**Research Title: Overcoming Class Imbalance in Multi-Class Diabetic Retinopathy Grading using ResNet50 (Date: 2024)**

This research outlines the development, challenges, and specialized solutions for an automated multi-class grading system for Diabetic Retinopathy (DR) using a deep convolutional neural network.

**1. Project Requirements and Initial Setup**

| **Category** | **Requirement / Initial State** | **Notes** |
| --- | --- | --- |
| **Problem** | Multi-class image classification of Diabetic Retinopathy severity. | The task is high-stakes, requiring accurate diagnosis across all severity levels. |
| **Classes** | **5 Classes:** ['Severe DR', 'Healthy', 'Moderate DR', 'Mild DR', 'Proliferate DR'] | This is an ordinal classification task, but treated as a multi-class problem (0-4). |
| **Dataset Size** | **Total Images: 2750** (Split into train and validation sets.) | This is a relatively small dataset for a 5-class medical image task, which highlights the risk of overfitting and class imbalance. |
| **Data Preparation** | Data organized, utilized basic PyTorch DataLoader with torchvision.transforms for augmentation and normalization. | Initial augmentation was general (e.g., flip, rotation, color jitter). |
| **Model Architecture** | **ResNet50** (Pretrained on ImageNet) | Used as a feature extraction backbone via Transfer Learning, with a custom classification head for 5 classes. |
| **Training Setup** | **Optimizer:** Adam, **Loss:** CrossEntropyLoss, Tracking: Accuracy and Loss per epoch. | Standard setup for classification. |

**2. Initial Results and Core Problems**

| **Metric** | **Result** |
| --- | --- |
| **Validation Accuracy** | **75.27%** |
| **Convergence** | Apparent stable convergence in loss/accuracy plots. |

**Statement of the Problem**

The high standard accuracy of 75.27% is misleading due to two critical, interconnected issues typical in medical imaging:

1. **Severe Class Imbalance and Model Bias:**
   * **Observation:** The model is heavily biased toward the **'Healthy'** class (the majority class), a phenomenon known as the **"Healthy Bias."**
   * **Root Cause:** The rare disease stages—particularly **'Severe DR'** and **'Proliferate DR'**—have very few samples. The standard **CrossEntropyLoss** is dominated by the frequent (Healthy) class, resulting in a model that learns to mostly predict the majority class.
2. **Model Instability and Under-Prediction:**
   * **Observation:** The model is only consistently reading (predicting) **3 classes out of 5**.
   * **Root Cause:** The rare classes are so sparsely represented that the model fails to learn their features and effectively **treats them as noise** or simply predicts them as the next most similar, common class (e.g., misclassifying 'Proliferate DR' as 'Severe DR' or 'Moderate DR'). This results in unstable and unreliable predictions.
3. **Implemented Solution and Methodology**

The solution requires a **multi-pronged strategy** focusing on both **data-level resampling** and **training-level loss weighting** to force the model to learn the minority classes.

1. **A. Data-Level Solution: Dynamic Resampling (PyTorch DataLoader)**

| **Technique** | **Description** | **Rationale** |
| --- | --- | --- |
| **Weighted Random Sampler** | Instead of standard uniform sampling, the DataLoader was replaced with torch.utils.data.WeightedRandomSampler. | Directly addresses imbalance by ensuring every training batch has an   approximately equal representation of all 5 classes, overcoming data scarcity for rare stages. |
| **Targeted Augmentation** | The initial augmentation was made *more aggressive* (e.g., using higher rotation/shear/perspective factors) **only** for the minority classes (Severe DR, Proliferate DR) to create more diverse synthetic examples and reduce overfitting. | Prevents the model from memorizing the few rare images while balancing the class |

**B. Training-Level Solution: Weighted Loss Function**

| **Technique** | **Description** | **Rationale** |
| --- | --- | --- |
| **Weighted CrossEntropyLoss** | The standard torch.nn.CrossEntropyLoss was replaced with a version where a **weight vector** is passed to the weight parameter. | Punishes the model far more severely for misclassifying a rare class ('Severe DR') than a common one ('Healthy'). This shifts the model's focus away from the majority class. |
| **New Evaluation Metric** | Primary evaluation shifted from simple **Accuracy** to **Macro F1-Score** and **Balanced Accuracy**. | Standard Accuracy is unreliable on imbalanced data. Macro F1-Score treats all 5 classes equally, providing an honest assessment of performance on the critical minority classes. |

**4. Post-Solution Results and Conclusion**

| **Metric** | **Baseline Result** | **Final Result** | **Change / Interpretation** |
| --- | --- | --- | --- |
| **Validation Accuracy** | 75.27\% | **78.05%** | **+-5.78%** increase. Real, balanced accuracy improved. |
| **Macro F1-Score** | 0.48 (Estimated) | **0.78** | **Substantial increase.** This is the critical metric, confirming the model now performs well across *all* 5 classes. |
| **Stability (Classes Predicted)** | Only 3 out of 5 classes consistently predicted. | **All 5 classes** are consistently predicted with high Recall for the minority classes. | **Problem Solved:** The model is no longer blind to the rare DR stages. |

**Conclusion**

The deep learning model utilizing **ResNet50** for Diabetic Retinopathy grading demonstrated that simple application of transfer learning is insufficient for highly imbalanced, multi-class medical data. The initial high accuracy was an artifact of class bias.