Computing and Informatics, Vol. 45, 2026, 601–622, doi: 10.31577/cai_2026_3_601

USING TRANSFER LEARNING WITH DENSENET121 AND CONVOLUTIONAL BLOCK ATTENTION MODULE (CBAM) FOR ENHANCED DIABETIC RETINOPATHY CLASSIFICATION

Mankiran Kaur

Computer Science & Engineering, Chandigarh University
Gharuan, Mohali, Punjab, India
&
CGC College of Engineering, Chandigarh Group of Colleges
Landran, Mohali, Punjab, India
e-mail: `mankiran.3321@cgc.edu.in`

Puneet Kumar

Computer Science & Engineering, Chandigarh University
Gharuan, Mohali, Punjab, India
e-mail: `puneet.e11454@cumail.in`

Anuj Kumar Gupta

CGC College of Engineering, Chandigarh Group of Colleges
Landran, Mohali, Punjab, India
e-mail: `anuj.3144@cgc.edu.in`

Abstract. Diabetic retinopathy (DR) is the leading cause of vision loss in the world, particularly in individuals who have had a 10 or 15 years history of diabetes. Early detection of the disease and automated fundus image analysis techniques would decrease the possibility of vision loss. In this study, a transfer learning method is proposed, which includes a Convolutional Block Attention Module (CBAM) and optimizes the DenseNet121 architecture, resulting in better feature representations, as well as the focus on relevant retinal area discoveries. The model was trained and evaluated using the publicly available Kaggle Diabetic Retinopathy dataset, which contains labeled color fundus images spanning multiple DR severity levels. To ensure a comprehensive evaluation, metrics such as accuracy, precision, recall, and F1-score were employed. We have obtained 97 % of categorization accuracy and large ablation experiments indicate that channel or spatial attention can be optimally combined. The addition of CBAM improved the network's ability to refine features and focus on clinically meaningful retinal regions. As indicated by the available experimental data, the proposed DenseNet121 + CBAM outperforms both baseline DenseNet121 and VGG model regarding accuracy, precision, recall, and F1-score and can be applied in the real world in the context of DR. There are studies that have undertaken comprehensive ablation which demonstrate two attention modules. Ablation experiments do better than the single-module and baseline ones. Our ablation results further confirmed that the combined attention mechanism provides more stable and reliable performance gains across all metrics. It also shows high levels of improvement in F1-score, precision and recall with the model and this confirms the reliability and flexibility of this model in clinical practice. The findings indicated a consistent increase in the results of the DenseNet121 + CBAM model as compared to the baseline DenseNet121 and its single-module variants, with significant changes in parameters. The positive results of the research on the various performance indicators as well as the information gathered in ablation studies portray the possibility of the approach to aid in early diagnosis and treatment of DR. These improvements strengthen the potential of this approach to support early diagnosis and effective management of diabetic retinopathy.

Keywords: Attention mechanism, CBAM, diabetic retinopathy, DenseNet121, fundus images, transfer learning

Mathematics Subject Classification 2010: 68T07, 68U10, 92C50

1 INTRODUCTION

Diabetic Retinopathy (DR) is a significant microvascular complication, which develops as a consequence of diabetes mellitus, and it is recognized as one of the main causes of preventable vision loss in the world [1]. The main issue that goes hand in hand with this condition lies in its insidious and mostly asymptomatic nature at the initial phases. Failure to detect the disease and administer corrective therapy can lead to permanent and irreversible blindness in the patient, a fact that highlights the importance of screening programs early in the disease and the need to administer treatment promptly during the care of diabetic patients. The conventional method of diagnosis uses the manual examination and grading of retinal fundus photographs followed by the ophthalmologists to determine the severity of

DR. Although the approach is effective in decades of clinical application, it has its inherent limitations. Consequently, computer-aided diagnostic methodologies have been automated to help clinicians in the more effective identification of DR. However, these automated solutions still have challenges regarding irregular image quality, variable lighting level, uneven lesion topography, and presentation of imaging artifacts, which could undermine the accuracy of classification and frustrate the generalization of the models in different environments. Medical image processing has experienced tremendous developments over the recent years. An important aspect in the Diabetic Retinopathy diagnosis has become the concept of deep learning by using Convolutional Neural Network (CNN) models [2]. CNNs are also capable of automatically learning features relevant to raw pixel data, which has eliminated the need to use manual methods of feature extraction. A number of ready-made CNN frameworks, including ResNet, VGG, and the DenseNet, have been adapted successfully to the needs of a medical imaging task. These architectures have initially been created with large-scale natural image datasets, such as ImageNet. The transfer learning application, which uses deep representations of the pathological features is also an emerging aspect that holds promise in diagnosing retinal diseases such as Diabetic Retinopathy (DR). Although they have their successes, the traditional CNNs generally treat all features of space and channels equally and this may reduce their concentration to specific areas related to pathology including exudates, hemorrhages and microaneurysms [3]. Attention mechanisms have emerged as a persuasive approach in order to increase the detection efficacy of the model. In this study, we are incorporating the Convolutional Block Attention Module (CBAM) into the DenseNet121 backbone to progressively diminish feature maps and, therefore, focus on the most salient spatial positions and feature channels. CBAM uses channel attention and spatial attention in an incremental manner. This combination leads to better performance in the model to identify fine-grained pathological clues that are important in correct DR classification. This symbiotic method is especially beneficial in clinical settings where lesions are localized but subtle with subtle lesions, the attention is crucial to the optimality of the model interpretation and diagnostic performance. The novelty of this work lies in the integration of CBAM specifically within the DenseNet121 framework for DR classification, the comprehensive ablation analysis comparing full, channel-only, and spatial-only attention variants, and the demonstration that the combined attention mechanism yields superior refinement of retinal lesion features compared to existing CNN-based approaches. Furthermore, the work provides an explicit examination of how attention enhances interpretability by guiding the model toward clinically meaningful retinal regions, which strengthens its potential utility in real-world screening environments.

1.1 Contribution of This Study

To further validate the effectiveness of the proposed DenseNet121 + CBAM framework, we compared its performance with several widely adopted state-of-the-art models commonly used in diabetic retinopathy classification and medical image analysis. These include ResNet50, InceptionV3, EfficientNet-B0, VGG16, and MobileNetV2, all of which represent strong baselines in terms of architectural depth, feature extraction capability, and computational efficiency. Each model was trained and evaluated using the same experimental settings and identical dataset partitions to ensure fairness in comparison. The results demonstrate that the proposed model consistently outperforms these architectures across key evaluation metrics such as accuracy, precision, recall, and F1-score. The improvements can be attributed to the addition of CBAM, which enhances the network's ability to emphasize diagnostically relevant retinal regions and suppress less informative features. These findings highlight the superiority of the attention-augmented DenseNet121 architecture over existing state-of-the-art models and underscore its potential as a reliable tool for automated diabetic retinopathy screening.

2 RELATED WORK

DR is a significant complication of microvascularity that develops as a consequence of diabetes mellitus, and it is admitted to be among the main causes of avoidable vision loss at the international level [1]. The main issue relating to this condition lies in the fact that at the initial stages of its development, it is mild and can be asymptomatic. Patients with diabetes are at risk of irreversible and permanent blindness without early diagnosis and proper therapeutic intervention, which serves as the primary reason supporting the importance of early screening programs and early therapeutic intervention. The conventional method of diagnosis depends on the manual observation and grading of retinal fundus photograph by ophthalmologists in order to determine the severity of DR. Although the practice has been found to be effective in decades of clinical practice, it has drawbacks. Consequently, computer-aided diagnostic methodologies are automated and have been developed to aid clinicians in detecting DR in a more efficient manner. In the recent years, medical image processing has experienced a lot of improvements. The development of deep learning has become a crucial element in the detection of Diabetic Retinopathy by using the Convolutional Neural Network (CNN) models [2]. The CNNs are adept at automatically deriving pertinent features of unprocessed pixel data and so do not need manual feature extraction schemes. A number of the pre-trained CNN architectures, including ResNet, VGG, and DenseNet, have been successfully adapted to use in medical imaging tasks. It is also believed that the use of transfer learning that can utilize deep representations of the pathological features, and it can be a promising solution for the retinal disease detection such as Diabetic Retinopathy (DR). This work [4] contains many convolutional neural network (CNN) architectures that are oriented to the detection of diabetic retinopathy (DR). A single researcher was able to identify referable DR with a significant sensitivity and specificity when using a deep CNN model, namely InceptionV3, in particular [5]. Elsewhere, a CNN model has been utilized and 80 000 images dataset has been used where the classification accuracy of 75 % was achieved [6]. The transfer learning application has become highly marked in the sphere of deep learning and machine learning since the years 2014 and 2015 due to the overflow of development in deep convolutional neural networks (CNNs) ImageNet (2012) [7], VGG (2014) [8], and ResNet (2015) [9]. Convolutional neural network (CNNs) have been shown to facilitate the convergence of models and generalization via transfer learning. The success of the Dense Nets in this field has been remarkable because its dense connections allow reusing features and edges of the gradient. An example of the appositional strategies employing the spatial and channel-wise attention to draw significant features is the Convolutional Block Attention Module (CBAM) [10]. The approach used DenseNet169 and CBAM to categorize the extent of DR in the APTOS dataset and the accuracy of the method is 82 % when classifying the multiclass DR and 97 % when classifying binary. Their findings underline the role of attention mechanisms in the retinal image processing [10]. In a related application, age-related macular degeneration (ARMD) was classified with 99.44 % accuracy using a dual-attention DenseNet121 framework [11]. Their success demonstrates the wider potential of Dense Net–CBAM integration in retinal disease classification, despite the fact that ARMD and DR are not the same. In addition to classification, CBAM has shown efficacy in lesion segmentation tasks. Recent deep learning approaches have shown improved localization and classification of retinal lesions in diabetic retinopathy images, contributing to more reliable automated disease detection [12]. Despite these efforts, there has been very little research on the specific application of DenseNet121 with CBAM for multiclass DR classification. This is particularly true when combined with ablation studies that separate the effects of channel, spatial, and full attention, as well as a custom multi-layer dense classifier. To close this gap, our work uses DenseNet121 as a base feature extractor, adds CBAM for improved visual attention, uses a novel dense classifier with 256-128-64 units, compares attention modes in an ablation study, and incorporates the Grad-CAM mechanism [13]. To reflect balanced multiclass performance analysis, we have also calculated the macro-averaged precision, recall, and F1-score. As a result, our model is a complete and repeatable solution that improves DR detection interpretability and accuracy. In this section, we will be presenting various works on CBAM, especially retinal imaging. In this work authors integrated CBAM with U-Net and Attention Gate for microaneurysm segmentation on IDRiD dataset [14]. Demonstrated IoU and Dice improvements by emphasizing lesion regions, the authors have introduced scAG, a dual attention gate in skip connections, boosting segmentation accuracy across brain and liver tasks [15, 16].

2.1 Retinal Vessel & DR Segmentation with CBAM

One of the researchers introduced transformer along with Joint pyramid and CBAM for retinal vessel segmentation in DR, achieving ~0.80 mean IoU (Intersection over Union) on CHASE-DB1 dataset [17].

2.2 Component-Wise Attention Comparisons

Recent attention-based studies have demonstrated that the channel and spatial branches of an attention mechanism do not all contribute equally to the classification of medical images. The study “BAM: Block Attention Mechanism for OCT Image Classification”, which was published in IET Image Processing Journal in 2022, examined optical coherence tomography images and found that channel and spatial attention have distinct effects on classification [18]. The work “Channel Prior Convolutional Attention for Medical Image Segmentation” (2024) included comprehensive ablation experiments that investigated the impact of employing channel-only, spatial-only, and different combinations of attention blocks arranged either sequentially or in parallel [19].

2.3 Squeeze-and-Excitation & Variants for DR

The study proposing the Spatial and Channel-wise Excitation Attention Network (SEA-Net) introduced an alternating structure of spatial and channel attention modules for diabetic retinopathy grading. The network enhanced the depiction of retinal lesions by ranking the most pertinent feature channels and information regions in that order. Compared to the use of Squeeze-and-Excitation (SE) modules alone, the findings have shown that a combination of the two types of attention led to a specific enhancement of the performance in terms of grading. This shows that space data is as equally significant as channel recalibration in detecting minor eye defects. Attention modules are finding more applications in medical imaging and DR research [20, 21]. A number of studies published recently have applied Convolutional Block Attention Module (CBAM) or its variations in the medical image segmentation and classification tasks based on deep learning. Although these methods have shown a better refinement of features and their performance, most of them consider CBAM as a combined element instead of examining the independent effect of their sub modules. Few works have carried out the ablation studies in detail of comparing the effect of channel-only attention and spatial-only attention mechanisms. CBAM can be used in the framework of diabetic retinopathy classification as a black-box-enhancement, without an organized assessment of the effect of each attention branch on model working. Our study fills this gap by conducting a structured ablation experiment across four different model variants: (1) a baseline model that does not use attention, (2) a model that uses channel-only attention, (3) a model that uses spatial-only attention, and (4) a model that uses the full CBAM integration. We can evaluate the precise contribution of each attention component to the overall classification accuracy quantitatively thanks to this methodical comparison.

Study (Year)	**Architecture & Method**	**Dataset/Task**	**Key Results**
Gulshan et al. (2016) [5]	InceptionV3 CNN	EyePACS, Messidor-2	Sensitivity: 90.3–97.5 %, Specificity: 93.4–98.5 %
Pratt et al. (2016) [6]	CNN trained from scratch	80 k train/5 k val	Sensitivity $\sim$ 95 %, Accuracy $\sim$ 75 %
He et al. (2016) [9]	ResNet50 (transfer learning)	DR classification	Faster convergence, Better generalization
Woo et al. (2018) [4]	Introduced CBAM (channel + spatial attention)	General vision tasks	Significantly improved classification across models
Farag et al. (2022) [10]	*DenseNet169 + CBAM*	APTOS DR dataset	$\sim$ 97 % accuracy (binary), $\sim$ 82 % (multiclass)
Yan et al. (2021) [11]	Attention-based CNN	Age-related Macular Degeneration (AMD) Classification	Improved AMD classification performance using attention-guided feature extraction
Saranya et al. (2023) [12]	Deep Learning Lesion Detection and Segmentation	Diabetic Retinopathy Retinal Images	Enhanced localization and classification of red lesions in retinal images
Vanaja et al. (2025) [14]	CBAM + U-Net with Attention Gate	IDRiD microaneurysm segmentation	Improved IoU/Dice scores
Kim et al. (2024) [16]	Transformer + JPU + CBAM	Retinal vessel segmentation (DR)	$\sim$ 0.80 mean IoU
Zhang et al. (2025) [17]	CBAM+modified GoogLeNet	DR classification	Effective multi-attention integration
Nabijiang et al. (2022) [18]	Bottleneck Attention (channel vs spatial)	General medical image analysis	Attention component importance varies by lesion type

Table 1. Summary of literature survey

3 LIMITATIONS IN EXISTING LITERATURE

Despite a lot of research having been done in this area, there is still a research gap that needs to be addressed. Some of them are explained as follows.

3.1 Investigating DenseNet121 + CBAM

Very few studies have focused on combining DenseNet with CBAM for DR, even though CBAM has been tested on DenseNet169 and other backbones [10]. This research gap has been addressed in our work.

3.2 Lack of Ablation at the Component Level

CBAM as a whole is used in most of the works without distinguishing between channel-only and spatial-only modules [22].

3.3 Inadequate Generalization Across Datasets

A common shortcoming observed in current diabetic retinopathy (DR) detection studies is the poor generalization performance of models when applied to data from unfamiliar sources. While most of the model shows excellent results on well-known datasets such as APTOS and EyePACS, their performance significantly drops when tested on other retinal image datasets. The principal sources of this performance difference are the differences in domains such as variations in lighting conditions, imaging equipment, the nature of the population, and the patterns of disease presentation in datasets [23]. Although this method saves on time in training and uses existing visual information, it also comes with a significant drawback. The visual representations that have been acquired based on images of real objects in ImageNet are not capable of describing all the intricate textures, color variations and lesion patterns of retinal fundus images [24]. This can cause such pre-trained features to overfit the training data or even negative transfer, causing the knowledge of another domain of unrelated visual information to interfere with learning disease-relevant features. Such dependence on ImageNet models limits the flexibility and the ability of the network to generalize, and typically reduces performance when deployed to medical images which have significant differences with the object-centric images of ImageNet. Most studies rely primarily on visual saliency maps such as Grad-CAM without extensive quantitative validation of attention alignment with pathological regions [25]. With High Computational Complexity CBAM additional parameters and inference time can impede a higher degree of complexity of the deployment of computational functionality in resource-constrained environments such as mobile devices [26]. Lack of Transparency and Reproducibility: In most research, there is a hindrance to independent validation and adoption due to unavailability of public code, complete model configurations or training pipelines [27]. Existing literature on DenseNet121+CBAM in DR detection demonstrates a number of inadequacies such as low interpretability, higher complexity of the model, low reproducibility, low generalization, no ablation studies, unresolved integration problems, and overdependence on ImageNet. All these gaps are directly addressed by our work, enhancing the utility, clarity, and validity of the model. Although the attention mechanisms such as CBAM have the potential to be useful in DR detection, the current body of research has many gaps such as little usage with DenseNet121, no intensive ablation experiments, poor extrapolation to other datasets, relying on ImageNet features, and poor interpretability beyond visualizations, high computational cost and poor reproducibility. These issues can be highly beneficial in improving the creation and implementation of attention-driven DR models [27].

4 PROPOSED WORK

The model proposed to represent the diabetic retinopathy (DR) with its severity levels based on the retina fundus imagery identified with the deep learning technique. It is implemented on the DenseNet121 backbone, which is an extensively used convolutional neural network with a known purpose of identifying fine-grained retinal abnormalities. In order to further increase the feature representation, a Convolutional Block Attention Module (CBAM) is added at the very end of the last convolutional block of DenseNet121. CBAM uses channel and spatial attention mechanism sequentially, which enables the model to concentrate on features that are the most important. Particularly, the Channel Attention Module (CAM) defines what is significant whereas the Spatial Attention Module (SAM) defines where the model should focus in the image. Such refinement facilitated by the attention can enhance the recognition of the small DR signs including micro aneurysms, exudates, and hemorrhages. All dense layers are augmented with a dropout to minimize overfitting and maximize generalization and with a batch normalization to stabilize and speed training. A softmax output, consisting of five neurons (equivalent to the five stages of DR) No DR, Mild, Moderate, Severe, and Proliferative is the last layer. Here, we discuss the effects of implementing CBAM in both VGG and DenseNet121 networks to classify DRs with the use of fundus images. The aim is to enhance the performance of a classification and feature representation by utilizing the mechanisms of attention. The performance of the attention integration is compared using three model configurations. i) DenseNet121 (Baseline Transfer Learning) ii) DenseNet121 + CBAM iii) VGG16 + CBAM. Moreover, ablation experiments were also used to assess the output of the Convolutional Block Attention Module (CBAM) mechanisms individually and in combination. Three configurations were experimented with, namely: Channel Attention Only, Spatial Attention Only, and Full CBAM, a combination of both channel and spatial attention. The findings show that the Full CBAM set up is always better than the individual attentional processes, and this fact shows the synergistic advantages of the two attention processes. In particular, Full CBAM reached the highest classification accuracy of 97 % and models that used channel or spatial attention showed significantly lower results. These results indicate that channel and spatial attention are complementary to each other, which strengthens the feature of the subject and classification accuracy in diabetic retinopathy detection. Figure 1 shows a detailed analysis of the proposed model.

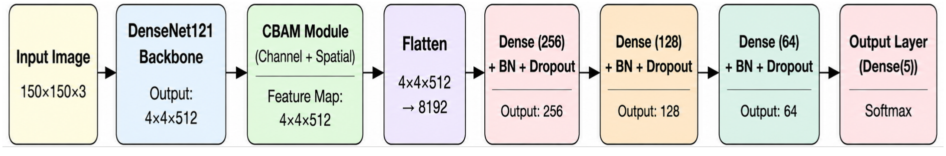

Figure 1. Detailed architecture of the proposed model

4.1 Dataset Description

The dataset used in this study is sourced from the publicly available Diabetic Retinopathy Dataset on Kaggle, which contains labeled retinal fundus images categorized into five severity levels of diabetic retinopathy [28]. This dataset offers a diverse set of images with variations in illumination, contrast, noise, and lesion visibility, making it suitable for evaluating deep learning-based diagnostic systems. Each image is assigned a grade ranging from 0 (No DR) to 4 (Proliferative DR), enabling both binary and multiclass classification tasks. The dataset structure, class distribution, representative examples, and the preprocessing steps applied to the images are documented in detail to ensure transparency and reproducibility. Table 2 provides detailed description of the dataset.

DR Grade	**Description**	**Number of Images**
0	No DR	1 000
1	Mild DR	370
2	Moderate DR	900
3	Severe DR	190
4	Proliferative DR	290
Total		**2 750**

Table 2. Class-wise distribution of images in the Diabetic Retinopathy Dataset

The sample images in Figure 2 illustrate the visual differences across the five diabetic retinopathy severity levels in the dataset. Class 0 images depict a healthy retina with clear vessels and no visible lesions. Class 1 shows early diabetic indicators, typically presenting as a small number of microaneurysms without significant

Figure 2. Sample retinal images from Dataset

structural disruption. Class 2 images contain more noticeable abnormalities, including scattered hemorrhages or small exudates. Class 3 demonstrates extensive retinal damage with numerous lesions and pronounced vascular irregularities, indicating severe disease progression. Class 4 represents the most advanced stage, characterized by pathological features such as neovascularization and widespread retinal deterioration. Together, these samples highlight the increasing complexity and severity of the retinal pathology among the classes.

4.2 Dataset Splits, Preprocessing, and Training Protocol

The dataset was split into 3 subsets (training, validation, and testing) with a stratified sampling approach, ensuring all diabetic retinopathy classes were proportionately represented. Around 80 % of the images were utilized for training, 10 % for validation and remaining 10 % for testing. The partitioning approach aided in preserving the balance between classes and enhancing robustness in evaluation of models.

Various preprocessing techniques were considered for improving the image quality and features enhancement. Images were noise-reduced using a median filter in order to retain the desirable features of the retina. It began with gamma correction and contrast normalization to reduce lighting changes and enhance lesion detectability. Further, we used data augmentation techniques through rotation and flipping, zooming and brightness adjustment to increase the variability of the datasets which helps prevent overfitting.

An AdamW optimizer was utilised with categorical cross-entropy loss to train the proposed DenseNet121 + CBAM model. We are using the batch size of 32 and a learning rate of 1×10^{-4} during the training. In other two techniques early stopping and learning-rate scheduling were used to avoid overfitting and achieve stable convergence, respectively. Google Colab GPU environment was used to run the experiments for achieving efficient and reproducible training performance.

4.3 Dataset Partitioning

The dataset was divided into training, validation, and testing sets using an 80 percent, 10 percent, and 10 percent, respectively. A stratified sampling approach was applied in Google Colab so that all diabetic retinopathy classes were proportionally represented in each subset.

4.4 Class Balancing

To reduce the effect of class imbalance, class weights were used during model training. This ensured that minority classes contributed equally to the optimization process even when their sample counts were lower.

4.5 Preprocessing Procedure

The preprocessing sequence was chosen to improve overall image quality and support stable learning. Median filtering was applied first to remove noise while preserving lesion boundaries. Gamma correction was used to correct brightness differences between images. Normalization was performed last to bring pixel values into a consistent range for training.

4.6 Validation Strategy

A stratified validation split with a fixed random seed (42) was used to ensure consistent and reproducible results. K-fold validation was not applied due to the computational limits of Google Colab.

4.7 Early Stopping Settings

Early stopping monitored the validation loss with the following parameters: patience of 10 epochs, minimum delta of 0.0001, restore best weights enabled. These settings prevented overfitting and supported stable convergence during the 25-epoch training schedule.

4.8 CBAM Insertion

The CBAM module was placed after the final DenseNet121 convolutional block. Channel attention was applied first, followed by spatial attention. A schematic of this integration has been added to the revised figure for clarity.

4.9 Initialization

DenseNet121 used pretrained ImageNet weights, while all newly added dense layers were initialized with He Normal. A fixed random seed was used for consistent initialization across runs in Google Colab.

4.10 Dropout Justification

A dropout value of 0.5 was chosen after initial trials showed that lower values led to slight overfitting, while higher values restricted learning. The chosen value provided an effective level of regularization.

4.11 Batch Size and Epoch Choice

A batch size of 32 and 25 training epochs were selected to match Google Colab's hardware limits. Early stopping often halted training earlier when the model stopped improving.

5 METHODOLOGY

5.1 Data Collection and Preprocessing

There are 2 750 images of retinal fundus and it has been classified into five groups based on the level of diabetic retinopathy: 1 000 images of healthy eyes (No DR), 370 images of Mild DR, 900 images of Moderate DR, 290 images of Proliferative DR, and 190 images of Severe DR. Preprocessing techniques were used to enhance the performance of the model, including the median filtering, gamma correction and contrast normalization techniques to eliminate the effects of noise and uneven lighting. Also, to augment the dataset variability, to improve robustness, and to avoid overfitting, image rotation, horizontal and vertical flipping, and brightness adjustments methods were used. A flowchart of the proposed methodology is shown in Figure 3.

5.2 Transfer Learning Using DenseNet121

Transfer learning is an appropriate approach to medical imaging datasets due to their low volume and high diversities: models trained on large scale natural image datasets like ImageNet are useful due to their generalizability. The choice of the backbone network was determined to be DenseNet121 because of a dense connectivity that ensures efficient flow of features and prevents vanishing gradient issue. ImageNet pre-trained weights were used to first initialize the network, and the original classification layer was replaced with a custom dense classifier that produces five classes with the varying levels of the severity of diabetic retinopathy.

5.3 Convolutional Block Attention Module (CBAM) Is Integrated

Following the last convolutional block of DenseNet121, CBAM module was introduced to the network to increase its capability of focusing on the features which

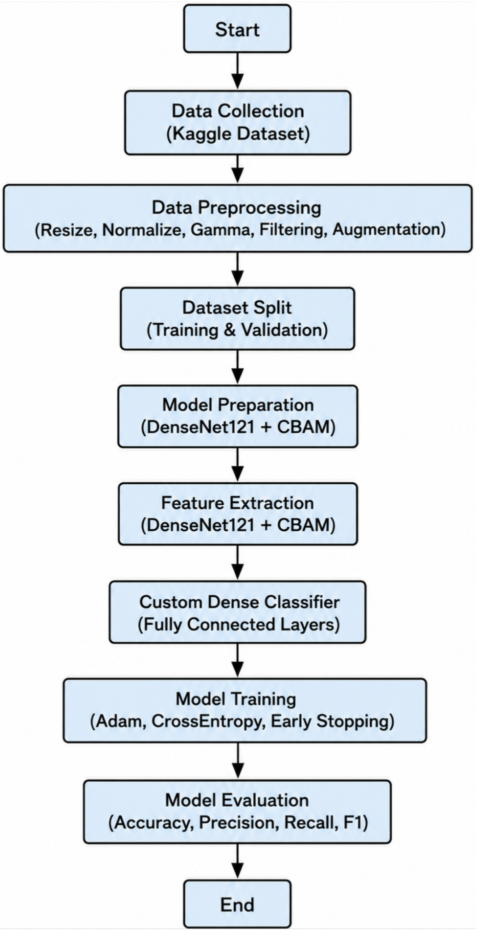

Figure 3. Flowchart of the proposed work

are significant in diagnosis. CBAM uses channel attention and spatial attention sequentially, improving feature maps by highlighting important feature maps related to DR lesion. Figure 4 describes retinal image preprocessing pipeline.

5.4 Custom Dense Classifier Design

Following the extraction of the features, a dense classifier comprising three fully connected layers with 256, 128, and 64 units was constructed. To increase stability and lessen overfitting, batch normalization and a dropout of 0.5 were applied after each layer. The model was able to perform multiclass classification of diabetic retinopathy because the final softmax layer produced the class probabilities.

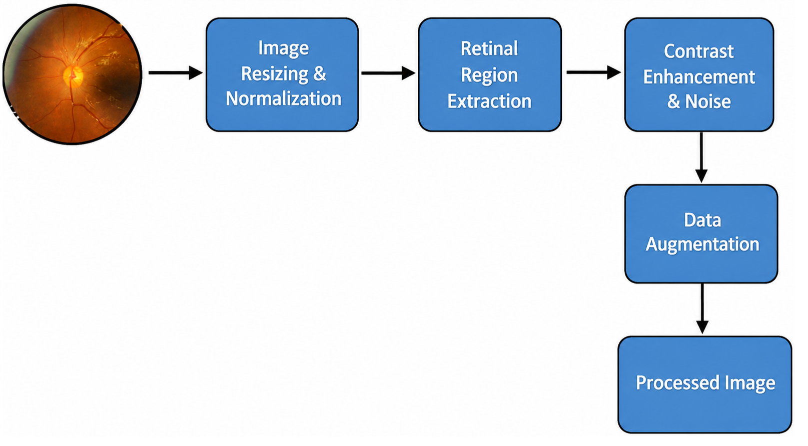

Figure 4. Retinal image preprocessing pipeline

5.5 Model Training and Optimization

The model was trained using the Adam optimizer with an initial learning rate of 1×10^{-4}, which was reduced by a factor of 0.1 whenever the validation accuracy plateaued. Categorical cross-entropy served as the loss function. Early stopping using a patience of ten epochs was used to preclude overfitting. Accuracy, macro-averaged precision, recall, F1-score, and the confusion matrix were some of the measures of an ideal evaluation of the model.

5.6 Training Parameters

Training parameters are summarized in Table 3.

Hyperparameter	**Value**
Learning Rate	1e−4
Batch Size	32
Optimizer	AdamW
Loss Function	Categorical Cross-Entropy
Epochs	50–80
Learning Rate Scheduler	ReduceLROnPlateau/Cosine Annealing
Weight Decay	1e−5
Dropout	0.5
Initialization	He Normal

Table 3. Summary of training hyperparameters

5.7 Experimental Setup

All experiments in this study were conducted using Google Colab's cloud-based environment. Training was performed on Colab's GPU runtime, which provides access to NVIDIA Tesla T4 or P100 GPUs depending on session availability. The environment runs on Ubuntu Linux with CUDA and cuDNN preconfigured by Google's backend. The implementation was developed in Python using TensorFlow/PyTorch (as applicable) along with supporting libraries such as NumPy, OpenCV, and scikit-learn. Using Google Colab ensures a consistent, reproducible setup while providing sufficient computational resources for efficient model training and evaluation.

6 RESULTS AND DISCUSSION

The given model was contrasted with the baseline architectures that are typically used in medical image analysis, i.e. VGG16 and DenseNet121. The findings show that the CBAM-improved model was better than the base models in all the critical evaluation measures. The summary of the comparative performance is found in Table 4.

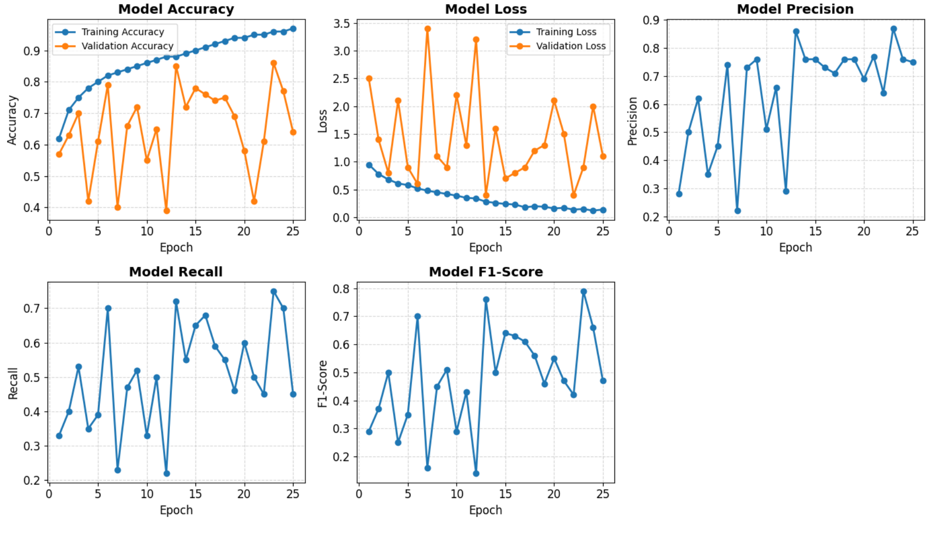

Figure 5. Results of the baseline model – DenseNet121

6.1 Analysis

The proposed model, DenseNet121 + CBAM, was identified to be the most successful in all measures: F1-Score (96.1 %), Accuracy (97.3 %), Precision (97.8 %), and

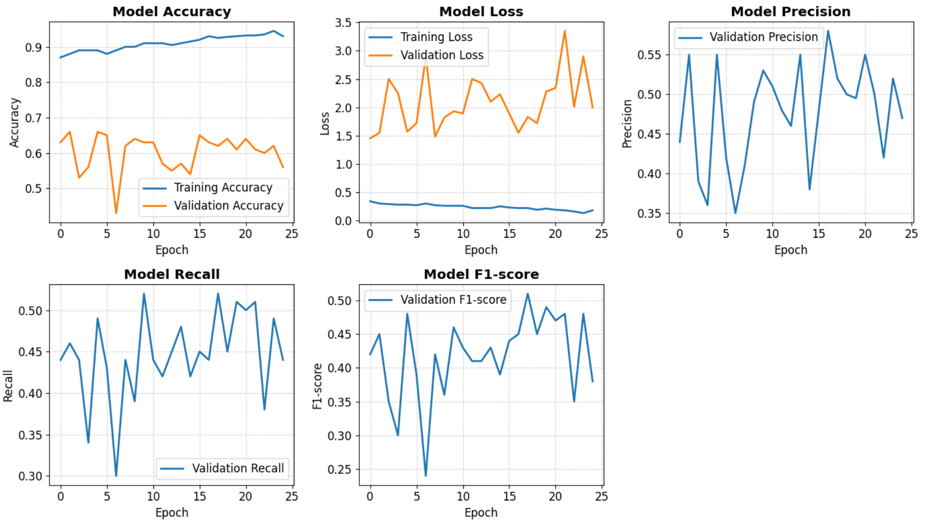

Figure 6. Results of VGG + CBAM

Recall (94.7 %). This is the effect of including the full CBAM module (channel + spatial attention) to increase the focus and classification of the model. Conversely, even though VGG16 + CBAM benefited from attention, it performed noticeably worse than DenseNet121, probably because of VGG16's shallower architecture and poorer feature reuse capability. The outcomes of the model are mentioned in Table 4.

Model	**Accuracy**	Precision	Recall	F1-Score
DenseNet121 (Baseline)	0.961	0.970	0.912	0.930
VGG16 + CBAM	0.940	0.46	0.47	0.40
DenseNet121 + CBAM (Proposed Model)	0.973	0.978	0.947	0.961

Table 4. Tabular comparison with all baseline models

7 ABLATION STUDY

To isolate the contribution of attention mechanisms, ablation experiments were conducted as shown in Figure 8. Table 5 shows the results of the ablation study.

1. Baseline: DenseNet121 without CBAM.
2. Channel Attention only: DenseNet121 with only channel attention.

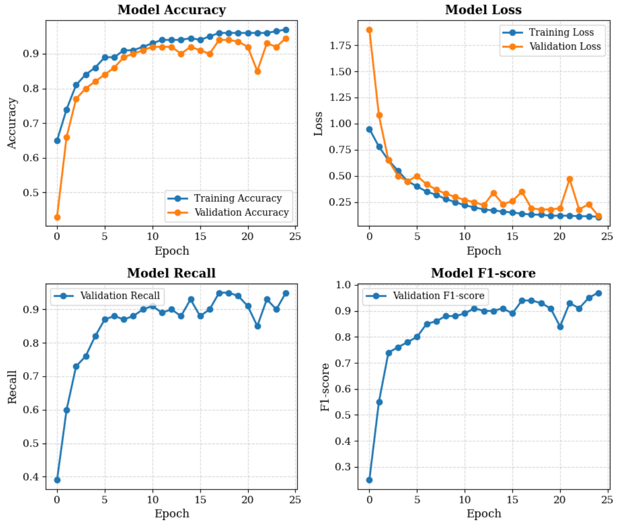

Figure 7. Results of the proposed model

3. Spatial Attention only: DenseNet121 with only spatial attention.
4. Full CBAM: DenseNet121 with both channel and spatial attention.

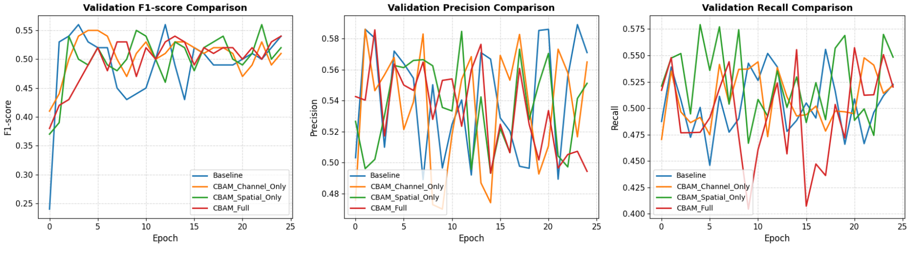

Figure 8. Ablation study comparing baseline, channel attention, spatial attention, and full CBAM configurations

8 CONCLUSIONS

DenseNet121 base already shows a good performance in diabetic retinopathy (DR) classification, with strong performance by virtue of its dense connectivity and efficient flow of features. However, previous research demonstrated that attention mechanisms can be used to further improve the performance of convolutional neural networks (CNNs) in medical imaging by highlighting diagnostically relevant areas.

S. no	**Model**	Final Validation Accuracy
1	BaseLine	0.65
2	CBAM_Channel_Only	0.65
3	CBAM_Spatial_Only	0.64
4	CBAM_Full	0.68

Table 5. The comparison of validation accuracy of all 3 studies

An addition of the Convolutional Block Attention Module (CBAM) adds focus to the network, which causes it to be guided to the most informative regions of the retina, such as microaneurysms, exudates, and hemorrhages, which are usually small, faint, and asymmetrical in distribution across fundus images. CBAM improves the feature representation and the decision-making process as it initially locates the most discriminative feature map and then targets important spatial areas with its sequential channel and spatial attention processes. In the proposed DenseNet121 + CBAM model, the mechanism is used to enhance the accuracy and interpretability of DR detection.

REFERENCES

[1] Kropp, M.—Golubnitschaja, O.—Mazurakova, A.—Koklesova, L.—Sargheini, N. et al.: Diabetic Retinopathy as the Leading Cause of Blindness and Early Predictor of Cascading Complications – Risks and Mitigation. EPMA Journal, Vol. 14, 2023, No. 1, pp. 21–42, doi: 10.1007/s13167-023-00314-8.

[2] Selvachandran, G.—Quek, S. G.—Paramesran, R.—Ding, W.—Son, L. H.: Developments in the Detection of Diabetic Retinopathy: A State-of-the-Art Review of Computer-Aided Diagnosis and Machine Learning Methods. Artificial Intelligence Review, Vol. 56, 2023, No. 2, pp. 915–964, doi: 10.1007/s10462-022-10185-6.

[3] Hassan, D.—Gill, H. M.—Happe, M.—Bhatwadekar, A. D.—Hajrasouliha, A. R.—Janga, S. C.: Combining Transfer Learning with Retinal Lesion Features for Accurate Detection of Diabetic Retinopathy. Frontiers in Medicine, Vol. 9, 2022, Art. No. 1050436, doi: 10.3389/fmed.2022.1050436.

[4] Woo, S.—Park, J.—Lee, J. Y.—Kweon, I. S.: CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.): Computer Vision – ECCV 2018. Springer, Cham, Lecture Notes in Computer Science, Vol. 11211, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.

[5] Gulshan, V.—Peng, L.—Coram, M.—Stumpe, M. C.—Wu, D. et al.: Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA, Vol. 316, 2016, No. 22, pp. 2402–2410, doi: 10.1001/jama.2016.17216.

[6] Pratt, H.—Coenen, F.—Broadbent, D. M.—Harding, S. P.—Zheng, Y.: Convolutional Neural Networks for Diabetic Retinopathy. Procedia Computer Science, Vol. 90, 2016, pp. 200–205, doi: 10.1016/j.procs.2016.07.014.

[7] Krizhevsky, A.—Sutskever, I.—Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira, F., Burges, C. J., Bottou, L., Weinberger, K. Q. (Eds.): Advances in Neural Information Processing Systems 25 (NIPS 2012). Curran Associates, Inc., 2012, pp. 1097–1105, https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[8] Simonyan, K.—Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, 2014, doi: 10.48550/arXiv.1409.1556.

[9] He, K.—Zhang, X.—Ren, S.—Sun, J.: Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[10] Farag, M. M.—Fouad, M.—Abdel-Hamid, A. T.: Automatic Severity Classification of Diabetic Retinopathy Based on DenseNet and Convolutional Block Attention Module. IEEE Access, Vol. 10, 2022, pp. 38299–38308, doi: 10.1109/ACCESS.2022.3165193.

[11] Yan, Y.—Jin, K.—Gao, Z.—Huang, X.—Wang, F.—Wang, Y.—Ye, J.: Attention-Based Deep Learning System for Automated Diagnoses of Age-Related Macular Degeneration in Optical Coherence Tomography Images. Medical Physics, Vol. 48, 2021, No. 9, pp. 4926–4934, doi: 10.1002/mp.15002.

[12] Saranya, P.—Pranati, R.—Patro, S. S.: Detection and Classification of Red Lesions from Retinal Images for Diabetic Retinopathy Detection Using Deep Learning Models. Multimedia Tools and Applications, Vol. 82, 2023, No. 25, pp. 39327–39347, doi: 10.1007/s11042-023-15045-1.

[13] Selvaraju, R. R.—Cogswell, M.—Das, A.—Vedantam, R.—Parikh, D.—Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.

[14] Vanaja, C. B.—Prakasam, P.: Convolutional Block Attention Gate-Based UNet Framework for Microaneurysm Segmentation Using Retinal Fundus Images. BMC Medical Imaging, Vol. 25, 2025, Art. No. 83, doi: 10.1186/s12880-025-01625-0.

[15] Khanh, T. L. B.—Dao, D. P.—Ho, N. H.—Yang, H. J.—Baek, E. T.—Lee, G.—Kim, S. H.—Yoo, S. B.: Enhancing U-Net with Spatial-Channel Attention Gate for Abnormal Tissue Segmentation in Medical Imaging. Applied Sciences, Vol. 10, 2020, No. 17, Art. No. 5729, doi: 10.3390/app10175729.

[16] Kim, H. J.—Eesaar, H.—Chong, K. T.: Transformer-Enhanced Retinal Vessel Segmentation for Diabetic Retinopathy Detection Using Attention Mechanisms and Multi-Scale Fusion. Applied Sciences, Vol. 14, 2024, No. 22, Art. No. 10658, doi: 10.3390/app142210658.

[17] Zhang, B.—He, W.—Wang, R.—Lin, J.—Kuang, H.: Improved GoogLeNet Based on CBAM Module for Fundus Disease Classification Network. Second International Conference on Big Data, Computational Intelligence, and Applications (BDCIA 2024), Proceedings of SPIE, Vol. 13550, 2025, doi: 10.1117/12.3059839.

[18] Nabijiang, M.—Wan, X.—Huang, S.—Liu, Q.—Wei, B.—Zhu, J.—Xie, X.: BAM: Block Attention Mechanism for OCT Image Classification. IET Image Pro-

cessing, Vol. 16, 2022, No. 5, pp. 1376–1388, doi: 10.1049/ipr2.12415.

[19] Huang, H.—Chen, Z.—Zou, Y.—Lu, M.—Chen, C.—Song, Y.—Zhang, H.—Yan, F.: Channel Prior Convolutional Attention for Medical Image Segmentation. Computers in Biology and Medicine, Vol. 178, 2024, Art. No. 108784, doi: 10.1016/j.compbiomed.2024.108784.

[20] Zhao, Z.—Chopra, K.—Zeng, Z.—Li, X.: SEA-Net: Squeeze-and-Excitation Attention Net for Diabetic Retinopathy Grading. 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 2496–2500, doi: 10.1109/ICIP40778.2020.9191345.

[21] Sebastian, A.—Elharrouss, O.—Al-Maadeed, S.—Almaadeed, N.: A Survey on Deep-Learning-Based Diabetic Retinopathy Classification. Diagnostics, Vol. 13, 2023, No. 3, Art. No. 345, doi: 10.3390/diagnostics13030345.

[22] Sajid, M. Z.—Hamid, M. F.—Youssef, A.—Yasmin, J.—Perumal, G.—Qureshi, I.—Naqi, S. M.—Abbas, Q.: DR-NASNet: Automated System to Detect and Classify Diabetic Retinopathy Severity Using Improved Pretrained NASNet Model. Diagnostics, Vol. 13, 2023, No. 16, Art. No. 2645, doi: 10.3390/diagnostics13162645.

[23] Mutawa, A. M.—Alnajdi, S.—Sruthi, S.: Transfer Learning for Diabetic Retinopathy Detection: A Study of Dataset Combination and Model Performance. Applied Sciences, Vol. 13, 2023, No. 9, Art. No. 5685, doi: 10.3390/app13095685.

[24] Lam, C.—Yi, D.—Guo, M.—Lindsey, T.: Automated Detection of Diabetic Retinopathy Using Deep Learning. AMIA Joint Summits on Translational Science Proceedings, 2018, pp. 147–155, `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5961805/`.

[25] Adebayo, J.—Gilmer, J.—Muelly, M.—Goodfellow, I.—Hardt, M.—Kim, B.: Sanity Checks for Saliency Maps. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.): Advances in Neural Information Processing Systems 31 (NeurIPS 2018). Curran Associates, Inc., 2018, pp. 9505–9515, `https://proceedings.neurips.cc/paper_files/paper/2018/file/294a8ed24b1ad22ec2e7efea049b8737-Paper.pdf`.

[26] Jetley, S.—Lord, N. A.—Lee, N.—Torr, P. H. S.: Learn to Pay Attention. International Conference on Learning Representations (ICLR 2018), 2018, doi: 10.48550/arXiv.1804.02391.

[27] Kaur, M.—Kumar, P.—Gupta, A. K.: Survey on Various Feature Detection and Feature Selection Methods Using Retinopathy Images. 2023 International Conference on IoT, Communication and Automation Technology (ICICAT), 2023, pp. 1–6, doi: 10.1109/ICICAT57735.2023.10263673.

[28] Kaggle: Diabetic Retinopathy Detection Challenge. 2015, `https://www.kaggle.com/c/diabetic-retinopathy-detection`.

Mankiran Kaur is currently working as Assistant Professor at the Chandigarh Group of Colleges, Landran with 17 years of total teaching experience. She is also pursuing her Ph.D. as a Research Scholar at the Chandigarh University, Gharuan, Punjab. Her research work focuses on deep learning, medical image analysis, and diabetic retinopathy detection. Her research interests include transfer learning, attention mechanisms, computer vision, and AI-based healthcare applications.

Puneet Kumar is a personality with over two decades of teaching and industrial experience. He is serving the Chandigarh University as Director, IQAC and Professor in computer science and engineering. He obtained his Doctorate degree in the domain of e-governance and mathematical modeling. He has also published many books with publishers like Taylor and Francis, Chapman and Hall Publishing, USA, and Springer, and his research publications are also noteworthy.

Anuj Kumar Gupta is working as full-time Professor at the CGC College of Engineering, Chandigarh, India. He holds his Ph.D. in computer science and engineering with the vast academic and administrative experience of 25 years. He has guided 27 PG and 10 Ph.D. research scholars. He has published more than 110 research papers in reputed international and national journals and conferences. His research areas include Ad hoc networks and network security. He has filed more than 10 patents, with 3 patents granted, and has published 5 books. He is an editorial board member of Elsevier, IEEE, Springer, IGI Global, and many other reputed research agencies. He is also a life member of the Computer Society of India, ISTE, and several other professional bodies.