Automated Diabetic Retinopathy Classification using Deep Learning

Pravin M (23BCE0935)
Mohit Chowdhary (23BCE2049)
GitHub: RetinoNet

***Abstract*—Diabetic Retinopathy (DR) is a leading cause of preventable blindness worldwide, particularly affecting working-age adults. This project builds an automated DR grading system using Convolutional Neural Networks (CNNs) with transfer learning. The model classifies retinal fundus images into five severity grades. A ResNet50/EfficientNet backbone was adopted, with class imbalance handling and Quadratic Weighted Kappa (QWK) as the primary metric. The Phase I results demonstrate promising performance with an accuracy of 71.6% and a QWK of 0.856.**

I. Introduction

Diabetic Retinopathy (DR) affects more than 93 million people worldwide [1]. It is a progressive eye disease that, if left untreated, may cause irreversible blindness. Early screening and grading are crucial for timely treatment, but manual grading requires ophthalmologists, making it costly and inconsistent. Automated grading systems can improve accessibility, reduce workload, and provide reliable second opinions.

This work uses transfer learning with CNN backbones to classify DR into five grades: No DR, Mild, Moderate, Severe, and Proliferative. The aim is not only to improve raw accuracy but also to handle the ordinal nature of the task where misclassifying between adjacent grades is less severe than misclassifying between extreme grades.

A. Key Modifications Based on Feedback

- Retained ResNet50 and EfficientNet backbones for feature extraction.
- Incorporated class imbalance handling via oversampling and weighted loss.
- Adopted Quadratic Weighted Kappa (QWK) as a key evaluation metric.
- Planned hybrid CNN–Vision Transformer (ViT) exploration for future phases.

II. Literature Review

Deep learning has been applied extensively to medical imaging tasks. CNN architectures such as ResNet, Inception, and EfficientNet have demonstrated strong performance for DR grading. EfficientNet, in particular, achieves a balance of accuracy and computational cost by compound scaling.

A. Ordinal Nature of DR

Unlike standard classification, DR grading is ordinal. Misclassifying "Mild" as "Moderate" is clinically less severe than misclassifying "Mild" as "Proliferative." Metrics such as QWK better capture this distinction compared to accuracy alone.

B. Hybrid Architectures

Recent studies suggest that Vision Transformers (ViTs) can capture global retinal context, while CNNs capture local lesion details. Hybrid CNN–ViT pipelines are increasingly being investigated for retinal disease classification.

III. System Design

Pipeline:

Input fundus images
$\downarrow$
Preprocessing (resize, normalize, augment, balance)
$\downarrow$
CNN Backbone (ResNet50 / EfficientNet)
$\downarrow$
Classification (Dense + Softmax)
$\downarrow$
Output DR Grade (0–4)

Modules include:

- **Preprocessing:** Image resizing (384×384), normalization, and augmentation.
- **Feature Extraction:** ResNet50/EfficientNet pretrained on ImageNet.
- **Classification:** Dense layers with softmax activation.
- **Evaluation:** Accuracy, F1, and QWK metrics.

IV. Dataset & Methodology

We used $\sim$5,500 images sampled from EyePACS and APTOS datasets. Images were normalized and augmented with random flips, rotations, and brightness adjustments to improve robustness.

A. Training Strategy

- Stage 1: Head training for 2 epochs (classifier only).
- Stage 2: Fine-tuning entire network for 15 epochs.
- Optimizer: Adam with learning rate scheduling.

V. Results (Phase I)

- Accuracy: 71.6%
- Macro F1: 0.64
- Weighted F1: 0.74
- QWK: 0.856

TABLE I
CLASS-WISE PERFORMANCE

Class	Precision	Recall	F1
0 (No DR)	0.96	0.79	0.87
1 (Mild)	0.36	0.74	0.49
2 (Moderate)	0.82	0.55	0.65
3 (Severe)	0.41	0.85	0.55
4 (Prolif.)	0.68	0.64	0.66

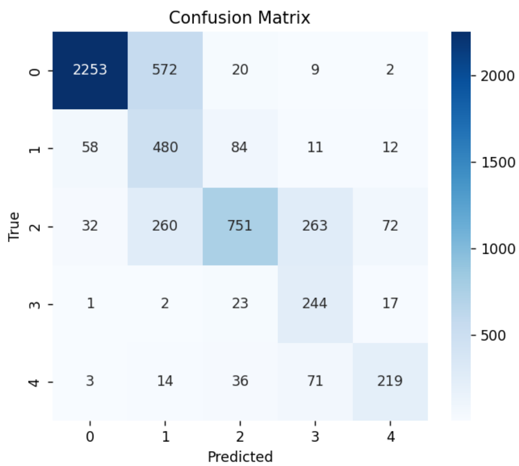

Fig. 1. Confusion Matrix highlighting adjacent misclassifications.

VI. DISCUSSION

The model performed strongly on “No DR” and moderately well on higher grades, but struggled with “Mild DR.” This is attributed to dataset imbalance and subtle visual differences between early grades. Incorporating ordinal-aware losses could reduce adjacent misclassifications.

VII. CONCLUSION & FUTURE WORK

This phase achieved 71.6% accuracy and a QWK of 0.856, demonstrating the potential of transfer learning for DR grading.

Future directions include:

- Implementing ordinal-aware loss functions.
- Exploring hybrid CNN–ViT models.
- External validation on hospital datasets.
- Addressing dataset imbalance through augmentation and generative models.

REFERENCES

[1] World Health Organization, “Global report on diabetes,” 2016.
[2] EyePACS dataset, https://www.kaggle.com/c/diabetic-retinopathy-detection
[3] APTOS 2019 dataset, https://www.kaggle.com/c/aptos2019-blindness-detection
[4] Tan, M. and Le, Q., “EfficientNet: Rethinking Model Scaling for CNNs,” ICML 2019.