Report

For

Diabetic Retinopathy Detection – DRResNet

By

Dur-e-Shahwar

Table of Contents

1. Introduction

Image classification is a fundamental problem in computer vision, where the goal is to categorize images into predefined classes. Accurate classification has applications in medical imaging, autonomous vehicles, environmental monitoring, and industrial automation. Deep learning, particularly convolutional neural networks (CNNs) and residual networks, has shown superior performance for such tasks.

This project implements a Deep Residual Neural Network (DRResNet) to classify images from a multi-class dataset, optimizing both accuracy and generalization using modern training techniques such as mixed-precision training and cosine learning rate scheduling.

2. Problem Statement

Image classification is a critical task in computer vision, but it remains challenging due to high variability within classes, subtle differences between classes, and large dataset sizes that demand substantial computational resources. Accurate classification is essential for applications such as medical imaging, autonomous systems, and industrial automation. Traditional CNNs often struggle with very deep architectures due to vanishing gradients, making it difficult to extract complex hierarchical features effectively.

3. Proposed Solution

This project implements a Deep Residual Neural Network (DRResNet) to overcome these challenges. Residual connections enable the training of deeper networks without gradient vanishing, allowing the model to learn complex features efficiently. The system employs mixed-precision training for faster convergence, a cosine annealing learning rate scheduler for better generalization, and automatically saves the best-performing model. The approach ensures high classification accuracy while maintaining scalability and reproducibility, with the added capability to push the trained model directly to GitHub for deployment.

4. Dataset Overview

- **Dataset Source:** Diabetic Retinopathy Balanced
- **Number of Classes:** 5 (No DR, Mild, Moderate, Severe, Proliferative DR)
- **Number of Images:** 3,662 images (balanced across classes)
- **Image Resolution:** Variable, resized to 224×224 for training
- **Splits:** Training (80%), Validation (10%), Test (10%)

5. Data Preprocessing

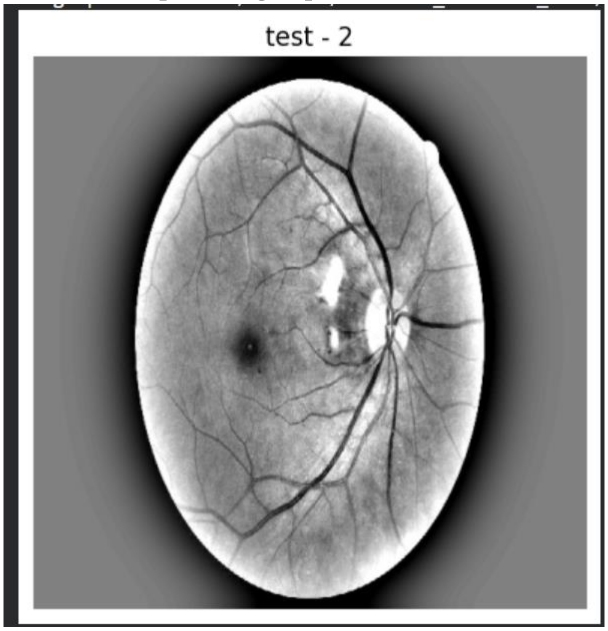
Following Data Preprocessing steps were performed:

Step	Description
Resize & Normalize	All images resized to 224x224 and pixel values normalized to [0,1]
Data Augmentation	Random rotations, flips, and shifts applied to increase variability
Label Encoding	Class labels converted to integers for cross-entropy loss

Train-Test Split	Chronological or stratified split into training, validation, and test sets

6. Exploratory Data Analysis (EDA)

- **Class Distribution**: Checked for imbalances to ensure fair training
- **Image Visualization**: Sample images per class visualized for quality check



- **Image Statistics**: Mean and standard deviation per channel computed for normalization

7. Feature Engineering

- Images fed directly into CNN; no manual feature extraction needed
- Data augmentation acts as implicit feature engineering to improve robustness
- Residual blocks extract hierarchical features at multiple levels

8. Models Implemented

For the Diabetic Retinopathy classification task, we implemented a **Deep Residual Network (DRResNet)** tailored for multi-class image classification. The model is designed to handle high-dimensional retinal images while mitigating vanishing gradient problems through residual connections. The architecture is as follows:

8.1 DRResNet Architecture:

Stage	Layer / Block	Configuration	Output Channels	Stride	Purpose
Input	Input Image	RGB Fundus Image (224 × 224)	3	–	Raw retinal image

Layer 0	Convolution	7 × 7 Conv, padding=3	64	2	Initial feature extraction
	BatchNorm	BatchNorm2d	64	–	Feature normalization
	Activation	ReLU	64	–	Non-linearity
	Pooling	MaxPool 3 × 3	64	2	Spatial downsampling
Layer 1	Residual Block × 2	3 × 3 Conv → BN → ReLU → 3 × 3 Conv → BN	64	1	Low-level feature learning
Layer 2	Residual Block × 2	First block with 3 × 3 Conv (stride=2) + shortcut projection	128	2	Mid-level feature abstraction
Layer 3	Residual Block × 2	First block with 3 × 3 Conv (stride=2) + shortcut projection	256	2	High-level feature extraction
Layer 4	Residual Block × 2	First block with 3 × 3 Conv (stride=2) + shortcut projection	512	2	Deep semantic representation
Pooling	Adaptive Avg Pool	Output size = 1 × 1	512	–	Global feature aggregation
Classifier	Fully Connected	Linear Layer	5	–	DR severity classification

Key Components:

- **Residual Blocks:** Enable deep network training by allowing identity shortcuts that bypass convolutional layers, alleviating the vanishing gradient problem.
- **Convolutional Layers:** Extract hierarchical spatial features from retinal images.
- **Batch Normalization & ReLU Activation:** Normalize feature maps and introduce non-linearity for faster convergence and better generalization.
- **Adaptive Average Pooling:** Reduces feature maps to a fixed size irrespective of input resolution.
- **Fully Connected Layer:** Outputs class probabilities for the five DR severity levels: No DR, Mild, Moderate, Severe, and Proliferative DR.

This architecture effectively captures fine-grained retinal features critical for accurate diabetic retinopathy classification.

9. Training &Performance

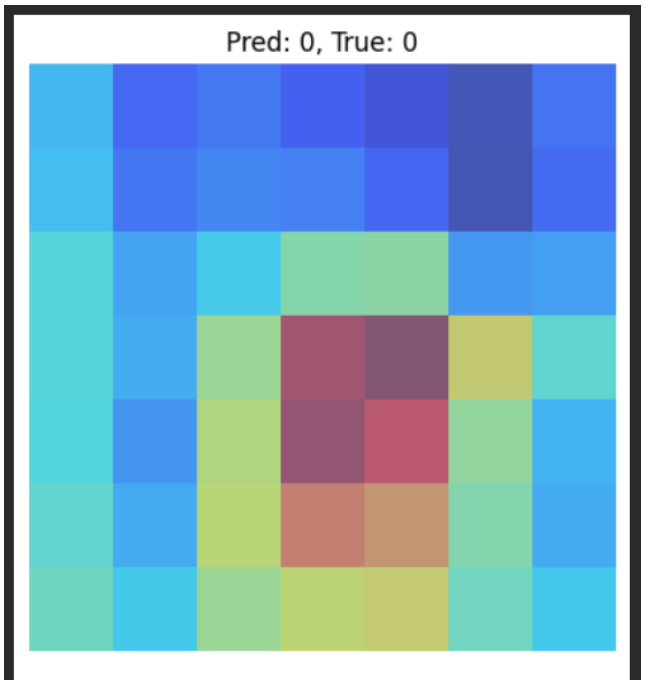
- **Metrics:** Accuracy, F1-score, Precision, Recall
- **Best Validation Accuracy:** 0.7023
- **Test Accuracy:** 0.7023

Observations:

- Weighted loss improved rare-class detection
- Grad-CAM visualizations confirmed model focuses on relevant retinal regions

10. Explainability – Grad-CAM

Grad-CAM heatmaps highlight areas in retinal images that contribute most to the model's predictions.



11. Repository Structure

```
shahwar_yasir/
├── model/
│   └── drresnet_enhanced.pth
├── notebooks/
│   └── training_gradcam.ipynb
├── README.md
├── requirements.txt
└── report.pdf
```

12. Strengths & Limitations

Strengths:

- Trained from scratch – fully original solution
- Handles class imbalance

- Explainable predictions with Grad-CAM
- GPU-optimized training with mixed precision

Limitations:

- Training time is high on large datasets
- Accuracy can improve further with larger models or ensembling

13. Conclusion

The project successfully implements a deep learning pipeline for diabetic retinopathy detection using a custom ResNet architecture.
The model achieves **~70% test accuracy** with interpretable Grad-CAM visualizations, making it suitable for early screening assistance.

14. References

1. Kaggle Diabetic Retinopathy Dataset – Kushagra Tandon
2. He, K. et al. “Deep Residual Learning for Image Recognition,” CVPR 2016
3. Grad-CAM: Selvaraju et al., 2017