuracies, respectively. In addition, the F1-score and Kappa-score are also reported. Our proposed method with MobileNetV3-small consumed 0.90 million parameters on the EYEPACS dataset. Overall, the proposed method produces results on par with recently reported works. Utilising synthetic datasets for initial deep model training, followed by fine-tuning authentic retinal fundus data, could enhance the overall performance in grading diabetic retinopathy. Hence, in the future, we aim to utilise neural diffusion models to generate high-quality retinal fundus images that can help in creating a generalised deep learning model for DR grading. This approach is expected to enhance model robustness and improve performance across diverse clinical scenarios.

Acknowledgments

Z. M. and H. A. designed the research. A. H. conducted the experiments. A. H., R. Q., H. A. conducted the analysis and interpretation of the results. H. A. and Z. M. supervised the work. A. H. and Z. M. wrote the manuscript. H. A. and R. Q. revised the manuscript. All authors reviewed the manuscript.

Author contributions

CRediT: **Abdul Hannan:** Methodology, Writing – original draft; **Zahid Mahmood:** Conceptualization, Resources, Supervision, Writing – original draft; **Rizwan Qureshi:** Formal analysis, Validation, Writing – review & editing; **Hazrat Ali:** Conceptualization, Formal analysis, Investigation, Supervision, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Availability of data and materials

All data analysed during this study are publicly available from https://www.kaggle.com/datasets/mariaherrerot/aptos2019 and https://www.kaggle.com/c/diabetic-retinopathy-detection.

References

Ahmad W, Ali H, Shah Z, Azmat S. 2022. A new generative adversarial network for medical images super resolution. Sci Rep. 12(1):9533. doi: 10.1038/s41598-022-13658-4.

Aptos. 2019. The aptos dataset. https://www.kaggle.com/datasets/mariaherrerot/aptos2019 .

Atwany MZ, Sahyoun AH, Yaqub M. 2022. Deep learning techniques for diabetic retinopathy classification: a survey. IEEE Access. 10:28642–28655. doi: 10.1109/ACCESS.2022.3157632.

Balagurunathan Y, Beers A, Mcnitt-Gray M, Hadjiiski L, Napel S, Goldgof D, Perez G, Arbelaez P, Mehrtash A, Kapur T, et al. 2021. Lung nodule malignancy prediction in sequential ct scans: summary of isbi 2018 challenge. IEEE Trans Med Imag. 40(12):3748–3761. doi: 10.1109/TMI.2021.3097665.

Bhatia K, Arora S, Tomar R. 2016. Diagnosis of diabetic retinopathy using machine learning classification algorithm. 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), Dehradun, India: IEEE. p. 347–351. doi: 10.1109/NGCT.2016.7877439.

Bhulakshmi D, Rajput DS. 2024a. Feddl: personalized federated deep learning for enhanced detection and classification of diabetic retinopathy. Peerj Comput Sci. 10:e2508. doi: 10.7717/peerj-cs.2508.

Bhulakshmi D, Rajput DS. 2024b. Privacy-preserving detection and classification of diabetic retinopathy using federated learning with feddeo optimization. Syst Sci & Control Eng. 12(1):2436664. doi: 10.1080/21642583.2024.2436664.

Chen R, Xu S, Ding Y, Li L, Huang C, Bao M, Li S, Wang Q. 2023. Dissecting causal associations of type 2 diabetes with 111 types of ocular conditions: a mendelian randomization study. Front Endocrinol. 14:1307468. doi: 10.3389/fendo.2023.1307468.

Dai L, Fang R, Li H, Hou X, Sheng B, Wu Q, Jia W. 2018. Clinical report guided retinal microaneurysm detection with multi-sieving deep learning. IEEE Trans Med Imag. 37(5):1149–1161. doi: 10.1109/TMI.2018.2794988.

Dutta S, Manideep BC, Basha SM, Caytiles RD, Iyengar NCSN. 2018. Classification of diabetic retinopathy images by using deep learning models. Int J Grid Distrib Comput. 11(1):99–106. doi: 10.14257/ijgdc.2018.11.1.09.

Elswah DK, Elnakib AA, Din Moustafa HE. 2020. Automated diabetic retinopathy grading using resnet. 2020 37th National Radio Science Conference (NRSC); Cairo, Egypt. 1–6. doi: 10.1109/NRSC49008.2020.9267116.

Esfahani MT, Ghaderi M, Kafiyeh R. 2018. Classification of diabetic and normal fundus images using new deep learning method. Leonardo Electron J Pract Technol. 17(32):233–248.

EyePACS. 2015. The eyepacs dataset. https://www.kaggle.com/c/diabetic-retinopathy-detection.

Farag MM, Fouad M, Abdel-Hamid AT. 2022. Automatic severity classification of diabetic retinopathy based on densenet and convolutional block attention module. IEEE Access. 10:38299–38308. doi: 10.1109/ACCESS.2022.3165193.

Federation ID. 2021. IDF diabetes atlas. 10th ed. Brussels, Belgium: International Diabetes Federation.

Fu H, Cheng J, Xu Y, Wong DWK, Liu J, Cao X. 2018. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Trans Med Imag. 37(7):1597–1605. doi: 10.1109/TMI.2018.2791488.

Fu H, Cheng J, Xu Y, Zhang C, Wong DWK, Liu J, Cao X. 2018. Disc-aware ensemble network for glaucoma screening from fundus image. IEEE Trans Med Imag. 37(11):2493–2501. doi: 10.1109/TMI.2018.2837012.

Gao Z, Li J, Guo J, Chen Y, Yi Z, Zhong J. 2019. Diagnosis of diabetic retinopathy using deep neural networks. IEEE Acces. 7:3360–3370. doi: 10.1109/ACCESS.2018.2888639.

Guo S. 2022. Lighteyes: a lightweight fundus segmentation network for mobile edge computing. Sensors (Switzerland). 22(9):3112. doi: 10.3390/s22093112.

Haq I, Mazhar T, Asif RN, Ghadi YY, Ullah N, Khan MA, Al-Rasheed A. 2024. Yolo and residual network for colorectal cancer cell detection and counting. Heliyon. 10(2):24403. doi: 10.1016/j.heliyon.2024.e24403.

He A, Li T, Li N, Wang K, Fu H. 2021. Cabnet: category attention block for imbalanced diabetic retinopathy grading. IEEE Trans Med Imag. 40(1):143–153. doi: 10.1109/TMI.2020.3023463.

He K, Gkioxari G, Dollar P. 2017. Mask r-cnn. Proceedings of ICCV; Venice, Italy. p. 2980–2988.

He K, Zhang X, Ren S. 2016. Deep residual learning for image recognition. Proceedings of CVPR; Las Vegas, NV, USA. p. 770–778.

Hu J, Wang H, Wang L, Lu Y. 2022. Graph adversarial transfer learning for diabetic retinopathy classification. IEEE AcceSs. 10:119071–119083. doi: 10.1109/ACCESS.2022.3220776.

Iqbal MS, Ali H, Tran SN, Iqbal T. 2021. Coconut trees detection and segmentation in aerial imagery using mask region - based convolution neural network. IET Comput Vision. 15(6):428–439. doi: 10.1049/cvi2.12028.

Iqbal T, Ali H. 2018. Generative adversarial network for medical images (mi-gan). J Med Syst. 42(11):231. doi: 10.1007/s10916-018-1072-9.

Jain A, Jalui A, Jasani J, Lahoti Y, Karani R. 2019. Deep learning for detection and severity classification of diabetic retinopathy. Proceedings 1st International Confernce on Innovation Information in Computing Technologies (ICIICT), Chennai, India, p. 1–6. doi: 10.1109/ICIICT1.2019.8741456.

Jiang H, Yang K, Gao M, Zhang D, Ma H, Qian W. 2019. An interpretable ensemble deep learning model for diabetic retinopathy disease classification. Proceedings 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society; (EMBC), Berlin, Germany; 2045–2048.

Kassani SH, Kassani PH, Khazaeinezhad R, Wesolowski MJ, Schneider KA, Deters R. 2019. Diabetic retinopathy classification using a modified xception architecture. 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Ajman, United Arab Emirates, p. 1–6. doi: 10.1109/ISSPIT47144.2019.9001846.

Khan Z, Khan FG, Khan A, Rehman ZU, Shah S, Qummar S, Ali F, Pack S. 2021. Diabetic retinopathy detection using vgg-nin a deep learning architecture. IEEE AccesS. 9:61408–61416. doi: 10.1109/ACCESS.2021.3074422.

Kumar SP, Dhiman G, Dhiman G, Vimal S, Viriyasitavat W. 2024. Ai-powered metaheuristic algorithms: enhancing detection and defense for consumer technology. IEEE Consum Electron Mag. 13(1):30–38. doi: 10.1109/MCE.2024.3442450.

Li X, Hu X, Yu L, Zhu L, Fu C-W, Heng P-A. 2020. Canet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. IEEE Trans Med Imag. 39(5):1483–1493. doi: 10.1109/TMI.2019.2951844.

Mahmood Z, Muhammad N, Bibi N, Ali T. 2017. A review on state-of-the-art face recognition approaches. Fractals. 25 (02):1750025. doi: 10.1142/S0218348X17500256.
Marmanis D, Datcu M, Esch T, Stilla U. 2016. Deep learning earth observation classification using imagenet pretrained networks. Ieee Geosci Remote S. 13(1):105–109. doi: 10.1109/LGRS.2015.2499239.
Mustafa H, Ali SF, Bilal M, Hanif MS. 2022a. Multi-stream deep neural network for diabetic retinopathy severity classification under a boosting framework. IEEE Access. 10:113172–113183. doi: 10.1109/ACCESS.2022.3217216.
Mustafa H, Ali SF, Bilal M, Hanif MS. 2022b. Multi-stream deep neural network for diabetic retinopathy severity classification under a boosting framework. IEEE Access. 10:113172–113183. doi: 10.1109/ACCESS.2022.3217216.
Patel R, Chaware A. 2020. Transfer learning with fine-tuned mobilenetv2 for diabetic retinopathy. Proceedings of the 2020 International Conference for Emerging Technology (INCET); Belgaum, India.
Qomariah DUN, Tjandrasa H, Fatichah C. 2019. Classification of diabetic retinopathy and normal retinal images using CNN and SVM. In: Proceedings 12th International Conference on Information and Communication Technologies. Surabaya, Indonesia: ICTS; p. 152–157. doi: 10.1109/ICTS.2019.8850940.
Qummar S, Khan FG, Shah S, Khan A, Shamshirband S, Rehman ZU, Ahmed Khan I, Jadoon W. 2019. A deep learning ensemble approach for diabetic retinopathy detection. IEEE Access. 7:150530–150539. doi: 10.1109/ACCESS.2019.2947484.
Raghavendra U, Fujita H, Bhandary SV, Gudigar A, Tan JH, Acharya UR. 2018. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. Inf Sci (NY). 441:41–49. doi: 10.1016/j.ins.2018.01.051.
Reddy GT, Bhattacharya S, Ramakrishnan SS, Chowdhary CL, Hakak S, Kaluri R, Reddy MP. 2020. An ensemble based machine learning model for diabetic retinopathy classification. 2020 International Conference on Emerging Trends in Information Technology and Engineering (IC-ETITE), Vellore, India: IEEE. p. 1–6. doi: 10.1109/ic-ETITE47903.2020.235.
Ren S, He K, Girshick R, et al. 2015. Faster r-cnn: towards real-time object detection with region proposal networks. Proceedings of NIPS; Montreal, (QC), Canada; 91–99.
Singh SP, Kumar N, Alghamdi NS, Dhiman G, Viriyasitavat W, Sapsomboon A. 2024. Next-gen WSN enabled IoT for consumer electronics in smart city: elevating quality of service through reinforcement learning-enhanced multi-objective strategies. IEEE Trans Consumer Electron. 70(4):6507–6518. doi: 10.1109/TCE.2024.3446988.
Singh SP, Kumar N, Dhiman G, Vimal S, Viriyasitavat W. 2024. AI-powered metaheuristic algorithms: enhancing detection and defense for consumer technology. IEEE Consum Electron Mag. 13(1):30–38.
Singh SP, Kumar N, Dhiman G, Vimal S, Viriyasitavat W. 2025. Ai-powered metaheuristic algorithms: enhancing detection and defense for consumer technology. IEEE Consumer Electron Mag. 14(3):44–52.
Tan JH, Fujita H, Sivaprasad S, Bhandary SV, Rao AK, Chua KC, Acharya UR. 2017. Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network. Inf Sci (NY). 420:66–76. doi: 10.1016/j.ins.2017.08.050.
Wu Y, Xia Y, Song Y, Zhang Y, Cai W 2020. NFN+: a novel network followed network for retinal vessel segmentation. Neural Networks. 126:153–162. doi: 10.1016/j.neunet.2020.02.018.
Yan Z, Yang X, Cheng KT. 2018. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. IEEE Trans Biomed Eng. 65(9):1912–1923. doi: 10.1109/TBME.2018.2828137.
Yang S, Jin Y, Lei J, Zhang S. 2024. Multi-directional guidance network for fine-grained visual classification. Vis Comput. 40 (11):8113–8124. doi: 10.1007/s00371-023-03226-w.
Yang Y, Li T, Li W, Wu H, Fan W, Zhang W. 2017. Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks. 20th International Conference on Medical Image Computing and Computer Assisted Intervention - MICCAI 2017; 2017 Sep 11–13; Proceedings, Part III. (QC) City, QC, Canada: Springer International Publishing. p. 533–540.
Yu T, Xu B, Bao M, Gao Y, Zhang Q, Zhang X, Liu R. 2022. Identification of potential biomarkers and pathways associated with carotid atherosclerotic plaques in type 2 diabetes mellitus: a transcriptomics study. Front Endocrinol. 13:981100. doi: 10.3389/fendo.2022.981100.
Zafar A, Aftab D, Qureshi R, Fan X, Chen P, Wu J, Ali H, Nawaz S, Khan S, Shah M. 2024. Single stage adaptive multi-attention network for image restoration. IEEE Trans Image Process. 33:2924–2935. doi: 10.1109/TIP.2024.3384838.
Zeng X, Chen H, Luo Y, Ye W. 2019. Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network. IEEE Access. 7:30744–30753. doi: 10.1109/ACCESS.2019.2903171.
Zhao H, Li H, Maurer-Stroh S, Cheng L. 2018. Synthesizing retinal and neuronal images with generative adversarial nets. Med Image Anal. 49:14–26. doi: 10.1016/j.media.2018.07.001.
Zhou S, Chen C, Han G, Xou, H. 2019. Deep convolutional neural network with dilated convolution using small size dataset. In: Proceedings of the 30th Chinese Control Conference (CCC). Guangzhou, China: p. 8568–8572. doi: 10.23919/ChiCC.2019.8865226.