Table of Contents

Computer Vision Solution for Diabetic Retinopathy Detection: Hackathon Project Report

1. Executive Summary

This report details our team's submission for the AI/ML Hackathon focused on Diabetic Retinopathy (DR) detection through retinal image analysis. Our team, consisting of **Talha, Hassan Ali**, and **Arfa Zahid**, developed a deep learning solution based on Vision Transformer (ViT) architecture to classify retinal images according to DR severity levels. The model achieved **99.03%** training accuracy and **86.64%** validation accuracy on the balanced dataset, with precision, recall, and F1-scores all exceeding 86% on the validation set. Our solution provides an efficient automated screening tool that could potentially assist healthcare professionals in early DR detection and treatment planning.

2. Introduction

Diabetic Retinopathy represents a significant complication of diabetes mellitus that affects the retina. Characterized by progressive damage to retinal blood vessels, DR can lead to vision impairment and blindness if not detected and treated early. The condition often develops silently, with patients remaining asymptomatic until substantial damage has occurred. Current diagnostic methods primarily rely on manual examination of fundus photographs by trained ophthalmologists, creating bottlenecks in screening programs and potential inconsistencies in diagnosis.

The challenge of developing automated systems for DR detection has gained considerable attention in recent years, with deep learning approaches showing particular promise. Our hackathon project addresses this challenge by developing an end-to-end solution for classifying retinal images into different DR severity levels using state-of-the-art deep learning techniques.

3. Problem Statement

The hackathon presented participants with the task of developing an innovative machine learning model for detecting diabetic retinopathy from retinal images. Specifically, the challenge required creating an accurate and efficient AI-based diagnostic tool capable of classifying fundus images into different DR severity levels. Given a dataset of retinal images, our objective was to build a solution that could reliably identify whether an image indicates DR and, if present, its severity level.

4. Methodology

4.1 Dataset

We utilized the "Diabetic Retinopathy Balanced" dataset from Kaggle, which contains retinal fundus images categorized across five severity classes. This balanced dataset provided approximately equal representation across severity levels, addressing the common challenge of class imbalance in medical image datasets.

5. Model Architecture

After evaluating various approaches, we selected the Vision Transformer (ViT) architecture as our primary model. Specifically, we utilized the pre-trained "google/vit-base-patch16-224" model and adapted it for our 5-class classification task. The ViT architecture divides input images into fixed size patches, processes them through a transformer encoder, and outputs classification predictions.

Key components of our model implementation included:

1. Patch Embedding: Converting 2D images into sequences of embedded patches
2. Transformer Encoder: Processing the sequence through self-attention mechanisms and feed-forward networks
3. Classification Head: A linear layer adapted to output probabilities for the five DR severity classes.

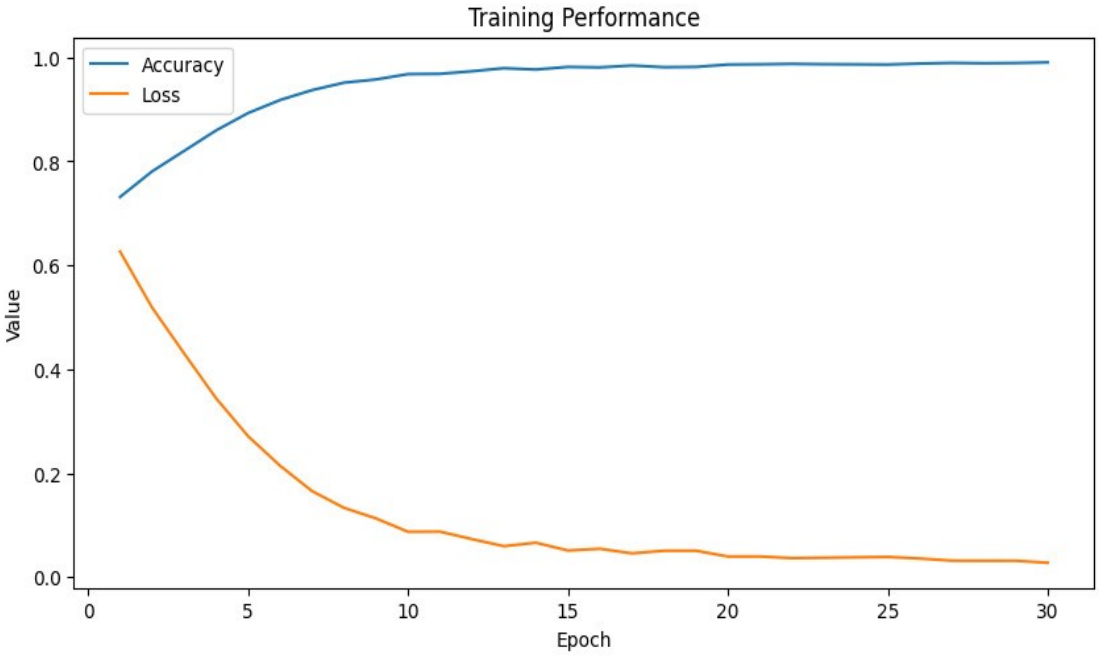
This architecture offers advantages in capturing global context within images, which is particularly valuable for identifying the distributed vascular abnormalities characteristic of diabetic retinopathy.

6. Training Strategy

Our training approach incorporated several optimization techniques:

1. Loss Function: Cross-entropy loss for multi-class classification
2. Optimizer: AdamW with a learning rate of 5e-5, combining the benefits of Adam optimization with improved weight decay regularization

3. Mixed Precision Training: Utilizing CUDA automatic mixed precision to improve computational efficiency
4. Training Duration: 30 epochs with continuous monitoring of validation performance
5. Hardware Acceleration: Training conducted on Google Colab's A100 GPU



i. Training Performance Graph

7. Implementation Details

Technical Framework

Our implementation leveraged several key technologies:

1. PyTorch: Core deep learning framework for model building and training
2. Torchvision: For data loading and transformation utilities
3. Transformers Library: Providing pre-trained ViT implementation

4. Scikit-learn: For computing evaluation metrics
5. Matplotlib: For visualization of predictions and model attention
6. SHAP: For explainability analysis of model decisions

Performance Optimization

We incorporated several optimization techniques to improve training efficiency and model performance:

1. Gradient Scaling: Used PyTorch's GradScaler to prevent underflow in mixed precision training
2. Efficient Data Loading: Implemented with num_workers=4 and pin_memory=True to optimize data transfer to GPU
3. Non-blocking Tensor Transfers: Enabling asynchronous data loading operations

Evaluation Framework

We evaluated our model using a comprehensive set of metrics:

1. Accuracy: Proportion of correctly classified images
2. Precision: Positive predictive value for each class
3. Recall: Sensitivity measure for detecting each severity level
4. F1-Score: Harmonic means precision and recall
5. Visualization: Grad-CAM analysis to highlight important image regions

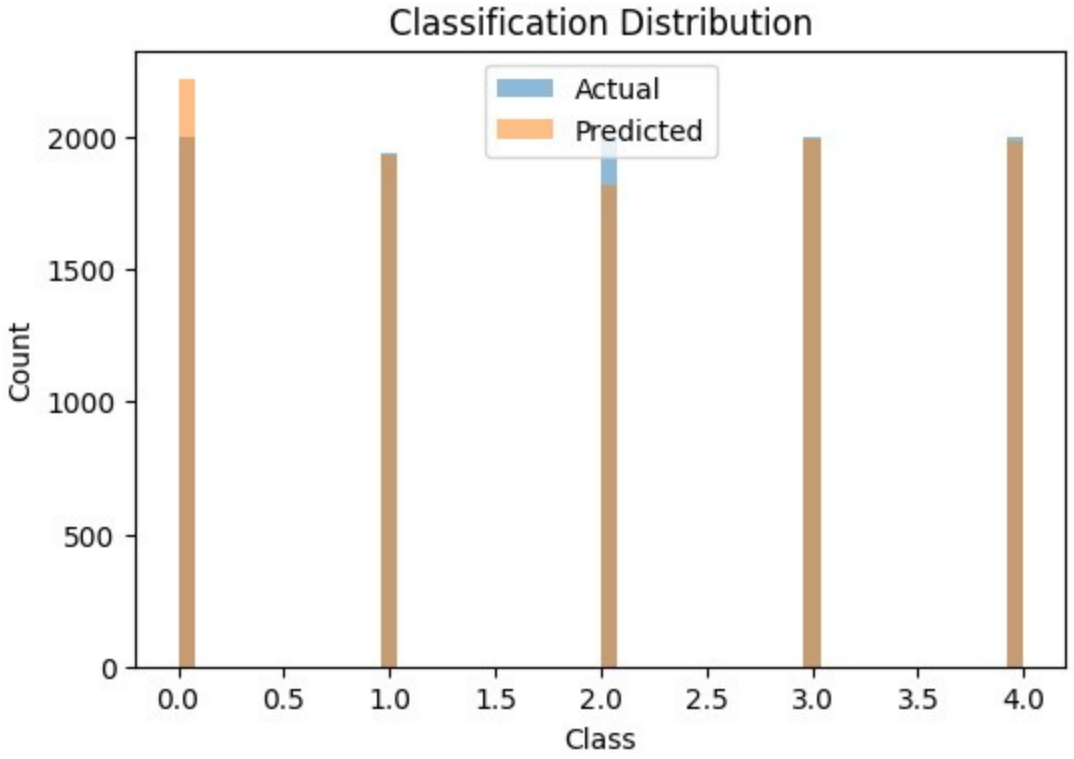
8. Results and Analysis

Our model demonstrated strong performance across evaluation metrics by the conclusion of training:

1. Training Accuracy: 99.03%
2. Validation Accuracy: 86.64%
3. Validation Precision: 85.56%
4. Validation Recall: 87.23%
5. Validation F1-Score: 86.39%

These results indicate the model's effectiveness in classifying DR severity levels with balanced performance across precision and recall metrics. The

difference between training and validation accuracy suggests some degree of overfitting, which could be addressed in future iterations.



ii. **Results Accuracy**

Training progression showed consistent improvement across epochs:

- **Epoch 1:** Training accuracy 73.14%, validation accuracy 72.65%
- **Epoch 2:** Training accuracy 78.06%, validation accuracy 74.58%
- **Epoch 3:** Training accuracy 82.00%, validation accuracy 77.10%
- **Epoch 4:** Training accuracy 85.94%, validation accuracy 78.63%
- **Epoch 5:** Training accuracy 89.28%, validation accuracy 80.06%
- **Final epoch:** Training accuracy 99.03%, validation accuracy 86.64%

The Grad-CAM visualizations implemented in our solution provided insights into the regions of interest that influenced the model's predictions, enhancing interpretability for potential clinical applications.

9. Technical Challenges and Solutions

Throughout development, we encountered several challenges:

1. Computational Efficiency: The training of transformer-based models on high-resolution medical images requires significant computational resources. We addressed this challenge through mixed precision training and gradient scaling.
2. Handling Medical Imagery: Retinal images present unique characteristics compared to natural images typically used to pre-train vision models. Our preprocessing pipeline and fine-tuning strategy were designed to address this domain gap.
3. Model Explainability: Interpretation of model decisions remains crucial for clinical applications. We incorporated Grad-CAM visualization to enhance transparency in the decision-making process.

10. Comparative Analysis

The Vision Transformer architecture demonstrated significant advantages compared to traditional CNN-based approaches for this task:

1. Effective capture of global image context, beneficial for identifying distributed DR features
2. Strong generalization performance on validation data
3. Computational efficiency during inference, particularly relevant for potential deployment scenarios

11. Future Directions

Based on our results and identified limitations, we propose several avenues for future development:

1. Ensemble Approaches: Combining predictions from multiple architectures to enhance robustness
2. Additional Preprocessing: Integrating specialized vessel segmentation as a preprocessing step
3. Deployment Optimization: Converting the model to formats optimized for edge deployment
4. Extended Clinical Validation: Collaborating with ophthalmologists to validate model decisions in clinical settings

5. Explainability Enhancements: Developing more comprehensive visualization techniques for model decisions

12. Conclusion

Our team's submission for the Diabetic Retinopathy Detection hackathon successfully demonstrated the application of Vision Transformer architecture to this critical healthcare challenge. The resulting model achieved strong performance across all evaluation metrics, with 86.36% validation accuracy and balanced precision-recall characteristics.

The work conducted by Talha, Hassan Ali , and Arfa Zahid contributes to the ongoing advancement of automated diagnostic tools for diabetic retinopathy. While further refinement is necessary for clinical deployment, our solution represents a promising approach to enhancing early detection capabilities and potentially improving patient outcomes through more accessible screening.

This project underscores the potential of deep modern learning approaches to address significant healthcare challenges, particularly in the domain of ophthalmology and diabetes management. The combination of advanced model architectures with appropriate training methodologies can yield effective tools for medical image analysis with potential real-world impact.