KSÜ Mühendislik Bilimleri Dergisi, 28(4), 2025
Araştırma Makalesi

KSU J Eng Sci, 28(4), 2025
Research Article

Kahramanmaras Sutcu Imam University
Journal of Engineering Sciences

Geliş Tarihi : 06.08.2025
Kabul Tarihi :14.11.2025

Received Date : 06.08.2025
Accepted Date : 14.11.2025

EYE DISEASE DETECTION WITH DEEP LEARNING MODELS SUPPORTED BY THE CBAM ATTENTION MECHANISM

CBAM DİKKAT MEKANİZMASI İLE DESTEKLENMİŞ DERİN ÖĞRENME MODELLERİYLE GÖZ HASTALIKLARININ TESPİTİ

Rıdvan COŞKUN [1*] (ORCID: 0000-0002-3673-6306)
Duygu KAYA [1] (ORCID: 0000-0002-6453-631X)
Hasan GÜLER [1] (ORCID: 0000-0002-9917-3619)

[1] Fırat University, Department of Electrical Electronics Engineering, Elazığ, Türkiye

*Sorumlu Yazar / Corresponding Author: Rıdvan COŞKUN, rcoskun@firat.edu.tr

ABSTRACT

Early diagnosis of eye diseases plays a critical role in treatment success and public health. With the widespread use of modern medical imaging methods, the development of automated diagnostic systems from retinal fundus images has become an important research area. In this study, the effects of integrating the Convolutional Block Attention Module (CBAM) into EfficientNetB0 and DenseNet121 architectures were investigated for the classification of cataract, diabetic retinopathy, glaucoma, and healthy subjects. Experimental results demonstrated that the CBAM attention mechanism enhances accuracy and generalization performance, particularly in distinguishing complex retinal findings. For DenseNet121, accuracy, precision, recall, and F1-score were obtained as 88.37%, 89.66%, 88.37%, and 88.52%, respectively. EfficientNetB0 achieved 96.32% accuracy, 96.34% precision, 96.32% recall, and 96.33% F1-score. After CBAM integration, the accuracy of DenseNet121 increased to 90.39% and its F1-score to 90.54%, while EfficientNetB0 improved to 96.56% accuracy and 96.57% F1-score. These results reveal that the incorporation of CBAM enhances the performance of deep learning models and significantly contributes to the development of reliable and clinically applicable systems for the automated detection of eye diseases.

Keywords: Deep learning, attention mechanism, CBAM, retinal fundus imaging, eye disease classification

ÖZET

Göz hastalıklarının erken teşhisi, tedavi başarısı ve toplum sağlığı açısından kritik bir rol oynamaktadır. Modern tıbbi görüntüleme yöntemlerinin yaygınlaşmasıyla birlikte, retina fundus görüntülerinden otomatik tanı sistemlerinin geliştirilmesi önemli bir araştırma alanı hâline gelmiştir. Bu çalışmada, katarakt, diyabetik retinopati, glokom ve sağlıklı bireylerin sınıflandırılması amacıyla EfficientNetB0 ve DenseNet121 mimarileri ile Convolutional Block Attention Module (CBAM) entegrasyonunun etkileri incelenmiştir. Deneysel sonuçlar, CBAM dikkat mekanizmasının özellikle karmaşık retinal bulguların ayırt edilmesinde doğruluk ve genelleme performansını artırdığını göstermiştir. DenseNet121 modeli için doğruluk %88,37, kesinlik %89,66, duyarlılık %88,37 ve F1 skoru %88,52 olarak elde edilmiştir. EfficientNetB0 modeli ise %96,32 doğruluk, %96,34 kesinlik, %96,32 duyarlılık ve %96,33 F1 skoruna ulaşmıştır. CBAM entegrasyonu sonrası DenseNet121'in doğruluğu %90,39'a, F1 skoru %90,54'e; EfficientNetB0'un doğruluğu %96,56'ya, F1 skoru ise %96,57'ye yükselmiştir. Sonuçlar, CBAM entegrasyonunun derin öğrenme modellerinin başarımını artırdığını ve otomatik göz hastalığı tespitinde güvenilir, klinik olarak uygulanabilir sistemlerin geliştirilmesine önemli katkı sağladığını ortaya koymaktadır.

Anahtar Kelimeler: Derin öğrenme, dikkat mekanizması, CBAM, retina fundus görüntüleme, göz hastalığı sınıflandırması

ToCite: COŞKUN, R., KAYA, D., & GÜLER, H., (2025). EYE DISEASE DETECTION WITH DEEP LEARNING MODELS SUPPORTED BY THE CBAM ATTENTION MECHANISM. *Kahramanmaras Sutcu Imam University Journal of Engineering Sciences,28(4),* 1983-1999.

INTRODUCTION

Timely identification of eye diseases is vital in reducing the risk of lasting vision impairment and enhancing overall patient well-being. Specifically, if pathological alterations in the retina are not recognized early, they can result in irreversible loss of sight. Conditions including diabetic retinopathy, glaucoma, cataract, and age-related macular degeneration impact millions globally and often develop stealthily, presenting no significant symptoms for extended periods. Prompt and precise detection is therefore essential to slow disease progression, improve treatment outcomes, and minimize the chances of permanent visual impairment (De Fauw et al., 2018).

As reported by the World Health Organization in 2019, an estimated 2.2 billion individuals worldwide are affected by some form of visual impairment, with early intervention and treatment having the potential to prevent over half of these cases (World Health Organization, 2019). Vision loss not only reduces an individual's quality of life but also results in serious consequences in terms of social productivity, healthcare system burden, and economic losses. Figure 1 presents the percentages of vision loss caused by undetected eye diseases at early stages.

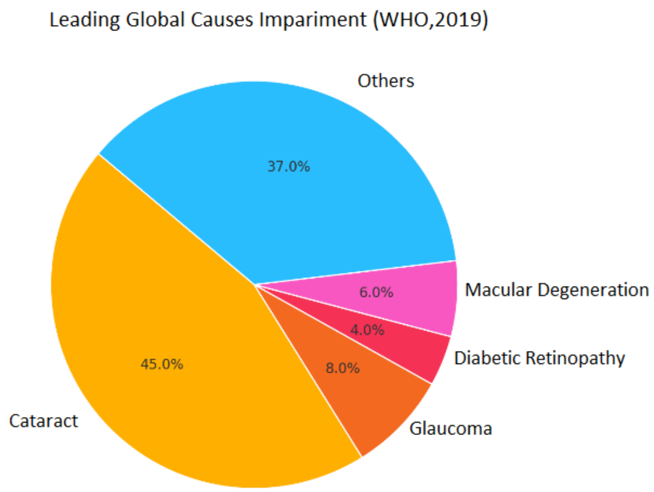

Figure 1. The Leading Causes of Vision Loss Worldwide (World Health Organization, 2019)

A significant advancement in the field of medical image analysis over the past decade is the increasing application of artificial intelligence, with deep learning techniques taking a central role. Deep learning models, which utilize multi-layered neural network architectures, are capable of independently identifying complex data structures within vast datasets. These models have achieved outstanding performance in a range of image processing tasks, including classification, segmentation, and feature detection, particularly when implemented via convolutional neural networks (CNN) (Jin and Ye, 2022). CNN architectures are capable of detecting and interpreting pathological findings in fundus images, such as minor vascular changes, microaneurysms, hemorrhages, and the optic disc, with a level of detail that exceeds the limits of human vision.

In recent years, the use of artificial intelligence and its subcomponents in medicine has significantly expanded, with applications ranging from classification to segmentation tasks. AI-assisted approaches are not only applied in ocular disease detection but have also been successfully implemented in other clinical scenarios. For example, in a study on wart treatment evaluation using optimum ensemble-based classification techniques, AI methods were shown to improve the accuracy and reliability of clinical decision support systems (Balcı and Alkan, 2024). Such studies highlight the applicability of artificial intelligence methods across different domains of medicine and emphasize the large-scale clinical potential of these approaches.

AI-assisted diagnostic systems not only offer high accuracy but also provide fast, repeatable, and objective decisions, leading to significant savings in time and cost during clinical processes (Ting et al., 2017). These systems hold considerable potential, particularly in developing countries, in regions with a shortage of specialists, and in screening programs. Recently, the use of machine learning in scientific research has become increasingly widespread, finding applications in fields such as text mining, spam filtering, video content recommendation, image analysis, and multimedia data interpretation (Al-Dulaimi et al., 2019; Amrit et al., 2017; Deldjoo et al., 2016; Hossain et al., 2019;

Rozenwald et al., 2020). Among various machine learning algorithms, deep learning is frequently used in such applications (Alom et al., 2019; Liu et al., 2017; Pouyanfar et al., 2018). Moreover, the collection and analysis of medical images in digital environments allow for the creation of large datasets, which in turn enhances model generalizability. Large-scale and publicly accessible fundus image databases such as Kaggle EyePACS, ODIR-5K, and Messidor enable researchers to work with rich datasets that represent different disease groups and populations, thereby facilitating the faster adaptation of developed AI models to real-world clinical applications (Campbell et al., 2018).

The main motivation of this study is to evaluate the performance impact of the Convolutional Block Attention Module (CBAM) on different deep learning architectures. For this purpose, CBAM integration was thoroughly investigated on DenseNet121 and EfficientNetB0 models. The experimental findings revealed that CBAM led to significant improvements in performance metrics such as accuracy and F1-score for the lower baseline model DenseNet121, while yielding smaller but consistent gains in EfficientNetB0. In contrast, CBAM did not provide notable improvements in other architectures such as ResNet and AlexNet. These results indicate that the effectiveness of attention mechanisms may be architecture-dependent and that CBAM plays a particularly important role in enhancing weaker models. The novelty of this study lies in demonstrating that CBAM's performance-boosting effect is not limited to strong models but can also provide clinically meaningful improvements in lower baseline architectures. In this respect, the study contributes a fresh perspective to the literature and offers an original contribution to increasing the reliability of AI-assisted clinical decision support systems.

To highlight the novelty of this research, the key contributions can be outlined as follows:

- In addition to performance metrics, this study also discusses the clinical implications of false positive and false negative outcomes, thereby assessing the practical applicability of AI models in healthcare.
- The integration of the Convolutional Block Attention Module (CBAM) not only provided minor improvements in strong models but also introduced significant enhancements in weaker baseline models, offering a rarely explored perspective in the literature.
- By comparing different deep learning architectures, the study reveals that the effect of CBAM is architecture-dependent and demonstrates that it can yield clinically meaningful improvements, particularly in lower-performing models.
- The proposed methodology not only demonstrates model performance but also incorporates discussions on real-world challenges such as data imbalance, limited datasets, and clinical adaptability.
- Unlike many previous works, this study emphasizes the role of attention mechanisms in enhancing the reliability of AI-assisted clinical decision support systems, highlighting their importance particularly in regions with limited healthcare resources.

The continuation of this study is structured as follows: a section providing essential information on CNN architectures and attention mechanisms; a materials and methods section that details the dataset used, the proposed methodologies, and the integration of the CBAM attention mechanism; an experimental results section presenting the training outcomes; a discussion section where the obtained findings are compared with the existing literature and the strengths and limitations of the study are highlighted; and finally, a conclusion section summarizing the results and providing insights into the limitations and potential directions for future research.

Literature Review / Benchmarking

The CBAM attention mechanism has achieved significant success across different medical imaging domains. For instance, in retinal imaging,(Vanaja and Prakasam, 2025) reported high IoU and Dice scores in microaneurysm segmentation using a U-Net framework enhanced with CBAM and Attention Gate. In neuroimaging, the study (Islam and Hossain, 2025) demonstrated improved accuracy for brain tumor segmentation on MRI images. In cardiac imaging, (Cao et al., 2025) incorporated CBAM into a Nested U-Net, achieving approximately a 1.05% increase in Dice score for left ventricular segmentation. Beyond medical applications, (Zhao et al., 2024) applied a CBAM-based ResUNet to rock thin-section image segmentation, yielding improvements in both accuracy and efficiency. These examples collectively highlight CBAM's versatility across imaging modalities and its effectiveness in improving segmentation precision and feature localization.

Previous research has consistently demonstrated the effectiveness of the CBAM attention mechanism across a variety of medical imaging tasks. For instance, (Vanaja and Prakasam, 2025) chieved superior performance in microaneurysm segmentation for diabetic retinopathy using a U-Net model enhanced with CBAM and Attention Gate. Similarly, (Huang and Prakash, 2025) compared SE, Transformer, and CBAM-based architectures for MRI tumor classification, concluding that CBAM provided a more balanced and stable performance. In another study, (Zhao et al., 2024) showed that integrating CBAM into a ResUNet architecture enhanced segmentation accuracy for microstructural image analysis. The findings of the present work are consistent with these earlier studies and further extend CBAM's applicability from segmentation tasks to the classification of retinal fundus images.

In the existing literature, numerous studies have investigated the automated classification of retinal fundus images, several of which have incorporated attention-based mechanisms. For instance, (Zhang et al., 2025) obtained 92.64% accuracy using an Improved GoogLeNet model augmented with CBAM. (Alsohemi and Dardouri, 2025) achieved 95.12% accuracy, with 95.21% precision, 94.88% recall, and an F1-score of 95.00% on the ODIR dataset for multi-disease classification. Additionally (Novely et al., 2025) reported that combining CBAM with skip connections improved the performance of pre-trained CNNs, reaching 95.18% accuracy.

In recent years, Vision Transformer (ViT) and visual attention approaches have gained significant momentum in the diagnostic analysis of retinal fundus images. For instance, (Yang et al., 2024) combined ViT with Masked Autoencoders (MAE) for referable diabetic retinopathy classification, achieving superior performance by effectively learning discriminative representations from large-scale fundus images. Similarly, (Cho et al., 2024) proposed an attention-based framework that fused multiple structural cues such as retinal vessels and optic disc features to enhance classification performance in glaucoma detection. In Scientific Reports, (Lin, 2025) introduced an efficient fusion transformer architecture that improved accuracy in intraocular disease classification, whereas (Goh et al., 2024) compared ViT and CNN models for referable diabetic retinopathy detection, emphasizing the competitive capability of ViT-based approaches. Furthermore, (Lee et al., 2025) presented an explainable Vision Transformer that successfully learned clinically meaningful biomarkers from fundus photographs through metabolic syndrome prediction, underscoring the suitability of ViTs for interpretable medical AI applications.

Our study specifically addresses the architecture-dependent nature of attention mechanisms, which has been relatively underexplored in previous research. This represents the key distinction and contribution of our work compared to existing studies.

CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES (CNN)

In the last ten years, a wide range of CNN architectures has emerged. The choice of network structure plays a key role in optimizing performance across various application domains. Since 1989, numerous modifications have been made to CNN architectures, including structural reformulations, regularization techniques, and parameter optimizations. However, the most significant advances in CNN performance have largely been achieved through the reorganization of processing units and the development of new architectural blocks. In particular, the most innovative improvements in CNN architectures have been related to the use of network depth. Starting with the AlexNet model introduced in 2012, the most popular CNN architectures have been examined. Examining features like input dimensions, network depth, and resilience allows researchers to determine which architecture best suits their specific objectives (Alzubaidi et al., 2021).

AlexNet

AlexNet has been widely acknowledged within deep CNN architectures for its notable advancements, especially in image recognition and classification domains. The initial introduction of AlexNet by Krizhevsky et al. (He et al., 2016) involved increasing the network's depth, which, together with the implementation of advanced parameter optimization methods, substantially improved the learning potential of convolutional neural networks (Alzubaidi et al., 2021).

During the development of AlexNet, the learning abilities of deep CNNs were largely restricted by available hardware. To address this, AlexNet's training process utilized parallel computing with two GPUs (NVIDIA GTX 580). Additionally, the number of feature extraction layers was expanded to broaden the range of image categories that CNNs could process. While increasing the network's depth enhanced its ability to generalize across different image resolutions, it also introduced challenges such as overfitting. To counteract this, dropout was employed during training to randomly deactivate some neurons, thus fostering more resilient feature learning. The adoption of the ReLU activation function, which does not saturate, also mitigated the vanishing gradient problem and hastened network convergence (Alzubaidi et al., 2021).

Visual Geometry Group (VGG)

Once the success of CNNs in image recognition became apparent, the Visual Geometry Group (VGG) introduced a streamlined and effective design for CNN architectures. This architecture stands out due to its significant depth, incorporating nineteen more layers than AlexNet to enhance representational power. A key feature of VGG is its use of smaller convolutional filters, which not only reduces parameter count but also simplifies the computational demands of the network. This approach set a new direction in CNN research, favoring smaller filters for efficiency. Additionally, VGG integrated 1×1 convolutions between layers to control model complexity and allow for flexible combinations of feature maps. The architecture also applies max pooling and padding to maintain spatial dimensions, leading to strong results in both localization and classification tasks (Alzubaidi et al., 2021).

ResNet

ResNet architectures vary widely in depth, with versions ranging from 34 to 1202 layers. Among them, ResNet50 is the most prevalent, featuring 49 convolutional layers and a final fully connected layer, and totaling about 25.5 million parameters and 3.9 million multiply-accumulate (MAC) operations. The distinguishing feature of ResNet is the incorporation of skip (bypass) connections, a concept previously explored in Highway Networks to tackle the training challenges of very deep networks. By embedding residual connections within the standard feedforward structure, ResNet enables the formation of multiple residual blocks whose operations may differ according to the specific architecture variant. Notably, ResNet with 152 layers secured first place at the 2015 ILSVRC, being substantially deeper than both VGG and AlexNet, yet achieving lower computational complexity than VGG (He et al., 2016; Urban et al., 2018).

DenseNet

In the DenseNet framework, each layer receives input from all preceding layers in a feed-forward sequence, enabling the transfer of feature maps from earlier layers to every subsequent layer. This connectivity allows the model to distinguish clearly between newly introduced and retained information, as DenseNet merges feature maps from prior layers by concatenation rather than addition. Nevertheless, the relatively slim architecture of DenseNet can lead to higher computational demands, stemming from both increased parameter counts and the expanding number of feature maps. The provision of direct gradient flow from the loss function to every layer promotes efficient information transmission across the network. Additionally, this architectural choice serves as a form of regularization, helping to alleviate overfitting, particularly in scenarios with limited training data (Rubin et al., 2018; Wu et al., 2018).

CONVOLUTIONAL BLOCK ATTENTION MODULE (CBAM)

Attention mechanisms are designed to prioritize the most informative areas within an image while filtering out irrelevant details. Similarly, the human visual system employs such strategies to effectively process and interpret intricate visual scenes. Inspired by this biological principle, the integration of attention modules into computer vision frameworks has led to marked improvements in overall performance. In the context of computer vision, attention mechanisms function by dynamically prioritizing features according to their relevance to the input data. The adaptive assignment of importance to different regions or channels has led to significant advancements in a wide variety of visual analysis tasks. These include, but are not limited to, classification, object localization, semantic segmentation, facial and person identification, re-identification, recognition of actions or events, few-shot learning, medical image interpretation, image generation, pose estimation, enhancing image resolution, 3D vision, and multi-modal processing (Carion et al., 2020; Dai et al., 2017; Dosovitskiy et al., 2021; Fu et al., 2019; Woo et al., 2018; Yuan et al., 2021).

Attention modules are closely linked to the processes of human perception and cognition, enabling individuals to focus on critical regions within an image to extract meaningful information (Xiao et al., 2021). Drawing on this idea, modern deep learning architectures have incorporated attention mechanisms that amplify salient features while reducing the impact of less relevant ones. One such widely utilized attention mechanism is the Convolutional Block Attention Module (CBAM). CBAM is comprised of two core components: the Channel Attention Module (CAM), which assigns importance across the channel dimension, and the Spatial Attention Module (SAM), which identifies significant spatial patterns (Woo et al., 2018; Xiao et al., 2021). The architecture of CBAM is depicted in Figure 2, where the input X is processed through Fc(X) for channel attention and Fs(X') for spatial attention. CBAM directs the network's attention toward the most relevant and distinguishing input features, thereby enhancing both classification accuracy and the model's overall performance.

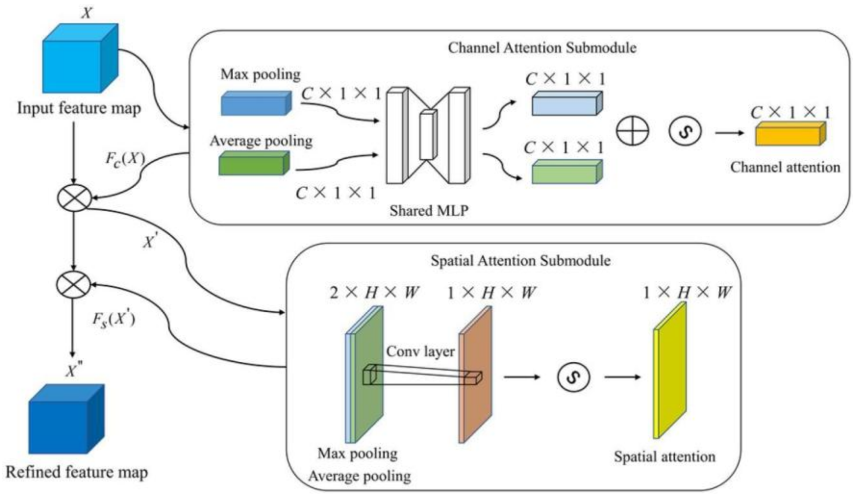

Figure 2. The Overall Structure of the CBAM Module (Ji et al., 2025)

In the Channel Attention Module (CAM), feature maps generated by the previous block are sent in parallel to both average pooling and max pooling layers. These layers transform the original feature map of dimensions C × H × W into a compressed C × 1 × 1 vector, allowing the spatial information from each channel to be effectively aggregated and resulting in more informative feature representations (Altuwaijri and Muhammad, 2022). The pooled outputs are then forwarded to a shared-weight Multi-Layer Perceptron (MLP) containing a hidden layer. The MLP applies a reduction ratio to initially shrink the channel dimension to C/r (where r = 16), before projecting it back to C, thereby reducing the overall parameter count. The results from the average and max pooling branches are combined using element-wise addition and processed through a sigmoid activation, yielding the channel attention map. Finally, the spatial attention weight coefficient (Fs) from the Spatial Attention Module is element-wise multiplied with its associated feature map to generate the final set of refined features (Ji et al., 2025).

MATERIALS AND METHODS

Eye diseases represent a significant cause of avoidable visual impairment and blindness worldwide, posing a substantial challenge to global healthcare systems. The World Health Organization reports that roughly 2.2 billion people globally experience vision impairment or loss, with early intervention able to prevent a considerable proportion of these cases. Conditions such as diabetic retinopathy, glaucoma, and cataract may lead to permanent sight loss as they gradually advance through pathological changes in the retina. Cataract is currently identified as the leading cause of blindness internationally, while diabetic retinopathy and glaucoma are also of great concern for early detection because of their progressive course and lack of early symptoms (Flaxman et al., 2017; Foster and Resnikoff, 2005).

Artificial intelligence, especially deep learning algorithms, has shown performance in medical imaging that rivals, and in many cases surpasses, that of human experts. Among deep learning architectures, convolutional neural networks (CNNs) have gained particular prominence and have demonstrated considerable success in areas like the detection of pathological features, segmentation, and categorization of retinal images. Recent large-scale investigations indicate that automated diagnostic tools built on CNN frameworks can reach accuracy rates similar to those obtained by expert ophthalmologists, particularly in identifying diseases such as diabetic retinopathy, glaucoma, and cataract (Bourne et al., 2021).

Recent studies have compared different attention mechanisms in CNN-based medical imaging models. The Squeeze-and-Excitation (SE) block introduces channel-wise recalibration by aggregating spatial information into a global descriptor, enabling the network to emphasize the most relevant feature maps. However, SE only focuses on channel dependencies and ignores spatial distributions. The Bottleneck Attention Module (BAM) incorporates both channel and spatial attention but in a parallel manner and with a more constrained structural design, limiting its adaptability in deep architectures. Transformer-based attention mechanisms, on the other hand, have shown remarkable performance in vision tasks due to their ability to capture long-range dependencies, but their high computational cost and data requirements make them less suitable for relatively smaller medical datasets. In contrast, CBAM sequentially combines channel and spatial attention, allowing the model not only to identify ‘what’ features are important but also ‘where’ they are located. This dual refinement has been demonstrated to enhance classification and segmentation performance in medical imaging, such as MRI tumor detection and retinal disease recognition tasks (Huang and Prakash, 2025). By balancing accuracy improvements with computational efficiency, CBAM provides a more practical solution for clinical deployment compared to alternative attention mechanisms.

In this study, the classification of human retinal images belonging to the normal, diabetic retinopathy, cataract, and glaucoma categories was targeted using the DenseNet and EfficientNet deep learning architectures, based on open-access databases. Afterward, the CBAM, representing one type of attention mechanism, was incorporated into these network architectures to evaluate its effect on classification outcomes. Consistent with recent advances in the field, the primary objective of this research is to establish a diagnostic system that is both highly accurate and broadly applicable, with potential for real-world clinical adoption.

The dataset used in this study consists of four classes: cataract (1038), diabetic retinopathy (1098), glaucoma (1007), and healthy (1074). The class distribution is relatively balanced and thus no severe imbalance exists. Nevertheless, considering that even minor differences in class size may affect model performance, we reported not only accuracy but also balanced metrics such as F1-score, precision, and recall. Furthermore, data augmentation techniques were applied during training to minimize the risks of overfitting or neglecting minority classes. This approach enhanced the generalization capability of the models and contributed to more reliable outcomes in clinical applications.

Dataset and Methodology

In this study, human eye diseases were detected using retinal images. The retinal images were obtained from the publicly available Kaggle Eye Diseases Classification dataset (*Eye_diseases_classification*, n.d.). The dataset contains retinal images from individuals diagnosed with cataract, diabetic retinopathy, and glaucoma, as well as from healthy subjects. It comprises a total of 1,038 cataract, 1,098 diabetic retinopathy, 1,007 glaucoma, and 1,074 healthy retinal images. Representative retinal images from each class are presented in Figure 3.

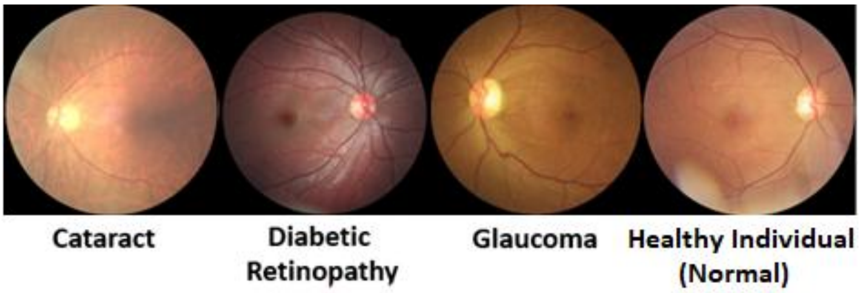

Figure 3. Retinal İmages for Each Class

Initially, the retinal images were trained using the deep learning models DenseNet and EfficientNet, and the performance of the models was observed. To enhance the system's performance, data augmentation was applied. Specifically, the images were randomly rotated by approximately 30º along the horizontal axis. This approach not only increased the number of training samples but also enabled the model to make accurate predictions for images taken from various angles. It can be stated that data augmentation had a positive impact on the training accuracy of the system. After analyzing the initial results, the Convolutional Block Attention Module (CBAM), one of the attention mechanisms, was integrated into the model to further improve performance. The flowchart of the proposed model is presented in Figure 4.

Training Results

The model training was performed on a system featuring an NVIDIA GeForce GTX 1060 GPU and 16 GB of RAM. All training procedures were executed in Python, initially employing the EfficientNetB0 model and subsequently DenseNet121. Key training parameters and hyperparameters are summarized below. For every experiment, the dataset was partitioned into 80% for training and 20% for validation. Images were uniformly resized to 256 × 256 pixels before input. A batch size of 32 was chosen, ensuring that each iteration during training and validation processed 32 images. To enhance model robustness, data augmentation was applied to the training set, incorporating random horizontal flips (RandomFlip('horizontal')) and random rotations between −0.5 and +0.5 radians (about −28.6° to +28.6°) using RandomRotation (0.5). Both EfficientNetB0 and DenseNet121 models were initialized with ImageNet pre-trained weights, and their base layers were kept non-trainable during training (base_model.trainable = False). Feature extraction was followed by a GlobalAveragePooling2D layer and a dense layer with 128 units utilizing the Leaky ReLU activation function (alpha = 0.01). The network's output layer was a Dense layer matching the number of target classes, producing logits.

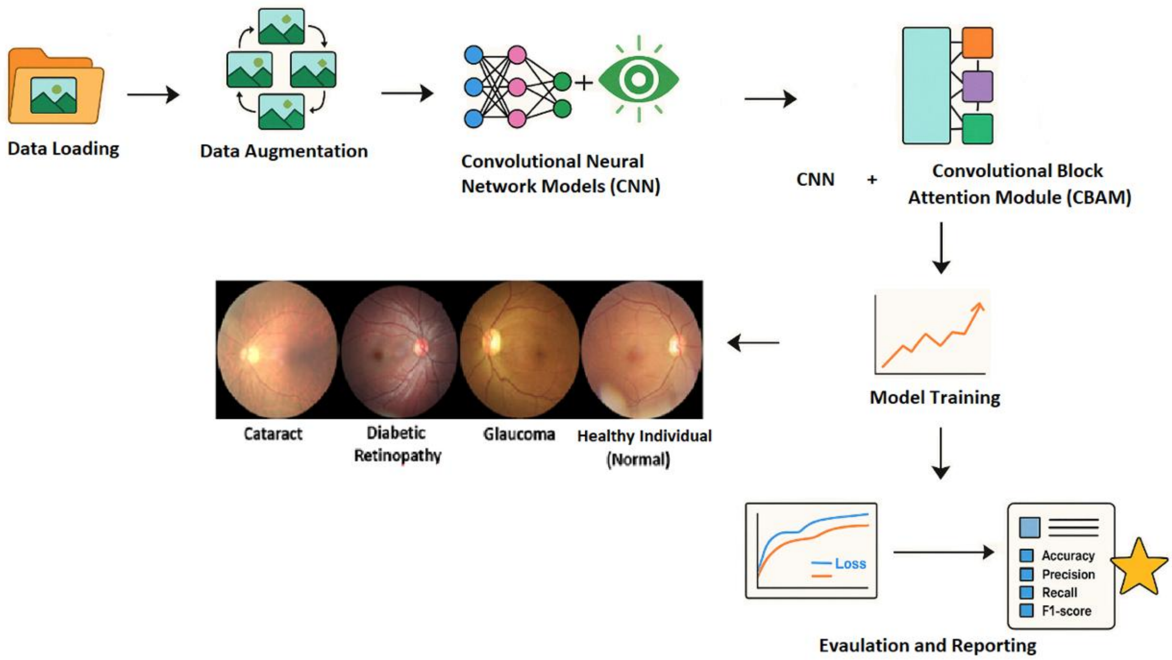

Figure 4. Flowchart of the Proposed Model

Model compilation was carried out with the Adam optimizer (learning rate = 0.001), using SparseCategoricalCrossentropy (from_logits=True) as the loss function. Overfitting was addressed with a Dropout layer set at 0.2. Training was automatically halted if validation loss or accuracy failed to improve for 10 consecutive epochs, through the EarlyStopping callback (patience = 10). The maximum number of training epochs was limited to 70, but early stopping minimized unnecessary epochs. During training, the evolution of accuracy and loss was tracked across all epochs for both training and validation sets. At the end of training, metrics such as accuracy, precision, recall, F1-score, and confusion matrix were computed using the validation data. Throughout the training process, emphasis was placed on optimal parameter selection, maintaining data variety, and implementing strategies to enhance generalizability.

The dataset was classified using the DenseNet121 model. Upon training and evaluation, the following results were obtained: accuracy = 0.8837, precision = 0.8966, recall = 0.8837, and F1-score = 0.8852. The accuracy and loss graphs corresponding to each epoch are shown in Figure 5.

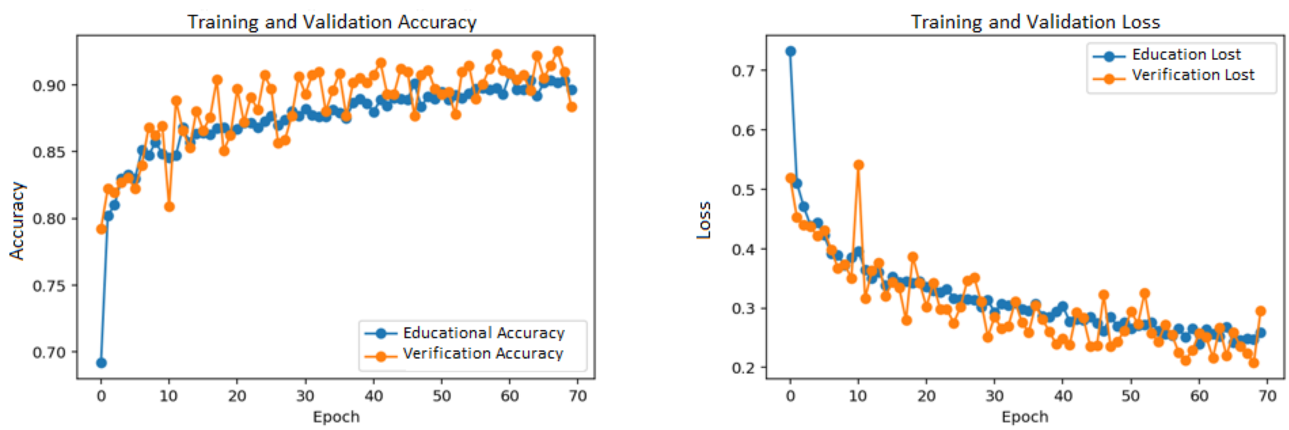

Figure 5. Accuracy and Loss Graphs of the DenseNet121 Model Training Results

To observe the classification accuracy at the end of training, the confusion matrix was obtained as shown in Figure 6.

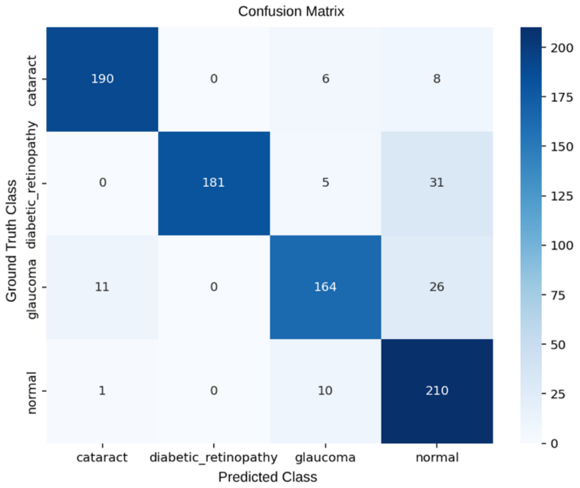

Figure 6. Confusion Matrix obtained from the Training Results of the DenseNet121 Model

The dataset was classified using the EfficientNetB0 model. Upon training and evaluation, the following results were obtained: accuracy = 0.9632, precision = 0.9634, recall = 0.9632, and F1-score = 0.9633. The accuracy and loss graphs corresponding to each epoch are presented in Figure 7.

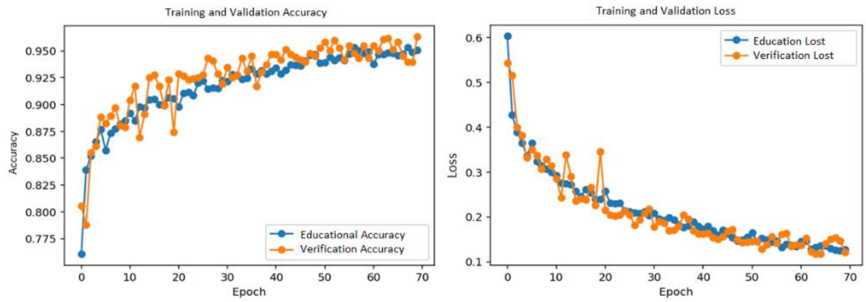

Figure 7. Accuracy and Loss Graphs of the EfficientNetB0 Model Training Results

To observe the classification accuracy at the end of training, the confusion matrix was obtained as shown in Figure 8.

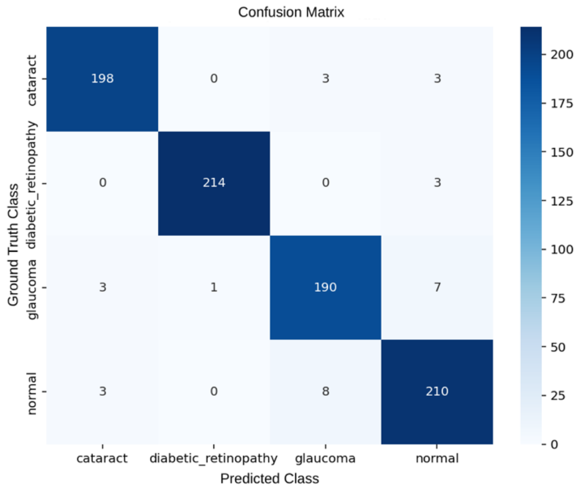

Figure 8. Confusion Matrix obtained from the Training Results of the EfficientNetB0 Model

Conventional deep learning models often analyze complex structures in medical images using local filters, which may cause them to overlook critical pathological details. Therefore, attention mechanisms such as CBAM dynamically determine which feature maps the model should focus on and which spatial regions of the image should be emphasized. This enables the model to achieve notable improvements in key performance metrics such as overall accuracy and recall, particularly when working with data that includes clinically significant and highly detailed structures like the retina. At this point in the research, the previously assessed models were further enhanced by incorporating the CBAM attention module, leading to an updated set of experimental outcomes. The dataset was classified using the DenseNet121 + CBAM model. Upon training and evaluation, the following results were obtained: accuracy = 0.9039, precision = 0.9097, recall = 0.9039, and F1-score = 0.9054. The accuracy and loss graphs of the DenseNet121 + CBAM model are presented in Figure 9.

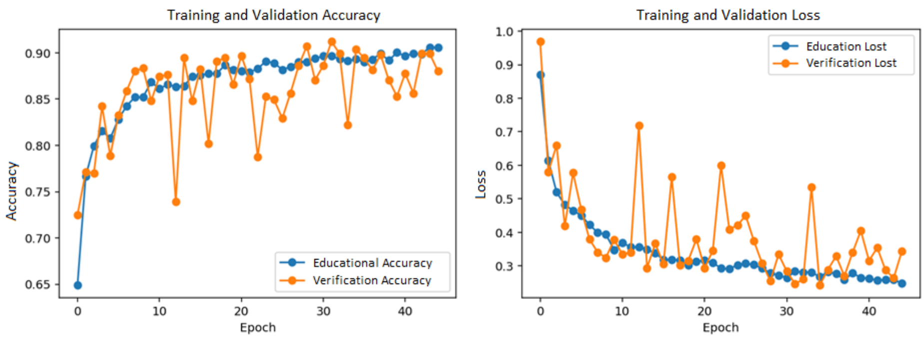

Figure 9. Accuracy and Loss Graphs of the DenseNet121+CBAM Model Training Results

To observe the classification accuracy at the end of training, the confusion matrix was obtained as shown in Figure 10.

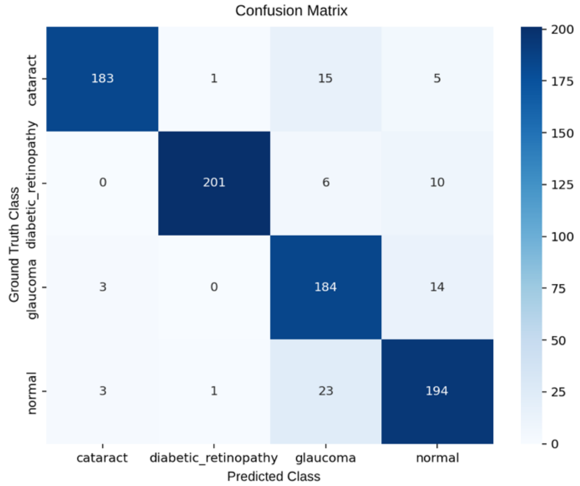

Figure 10. Confusion Matrix obtained from the Training Results of the DenseNet121+CBAM Model

The dataset was classified using the EfficientNetB0 + CBAM model. Upon training and evaluation, the following results were obtained: accuracy = 0.9656, precision = 0.9662, recall = 0.9656, and F1-score = 0.9657. The accuracy and loss graphs of the EfficientNetB0 + CBAM model are presented in Figure 11.

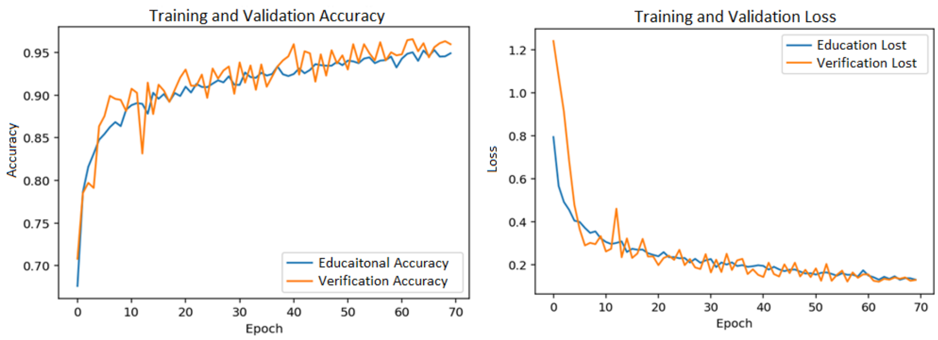

Figure 11. Accuracy and Loss Graphs of the EfficientNetB0+CBAM Model Training Results

To observe the classification accuracy at the end of training, the confusion matrix was obtained as shown in Figure 12.

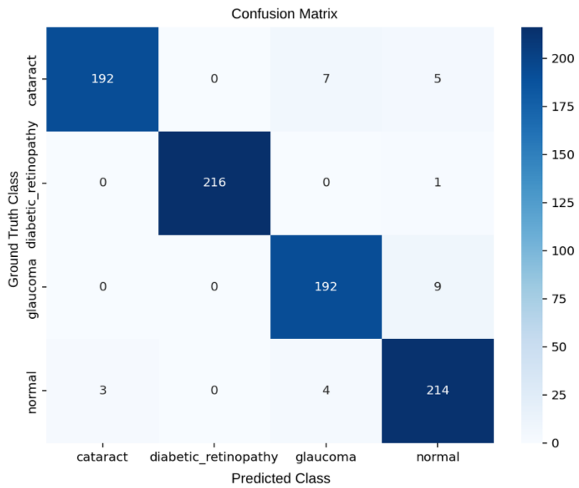

Figure 12. Confusion Matrix obtained from the Training Results of the EfficientNetB0+CBAM Model

Today, deep learning architectures, particularly in the field of medical image classification, have achieved significantly higher accuracy and generalization capabilities compared to traditional methods. However, in retinal images that contain complex and clinically critical details, it is essential for the model to effectively focus not only on specific pixels or prominent structural regions, but also on the most semantically meaningful areas. At this point, attention mechanisms play a vital role during the model training process by shifting the learning weights toward important regions and suppressing irrelevant noise, thereby making the model's decisions both more accurate and more interpretable.

Table 1. Model Performance Comparison Table

Model	Accuracy	Precision	Recall	F1 Score
DenseNet121	0.8837	0.8966	0.8837	0.8852
EfficientNetB0	0.9632	0.9634	0.9632	0.9633
DenseNet121 + CBAM	0.9039	0.9097	0.9039	0.9054
EfficientNetB0 + CBAM	0.9656	0.9662	0.9656	0.9657

Modern attention blocks, particularly CBAM, enhance the ability of deep learning models to extract meaningful information at both channel and spatial levels, thereby improving classification performance and strengthening model reliability in clinical applications. Incorporating these attention strategies into models contributes to notable improvements in both accuracy and overall system effectiveness. In Table 1, the key performance results of models with and without CBAM integration are presented in a comparative manner.

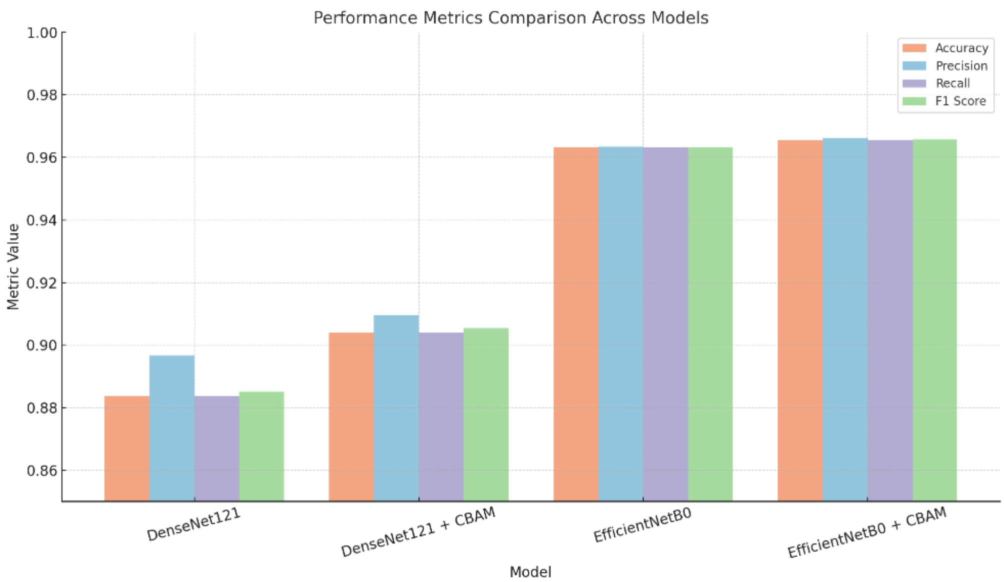

Figure 13. Performance Metrics Comparison Across Models

The training results of the models are presented comparatively in Figure 13. As observed in the figure, the CBAM attention mechanism improved performance in both the baseline (DenseNet121) and advanced (EfficientNetB0) architectures by enabling the models to focus on the most relevant regions within the visual information. The contribution of CBAM was particularly more evident in models with lower baseline accuracy, such as DenseNet121. These findings suggest that attention-based methods are especially valuable for boosting model effectiveness in demanding applications like medical image analysis.

DISCUSSION

This study investigated the effects of integrating the CBAM attention mechanism into deep learning-based ocular disease classification models. The results demonstrated that CBAM provided notable improvements, particularly in DenseNet121. The accuracy of DenseNet121 increased from 88.37% to 90.39%, and the F1-score improved from 88.52% to 90.54%. For EfficientNetB0, which already exhibited a high baseline performance, the accuracy increased from 96.3% to 96.6% and the F1-score from 96.3% to 96.6%. These findings indicate that attention mechanisms yield more substantial enhancements in models with lower baseline performance, while in high-performing architectures they provide more modest yet consistent gains.

When compared with previous studies, SE blocks have been reported to enhance channel-level representations but remain limited due to the absence of spatial contextual information (Huang and Prakash, 2025). BAM and Transformer-based attention mechanisms, while powerful in their representational capacity, often require extensive computational resources and large-scale datasets, which may restrict their feasibility in clinical environments. In this context, CBAM offers a more lightweight and flexible solution by sequentially combining channel and spatial attention.

Earlier studies have also demonstrated CBAM's effectiveness in diverse medical imaging domains. For instance, (Vanaja and Prakasam, 2025) achieved high performance in microaneurysm segmentation for diabetic retinopathy using a CBAM-enhanced U-Net with Attention Gate. (Huang and Prakash, 2025) compared SE, Transformer, and CBAM-based approaches for MRI tumor classification and reported that CBAM delivered more balanced performance. (Zhao et al., 2024) Further showed that a ResUNet-CBAM improved segmentation accuracy for microstructural image analysis. The findings of the present study align with these reports and extend CBAM's applicability to the classification of retinal fundus images.

In the literature, several studies have explored the automated classification of retinal fundus images, with some incorporating attention mechanisms. For instance, (Zhang et al., 2025) reported 92.64% accuracy using an Improved GoogLeNet with CBAM. (Alsohemi and Dardouri, 2025) achieved 95.12% accuracy, with a precision of 95.21%, a recall of 94.88%, and an F1-score of 95.00% on the ODIR dataset for multi-disease classification. (Novely et al., 2025) showed that combining CBAM with skip connections improved pre-trained CNNs, achieving 95.18% accuracy.

In our study, the DenseNet121 + CBAM combination achieved 90.39% accuracy, which is comparable to other CBAM-based approaches reported in the literature. Meanwhile, the EfficientNetB0 + CBAM achieved 96.56% accuracy, which is on par with recent EfficientNet and Transformer-based models. On the other hand, the lack of notable improvements in ResNet and AlexNet highlights the architecture-dependent effectiveness of attention mechanisms, addressing an important gap in the current literature.

The clinical implications of false negative (FN) and false positive (FP) errors are critical for assessing the reliability of AI-based diagnostic systems. In progressive diseases such as diabetic retinopathy and glaucoma, FN errors are particularly concerning, as they may lead to missed diagnoses and delayed treatment, resulting in irreversible vision loss. Such outcomes not only diminish individual quality of life but also pose significant public health challenges.

On the other hand, FP errors may cause unnecessary referrals and additional healthcare costs, while also imposing psychological stress on patients. Although their consequences are generally less severe compared to FN errors, they still contribute to inefficiencies within the healthcare system.

In this context, our findings highlight the potential of the CBAM attention mechanism in reducing FN rates. By lowering the likelihood of missed cases, CBAM integration facilitates earlier detection and timely intervention for progressive diseases, thereby enhancing the clinical reliability of AI-assisted diagnostic systems. Hence, beyond numerical accuracy metrics, considering the real-world clinical consequences of error types is essential for evaluating the true applicability and impact of deep learning models in healthcare.

In conclusion, CBAM's lightweight structure, which sequentially combines channel and spatial attention, provides notable advantages in terms of both performance improvement and computational efficiency. The findings suggest that CBAM is a promising approach for integration into clinical decision support systems, especially in resource-limited settings where rapid and accurate diagnostic processes are essential.

RESULTS

In this study, the performance of combining deep learning architectures with an attention mechanism was thoroughly analyzed for the automatic classification of retinal fundus images into cataract, diabetic retinopathy, glaucoma, and healthy subject classes. Model evaluations were conducted using four main architectures: classical DenseNet121, classical EfficientNetB0, DenseNet121 + CBAM, and EfficientNetB0 + CBAM. The performance of these models was comparatively reported using multiple metrics such as accuracy, precision, recall, and F1-score.

Looking first at the results obtained from the baseline models, the DenseNet121 model achieved an accuracy of 0.8837, precision of 0.8966, recall of 0.8837, and F1-score of 0.8852. These results indicate that while DenseNet121 provides a strong foundation for medical image classification, it still has potential for improvement in handling complex and detailed retinal features. In contrast, the EfficientNetB0 model achieved an accuracy of 0.9632, precision of 0.9634, recall of 0.9632, and F1-score of 0.9633. Due to its advanced parameter optimization and transfer learning capabilities, EfficientNetB0 demonstrated highly effective performance on the dataset.

In the second phase of the study, the CBAM attention mechanism was integrated into each model, and classification processes were repeated. The impact of the attention module on model performance was systematically evaluated. With the integration of CBAM, the DenseNet121 + CBAM model achieved an accuracy of 0.9039, precision of 0.9097, recall of 0.9039, and F1-score of 0.9054. Compared to the baseline DenseNet121 model, all metrics showed noticeable improvement after the addition of the attention module. This suggests that CBAM enhances the model's feature extraction capability, particularly in cases involving highly detailed and localized pathological findings.

The EfficientNetB0 + CBAM model achieved an accuracy of 0.9656, precision of 0.9662, recall of 0.9656, and F1-score of 0.9657. Although the baseline EfficientNetB0 already showed very high performance, the addition of CBAM resulted in a small but consistent improvement. This finding supports the notion that attention blocks can optimize decision confidence and overall accuracy, even in high-performing models, particularly when interpreting complex medical images.

Overall, the results demonstrate that attention modules such as CBAM significantly enhance the performance of deep learning-based medical image classification systems, especially in terms of recognizing complex structural details and subtle pathological findings. In particular, the integration of CBAM with DenseNet121 improved the baseline model across all key metrics, including accuracy, precision, recall, and F1-score. Similarly, the already strong performance of EfficientNetB0 was further improved with the attention mechanism. These findings highlight that incorporating attention-based modules into deep learning architectures can make a meaningful contribution to the development of clinically relevant, reliable, and high-accuracy diagnostic systems.

For the clinical applicability of the experimental findings, accuracy metrics alone are insufficient. Therefore, the clinical implications of false positive (FP) and false negative (FN) errors were analyzed. In particular, FN cases observed in diabetic retinopathy may result in missed diagnoses and delayed treatment, leading to severe consequences. Similarly, FN errors in glaucoma can increase the risk of irreversible vision loss. On the other hand, FP cases may cause patients to be unnecessarily referred to specialists, leading to additional costs and psychological stress. Hence, beyond achieving high accuracy, the clinical consequences of error types should also be considered, and the reliability of AI-based systems must be assessed within this context.

The significance of this work extends beyond the numerical improvements. By systematically examining CBAM across different architectures and demonstrating its ability to reduce false negatives, the study offers valuable evidence for the reliability of AI-based diagnostic systems in clinical practice. This is particularly important for progressive diseases such as diabetic retinopathy and glaucoma, where early detection is critical for preventing irreversible vision loss. Therefore, the study should be considered not only as a technical advancement but also as a promising step toward future clinical applications with direct benefits for public health.

Nevertheless, this work has limitations. Only a limited number of architectures were tested, the dataset may not fully represent the diversity of real-world clinical settings, and the evaluation was restricted to classification tasks. Validation on larger, multi-center datasets and exploration of additional imaging modalities (e.g., OCT combined with fundus photography) are needed to further strengthen generalizability.

Future research should focus on Vision Transformer-based models, self-supervised learning techniques, and multi-modal frameworks that integrate complementary data sources for more powerful and generalizable performance. Additionally, enhancing clinical adaptability through the development of lightweight, real-time systems for deployment on mobile devices and cloud platforms will be critical. Such innovations will enable practical integration into clinical workflows, particularly in resource-limited settings where rapid and accurate diagnosis is most needed.

In conclusion, this study highlights the value of CBAM integration and serves as a pioneering contribution that not only advances the scientific literature but also sets the stage for the development of more reliable, accessible, and scalable AI-driven solutions for ocular disease diagnosis in the future.

Artificial Intelligence Contribution Statement

This manuscript was entirely written, edited, analyzed, and prepared without the assistance of any artificial intelligence tools. All content, including text, data analysis, and figures, was solely generated by the authors.

REFERENCES

Al-Dulaimi, K., Chandran, V., Nguyen, K., Banks, J., & Tomeo-Reyes, I. (2019). Benchmarking HEp-2 specimen cells classification using linear discriminant analysis on higher order spectra features of cell shape. *Pattern Recognition Letters*, *125*, 534–541. https://doi.org/10.1016/j.patrec.2019.06.020

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. A. S., & Asari, V. K. (2019). A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics*, *8*(3), Article 3. https://doi.org/10.3390/electronics8030292

Alsohemi, R., & Dardouri, S. (2025). Fundus Image-Based Eye Disease Detection Using EfficientNetB3 Architecture. *Journal of Imaging*, *11*(8), 279. https://doi.org/10.3390/jimaging11080279

Altuwaijri, G. A., & Muhammad, G. (2022). Electroencephalogram-Based Motor Imagery Signals Classification Using a Multi-Branch Convolutional Neural Network Model with Attention Blocks. *Bioengineering*, *9*(7), Article 7. https://doi.org/10.3390/bioengineering9070323

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

Amrit, C., Paauw, T., Aly, R., & Lavric, M. (2017). Identifying child abuse through text mining and machine learning. *Expert Systems with Applications*, *88*, 402–418. https://doi.org/10.1016/j.eswa.2017.06.035

Balcı, M., & Alkan, A. (2024). Identification of wart treatment evaluation by using optimum ensemble based classification techniques. *Biomedical Signal Processing and Control*, *95*, 106491. https://doi.org/10.1016/j.bspc.2024.106491

Bourne, R., Steinmetz, J. D., Flaxman, S., Briant, P. S., Taylor, H. R., Resnikoff, S., Casson, R. J., Abdoli, A., Abu-Gharbieh, E., Afshin, A., Ahmadieh, H., Akalu, Y., Alamneh, A. A., Alemayehu, W., Alfaar, A. S., Alipour, V., Anbesu, E. W., Androudi, S., Arabloo, J., … Vos, T. (2021). Trends in prevalence of blindness and distance and near vision impairment over 30 years: An analysis for the Global Burden of Disease Study. *The Lancet Global Health*, *9*(2), e130–e143. https://doi.org/10.1016/S2214-109X(20)30425-3

Campbell, J. P., Brown, J., Chan, R. V. P., Dy, J., Ioannidis, S., Erdogmus, D., Kalpathy-Cramer, J., & Chiang, M. F. (2018). Automated diagnosis of plus disease in retinopathy of prematurity using deep learning. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *22*(4), e12. https://doi.org/10.1016/j.jaapos.2018.07.036

Cao, K., Zhao, M., Geng, M., Zheng, S., & Jung, H. (2025). Left ventricular segmentation method based on optimized UNet and improved CBAM: ESV and EDV tracking study. *PLOS ONE*, *20*(6), e0325794. https://doi.org/10.1371/journal.pone.0325794

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-End Object Detection with Transformers. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), *Computer Vision – ECCV 2020* (pp. 213–229). Springer International Publishing. https://doi.org/10.1007/978-3-030-58452-8_13

Cho, Y.-S., Song, H.-J., Han, J.-H., & Kim, Y.-S. (2024). Attention Mechanism-Based Glaucoma Classification Model Using Retinal Fundus Images. *Sensors*, *24*(14), 4684. https://doi.org/10.3390/s24144684

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., & Wei, Y. (2017). *Deformable Convolutional Networks*. 764–773. https://openaccess.thecvf.com/content_iccv_2017/html/Dai_Deformable_Convolutional_Networks_ICCV_2017_paper.html

De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., van den Driessche, G., Lakshminarayanan, B., Meyer, C., Mackinder, F., Bouton, S., Ayoub, K., Chopra, R., King, D., Karthikesalingam, A., … Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, *24*(9), 1342–1350. https://doi.org/10.1038/s41591-018-0107-6

Deldjoo, Y., Elahi, M., Cremonesi, P., Garzotto, F., Piazzolla, P., & Quadrana, M. (2016). Content-Based Video Recommendation System Based on Stylistic Visual Features. *Journal on Data Semantics*, *5*(2), 99–113. https://doi.org/10.1007/s13740-016-0060-9

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (No. arXiv:2010.11929). arXiv. https://doi.org/10.48550/arXiv.2010.11929

Eye_diseases_classification. (n.d.). Retrieved June 24, 2025, from https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification

Flaxman, S. R., Bourne, R. R. A., Resnikoff, S., Ackland, P., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., Keeffe, J., Kempen, J. H., Leasher, J., Limburg, H., Naidoo, K., Pesudovs, K., Silvester, A., Stevens, G. A., Tahhan, N., Wong, T. Y., Taylor, H. R., … Zheng, Y. (2017). Global causes of blindness and distance vision impairment 1990–2020: A systematic review and meta-analysis. *The Lancet Global Health*, *5*(12), e1221–e1234. https://doi.org/10.1016/S2214-109X(17)30393-5

Foster, A., & Resnikoff, S. (2005). The impact of Vision 2020 on global blindness. *Eye*, *19*(10), 1133–1135. https://doi.org/10.1038/sj.eye.6701973

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., & Lu, H. (2019). *Dual Attention Network for Scene Segmentation*. 3146–3154. https://openaccess.thecvf.com/content_CVPR_2019/html/Fu_Dual_Attention_Network_for_Scene_Segmentation_CVPR_2019_paper.html

Goh, J. H. L., Ang, E., Srinivasan, S., Lei, X., Loh, J., Quek, T. C., Xue, C., Xu, X., Liu, Y., Cheng, C.-Y., Rajapakse, J. C., & Tham, Y.-C. (2024). Comparative Analysis of Vision Transformers and Conventional Convolutional Neural Networks in Detecting Referable Diabetic Retinopathy. *Ophthalmology Science*, *4*(6). https://doi.org/10.1016/j.xops.2024.100552

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition*. 770–778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

Hossain, E., Khan, I., Un-Noor, F., Sikander, S. S., & Sunny, Md. S. H. (2019). Application of Big Data and Machine Learning in Smart Grid, and Associated Security Concerns: A Review. *IEEE Access*, *7*, 13960–13988. https://doi.org/10.1109/ACCESS.2019.2894819

Huang, K. A., & Prakash, N. (2025). Evaluating the Impact of Attention Mechanisms on a Fine-Tuned Neural Network for Magnetic Resonance Imaging Tumor Classification: A Comparative Analysis. *Cureus*. https://doi.org/10.7759/cureus.80872

Islam, R., & Hossain, S. (2025). Enhanced Brain Tumor Segmentation Using CBAM-Integrated Deep Learning and Area Quantification. *International Journal of Biomedical Imaging*, *2025*, 2149042. https://doi.org/10.1155/ijbi/2149042

Ji, Z., Li, S., Zhang, H., Chen, C., Xu, Q., Li, J., & Wang, H. (2025). CBAM-DeepConvNet: Convolutional Block Attention Module-Deep Convolutional Neural Network for asymmetric visual evoked potentials recognition. *Brain-Apparatus Communication: A Journal of Bacomics*, *4*(1), 2489396. https://doi.org/10.1080/27706710.2025.2489396

Jin, K., & Ye, J. (2022). Artificial intelligence and deep learning in ophthalmology: Current status and future perspectives. *Advances in Ophthalmology Practice and Research*, *2*(3), 100078. https://doi.org/10.1016/j.aopr.2022.100078

Lee, T. K., Kim, S. Y., Choi, H. J., Choe, E. K., & Sohn, K.-A. (2025). Vision transformer based interpretable metabolic syndrome classification using retinal Images. *NPJ Digital Medicine*, *8*, 205. https://doi.org/10.1038/s41746-025-01588-0

Lin, A. (2025). Efficient fusion transformer model for accurate classification of eye diseases. *Scientific Reports*, *15*(1), 36223. https://doi.org/10.1038/s41598-025-20226-z

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, *234*, 11–26. https://doi.org/10.1016/j.neucom.2016.12.038

Novely, N., Mahmud Shuvo, S., & Faruk, Md. F. (2025). Improving Pre-Trained CNNs with CBAM and Skip Connections for Multi-Class Retinal Diseases Classification using OCT Images. *Proceedings of the 3rd International Conference on Computing Advancements*, 946–953. https://doi.org/10.1145/3723178.3723304

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., Shyu, M.-L., Chen, S.-C., & Iyengar, S. S. (2018). A Survey on Deep Learning: Algorithms, Techniques, and Applications. *ACM Comput. Surv.*, *51*(5), 92:1-92:36. https://doi.org/10.1145/3234150

Rozenwald, M. B., Galitsyna, A. A., Sapunov, G. V., Khrameeva, E. E., & Gelfand, M. S. (2020). A machine learning framework for the prediction of chromatin folding in *Drosophila* using epigenetic features. *PeerJ Computer Science*, *6*, e307. https://doi.org/10.7717/peerj-cs.307

Rubin, J., Parvaneh, S., Rahman, A., Conroy, B., & Babaeizadeh, S. (2018). Densely connected convolutional networks for detection of atrial fibrillation from short single-lead ECG recordings. *Journal of Electrocardiology*, *51*(6, Supplement), S18–S21. https://doi.org/10.1016/j.jelectrocard.2018.08.008

Ting, D. S. W., Cheung, C. Y.-L., Lim, G., Tan, G. S. W., Quang, N. D., Gan, A., Hamzah, H., Garcia-Franco, R., San Yeo, I. Y., Lee, S. Y., Wong, E. Y. M., Sabanayagam, C., Baskaran, M., Ibrahim, F., Tan, N. C., Finkelstein, E. A., Lamoureux, E. L., Wong, I. Y., Bressler, N. M., … Wong, T. Y. (2017). Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA*, *318*(22), 2211–2223. https://doi.org/10.1001/jama.2017.18152

Urban, G., Subrahmanya, N., & Baldi, P. (2018). Inner and Outer Recursive Neural Networks for Chemoinformatics Applications. *Journal of Chemical Information and Modeling*, *58*(2), 207–211. https://doi.org/10.1021/acs.jcim.7b00384

Vanaja, C. B., & Prakasam, P. (2025). Convolutional block attention gate-based Unet framework for microaneurysm segmentation using retinal fundus images. *BMC Medical Imaging*, *25*(1), 83. https://doi.org/10.1186/s12880-025-01625-0

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). *CBAM: Convolutional Block Attention Module*. 3–19. https://openaccess.thecvf.com/content_ECCV_2018/html/Sanghyun_Woo_Convolutional_Block_Attention_ECCV_2018_paper.html

World Health Organization. (2019). *World report on vision*. World Health Organization. https://iris.who.int/handle/10665/328717

Wu, S., Zhong, S., & Liu, Y. (2018). Deep residual learning for image steganalysis. *Multimedia Tools and Applications*, *77*(9), 10437–10453. https://doi.org/10.1007/s11042-017-4440-4

Xiao, X., Xu, M., Han, J., Yin, E., Liu, S., Zhang, X., Jung, T.-P., & Ming, D. (2021). Enhancement for P300-speller classification using multi-window discriminative canonical pattern matching. *Journal of Neural Engineering*, *18*(4), 046079. https://doi.org/10.1088/1741-2552/ac028b

Yang, Y., Cai, Z., Qiu, S., & Xu, P. (2024). Vision transformer with masked autoencoders for referable diabetic retinopathy classification based on large-size retina image. *PLOS ONE*, *19*(3), e0299265. https://doi.org/10.1371/journal.pone.0299265

Yuan, Y., Huang, L., Guo, J., Zhang, C., Chen, X., & Wang, J. (2021). *OCNet: Object Context Network for Scene Parsing* (No. arXiv:1809.00916). arXiv. https://doi.org/10.48550/arXiv.1809.00916

Zhang, B., He, W., Wang, R., Lin, J., & Kuang, H. (2025). Improved GoogLeNet based on CBAM module for fundus disease classification network. *Second International Conference on Big Data, Computational Intelligence, and Applications (BDCIA 2024)*, *13550*, 511–517. https://doi.org/10.1117/12.3059839

Zhao, L., Zhang, H., Sun, X., Ouyang, Z., Xu, C., & Qin, X. (2024). Application of ResUNet-CBAM in Thin-Section Image Segmentation of Rocks. *Information*, *15*(12), 788. https://doi.org/10.3390/info15120788