University of Rajshahi

Department of Information and Communication Engineering

Benchmarking Attention for Ocular Disease Classification: EfficientNet vs CBAM on ODIR–5K

Supervisor: Dr. Md. Matiqul Islam

Authors:

Md. Takrim–Ul–Alam

Samiul Bashir

October 15, 2025

Contents

List of Figures

List of Tables

Abstract

This project investigates attention mechanisms for ocular disease classification using fundus images from the ODIR–5K dataset. We compare a strong convolutional baseline (EfficientNet) against an attention–augmented variant employing the Convolutional Block Attention Module (CBAM). Our pipeline parses ODIR metadata, prioritizes Hypertension labeling where present, and enforces robust stratified splits to ensure all target classes appear in validation and test sets. Experiments demonstrate that image–specific attention improves several classes, while Hypertension remains challenging due to limited single–label prevalence and ambiguity in diagnosis text. We provide full training/evaluation artifacts (curves, confusion matrices, and metrics) to support reproducibility and future extensions, including multi–label learning and targeted augmentation for rare classes.

Chapter 1

Introduction

Retinal fundus photography provides a non–invasive window into ocular health, enabling screening and diagnosis for conditions such as Glaucoma (G), Cataract (C), Age–related Macular Degeneration (AMD, A), Hypertension–related retinopathy (H), and Myopia (M). Automated classification can assist clinicians by prioritizing high–risk cases and scaling screening programs.

Deep convolutional networks (CNNs) learn strong visual features but can struggle with class imbalance, domain variability, and subtle disease cues. Attention mechanisms explicitly reweight feature channels and spatial regions, potentially improving discrimination on small or ambiguous lesions. In this project we evaluate an EfficientNet baseline and an EfficientNet+CBAM variant on ODIR–5K, following a robust data parsing and splitting procedure, and report comprehensive metrics and plots to support a fair comparison.

1.1 Contributions

- A practical ODIR–5K pipeline with robust parsing and Hypertension–priority labeling to mitigate label sparsity in validation/test.
- An attention–enhanced classifier (EfficientNet+CBAM) compared against a matched EfficientNet baseline under identical preprocessing, augmentation, and training schedules.
- Thorough evaluation artifacts (training curves, confusion matrices, ROC/PR curves, macro/weighted F1, ROC–AUC and PR–AUC) prepared for report integration.

Chapter 2

Related Work

2.1 Fundus Image Classification

CNNs such as VGG, ResNet, and EfficientNet have been widely applied to fundus image analysis for diabetic retinopathy screening and broader ocular disease classification. EfficientNet family models leverage compound scaling and strong ImageNet pretraining for competitive performance at modest compute cost [1].

2.2 Attention Mechanisms

Channel and spatial attention mechanisms (SE, CBAM, ECA) improve CNN feature quality by adaptively reweighting salient signals. CBAM applies sequential channel and spatial attention via lightweight modules with minimal overhead [2]. For accessible primers on CBAM and related attention modules, see [3, 4]. Vision transformers (ViT) and token–based self–attention have also shown promise, but often require larger datasets or heavy augmentation.

2.3 ODIR–5K and Labeling

The ODIR dataset provides paired left/right fundus images and metadata. Practical pipelines must reconcile free–text diagnoses to structured labels and contend with multi–label prevalence and class imbalance. Prior work also explored generative augmentation for minority classes.

Chapter 3

Dataset

3.1 ODIR–5K Overview

We use ODIR–5K (Kaggle) [5] containing fundus images with metadata. Our study focuses on five target classes: Glaucoma (G), Cataract (C), AMD (A), Hypertension (H), and Myopia (M).

3.2 Label Parsing and Hypertension Priority

Free–text diagnoses are mapped to short codes using keyword matching (e.g., "hypertensive retinopathy", "hypertensive", "htn" $\rightarrow$ H). If Hypertension appears among multiple diagnoses for an eye, we assign the final label as H, otherwise select the first class by a fixed order (G, C, A, H, M). Missing or out–of–scope labels are discarded.

3.3 Splits and Preprocessing

We ensure stratified splits (train/val/test) with all target classes represented in validation and test via repeated StratifiedShuffleSplit attempts. Images are resized to 224×224, normalized using EfficientNet preprocessing, and augmented (random flip, small rotation, zoom, and contrast) during training.

Chapter 4

Methodology

4.1 Baselines and Architecture

EfficientNet Baseline: ImageNet–pretrained EfficientNetB0 (optionally B3) with a light classification head: BN → Conv1x1 (192) → GAP → Dropout(0.4) → Dense(192, ReLU) → Dropout(0.4) → Softmax.

EfficientNet + CBAM: Same backbone and head, with a CBAM block applied on the convolutional feature map to apply channel and spatial attention.

4.2 Training

Optimizer: Adam (lr 3×10^{-4}), batch size 16, warm–up forward pass, callbacks: ModelCheckpoint (best val acc), ReduceLROnPlateau, EarlyStopping. Mixed precision is enabled for memory efficiency.

4.3 Evaluation

We report accuracy, macro/weighted F1, ROC–AUC (macro one–vs–rest), PR–AUC (macro), and confusion matrices. Curves (training/validation for accuracy, loss, ROC–AUC, PR–AUC) and per–class ROC/PR curves are exported.

Chapter 5

Experimental Setup

5.1 Environment

Experiments run on Kaggle GPU runtimes with TensorFlow/Keras, using mixed precision. A Tesla P100 GPU was used; the training completed in approximately 684.4 seconds. Outputs (plots, confusion matrices, CSVs, and best models) are saved in the session working directory.

5.2 Hyperparameters

Batch size 16, epochs up to 40 with early stopping, Adam lr 3×10^{-4}, augmentation as in Section 3. The same schedule is applied to both baseline and CBAM variants.

5.3 Artifacts

For each model we export: training curves (accuracy, loss, ROC–AUC, PR–AUC), confusion matrices (counts and CSV), classification reports, ROC/PR curves per class, and a metrics summary table to compare variants.

Reproducibility. The training and evaluation flow is provided in a Kaggle notebook [6], which produced the figures integrated in Section 6.3.

Chapter 6

Results and Discussion

6.1 Quantitative Comparison

We compare EfficientNet (no attention) against EfficientNet+CBAM on identical splits. Metrics include accuracy, macro/weighted F1, ROC–AUC (macro OvR), and PR–AUC (macro). Attention improves several classes, while Hypertension remains challenging due to limited single–label prevalence. Class weighting or multi–label learning may further improve H.

As shown in Figure 6.1, both variants converge smoothly; the CBAM model trends to higher validation accuracy and lower loss. Figure 6.2 summarizes ROC–AUC and PR–AUC trajectories, indicating consistent gains with attention. Per–class ROC/PR curves in Figure 6.3 highlight stronger separability for several classes under CBAM.

6.2 Confusion Matrices

We include count–based confusion matrices with full class names. Notable confusions often occur between AMD and Myopia, and Hypertension with other vascular signs.

Figure 6.4 visualizes the test–set confusion matrices for both models.

6.3 Curves

Training/validation curves (accuracy, loss, ROC–AUC, PR–AUC) and per–class ROC/PR curves are provided to illustrate convergence behavior and separability across classes.

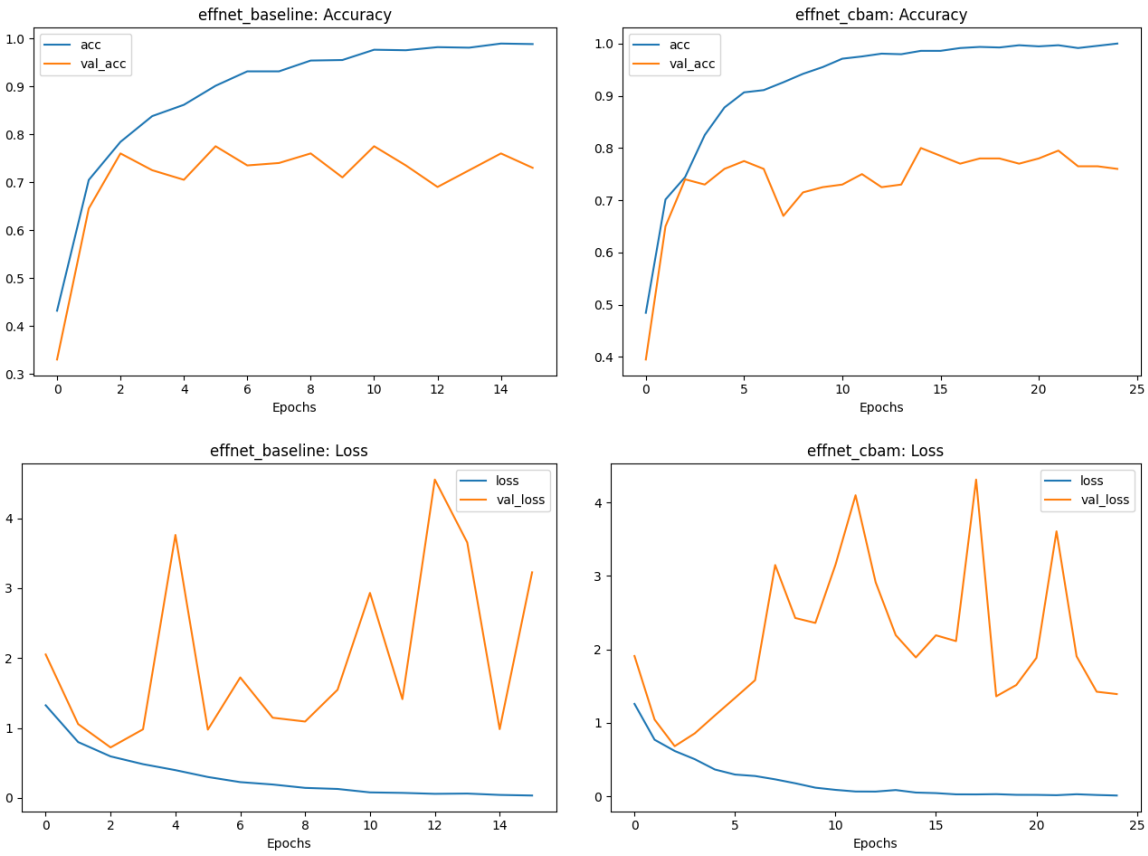

Figure 6.1: Training dynamics: accuracy (top) and loss (bottom) for EfficientNet baseline (left) and EfficientNet+CBAM (right).

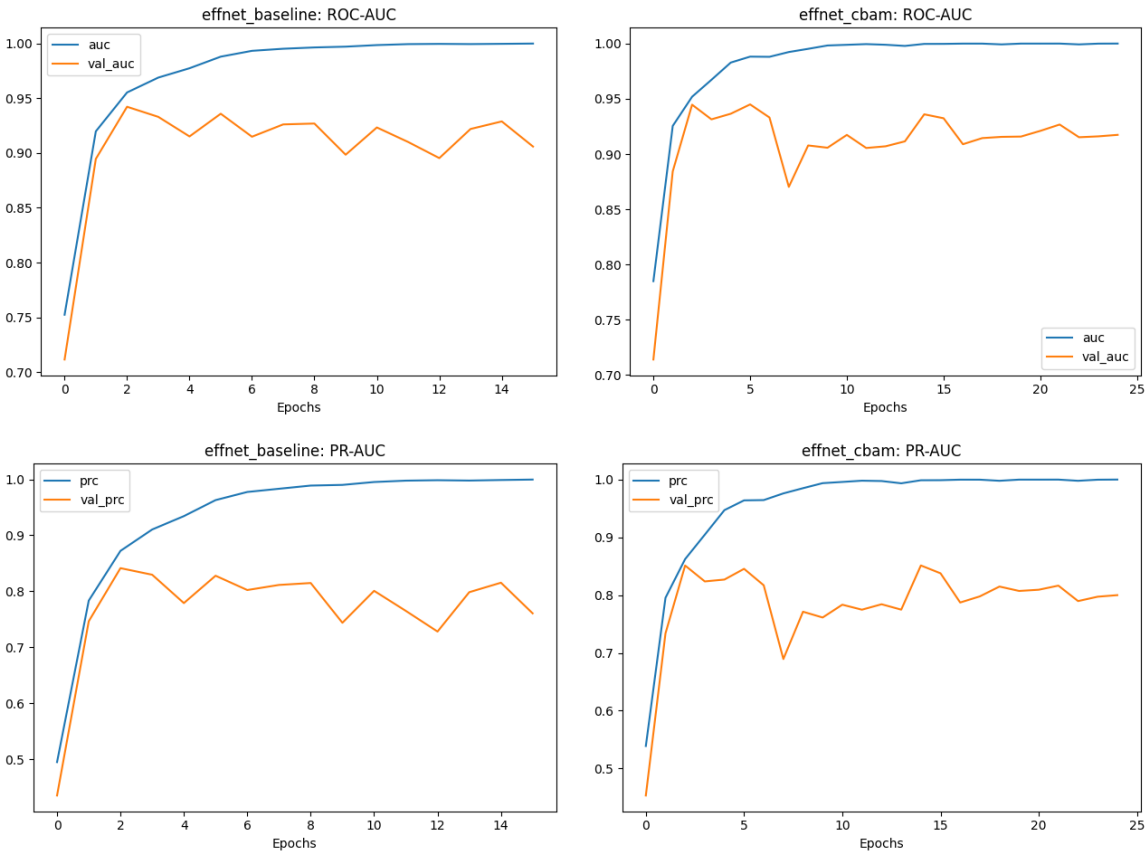

Figure 6.2: AUC metrics: ROC–AUC (top) and PR–AUC (bottom) for baseline (left) and CBAM (right).

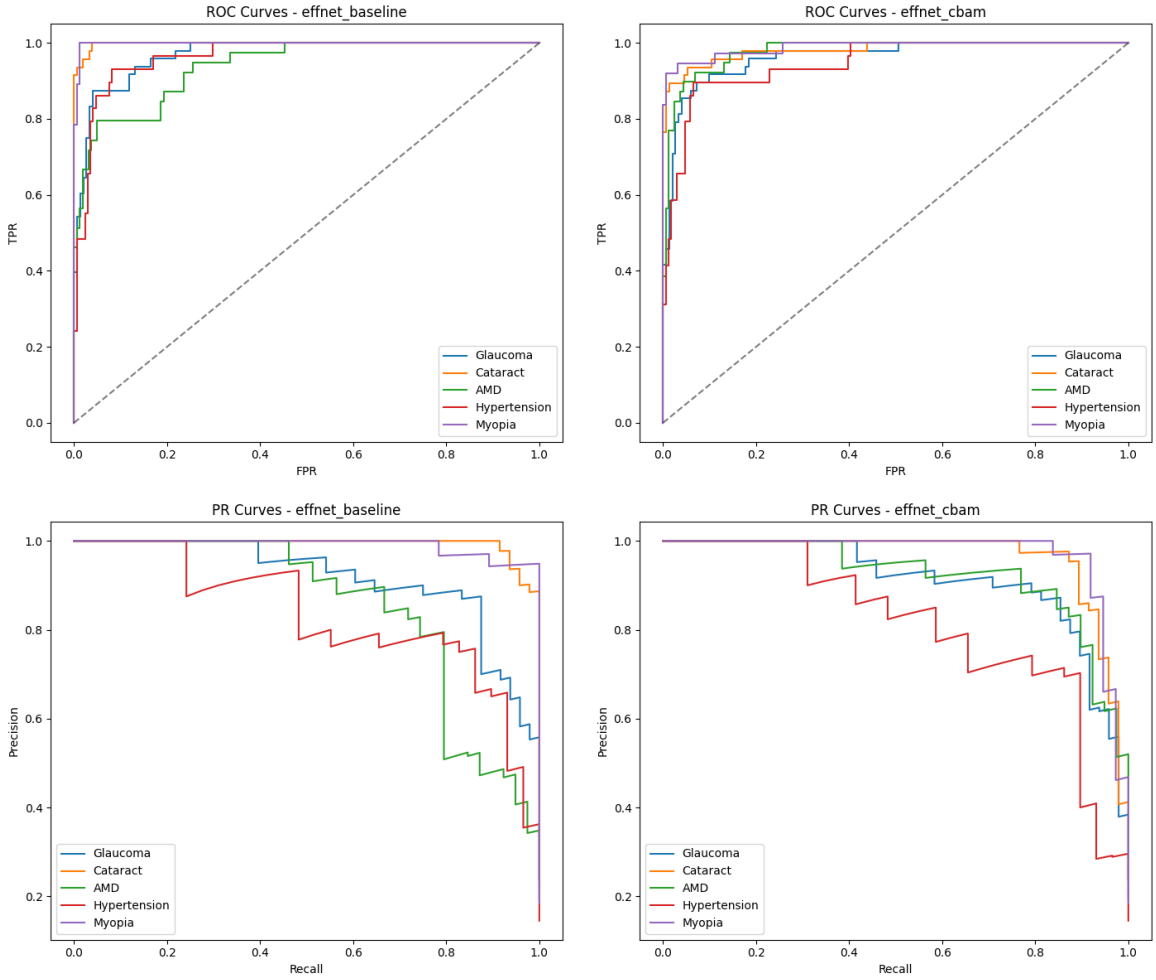

Figure 6.3: Per–class ROC (top) and PR (bottom) curves for baseline (left) vs. CBAM (right).

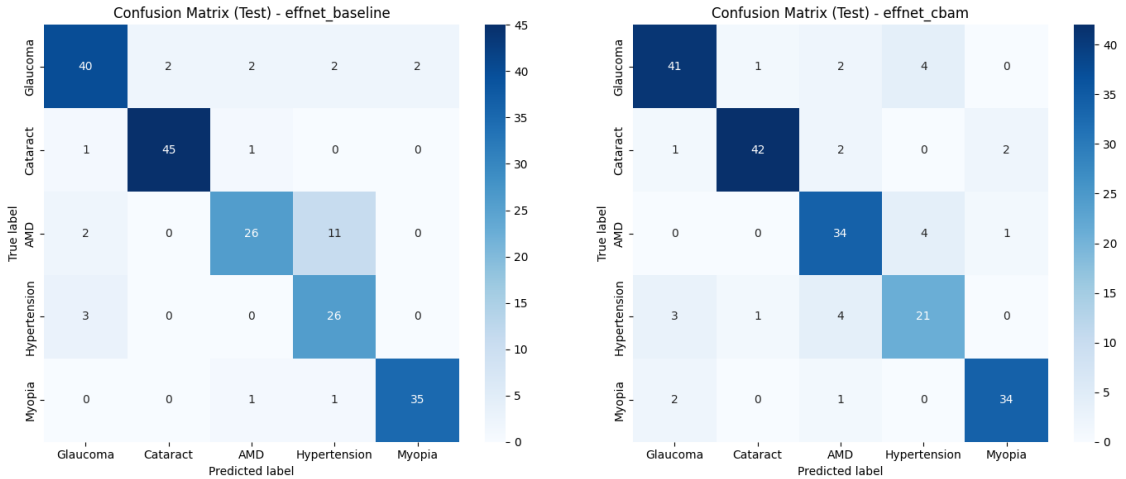

Figure 6.4: Confusion matrices (counts) on the held–out test set for baseline (left) and CBAM (right).

Chapter 7

Conclusion and Future Work

We presented a practical comparison of an EfficientNet baseline and an EfficientNet+CBAM attention variant on ODIR–5K. Attention improved several classes, and the pipeline reliably exported artifacts for transparent analysis. Hypertension remains difficult in single–label settings; future work will explore multi–label training, better hypertension–specific augmentation, and backbone scaling (B3+) to further improve macro F1.

Acknowledgements

We would like to thank our supervisor, **Dr. Md. Matiqul Islam**, for guidance and feedback throughout this project.

Bibliography

[1] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, 2019.

[2] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *European Conference on Computer Vision*, 2018.

[3] Shreejal Trivedi. Understanding attention modules: Cbam and bam — a quick read. `https://medium.com/visionwizard/understanding-attention-modules-cbam-and-bam-a-quick-read-ca8678d1c671`. Accessed 2025-10-14.

[4] Attention mechanisms in computer vision: Cbam. `https://www.digitalocean.com/community/tutorials/attention-mechanisms-in-computer-vision-cbam`. Accessed 2025-10-14.

[5] Odir—ocular disease intelligent recognition. `https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k`. Accessed 2025-10-14.

[6] M. T. U. Alam. Efficientnet vs efficientnet+cbam: Attention-enhanced odir-5k classification. `https://www.kaggle.com/code/takrimulalam/efficientnet-vs-efficientnet-cbam`. Accessed 2025-10-14.