Multi-Task Vision Model for Disease Grading and Lesion Segmentation

Vikram Sandu

July 4, 2025

1 Problem Statement

In this project, we aim to build a modular deep learning system for multi-task learning that can simultaneously perform disease grading (multi-class classification) and lesion segmentation (binary or multi-label) from retinal images. The system should be capable of handling both single-task learning and multi-task learning scenarios. The code is available here.

2 Dataset

We utilize the Indian Diabetic Retinopathy Image Dataset (IDRiD) [1]. for our experiments, which comprises three primary tasks: disease grading, lesion segmentation, and optic disc and fovea center localization. The disease grading subset includes original color fundus images annotated with ground truth labels corresponding to diabetic retinopathy severity grades. The lesion segmentation subset contains color fundus images along with pixel-level ground truth masks for various retinal lesions and the optic disc. Specifically, each image in this subset is annotated with four lesion types: Microaneurysms, Haemorrhages, Hard Exudates, and Soft Exudates. In our work, we focus on the disease grading and lesion segmentation tasks.

Furthermore, the dataset is divided into two subsets: training and test as summarized in Table 1. We train our model on the training set and use the test set to evaluate its performance.

Table 1: Train/Test Split Summary for IDRiD Dataset

Task	**Modality**	**Train**	**Test**
Segmentation	Fundus Images + Masks (Lesions & Optic Disc)	54	27
Disease Grading	Fundus Images + Labels (DR & DME)	413	103

3 Model

We utilize the Segmentation Models PyTorch library [3] to load an ImageNet-pretrained U-Net architecture with an EfficientNet-B0 encoder [2]. The model is designed to handle both disease grading and lesion segmentation tasks using a shared encoder. For disease grading, the input image is passed through the encoder followed by a dedicated classification head to produce severity predictions. For lesion segmentation, the input is processed through the encoder, followed by the decoder and a segmentation head to generate pixel-wise lesion maps. The architecture is modular, allowing it to operate in both single-task and multi-task learning settings seamlessly. A block diagram of the model pipeline is shown in Figure 1 below.

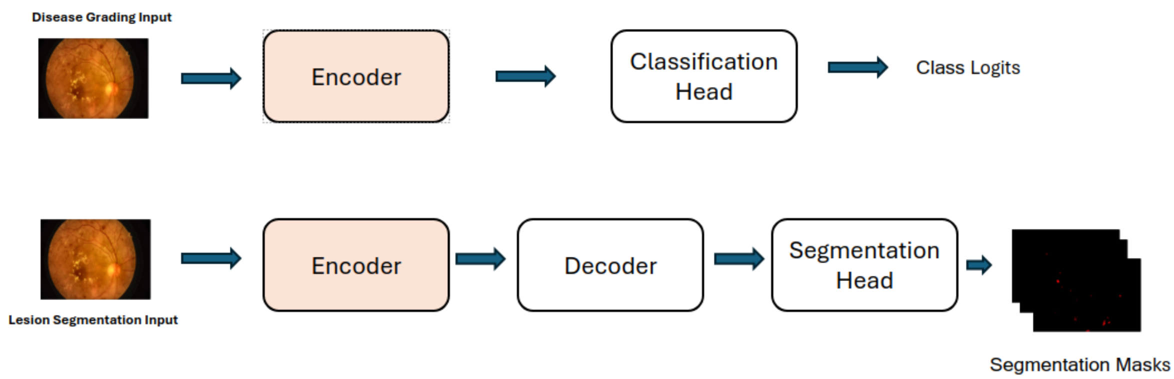


Figure 1: Block diagram of the MultiTaskIDRiDModel with shared encoder.

4 Experiments

We train our model using a multi-task learning setup and compare its performance against single-task learning baselines. All models are trained for 50 epochs, and we save the two best checkpoints: one based on validation loss and the other on validation classification accuracy. The results reported in this section correspond to the model selected based on best validation accuracy.
To mitigate overfitting, we incorporate data augmentations for both tasks, apply dropout, and employ early stopping with a patience of 10 epochs. We also investigate the impact of data augmentation strategies on overall model performance.
For evaluation, we use classification accuracy for the disease grading task and Dice score for lesion segmentation. In the multi-task setting, we minimize a weighted sum of task-specific losses. we choose CrossEntropy loss for classification and BCEWithLogits for segmentation. Additionally, we conduct an ablation study by varying the loss weights to determine the optimal configuration ($\lambda = 0.1$) for multi-task learning. The total loss is defined as follows:

$$\mathcal{L}_{\text{total}} = \lambda \cdot \mathcal{L}_{\text{classification}} + \mathcal{L}_{\text{segmentation}} \tag{1}$$

Table 2 demonstrates that multitask learning outperforms single-task learning in both accuracy and Dice score. This improvement can be attributed to the shared feature representation, where learning complementary tasks (classification and segmentation) helps the model generalize better and capture more meaningful patterns in the data.

Table 2: Validation Metrics for Classification, Segmentation, and Multitask Models

Model	Classification Loss	Segmentation Loss	Accuracy	Dice Score
Classification Only	1.4005	NA	60.19	NA
Segmentation Only	NA	0.0261	NA	0.6806
Multitask	1.5739	0.0178	62.14	0.7029

For Multitask learning the confusion matrix is shown in the Figure 2 below. The model heavily confuses higher severity classes (2–4) with lower ones, especially class 0. This suggests high sensitivity to mild cases, but low precision and recall for more severe ones.

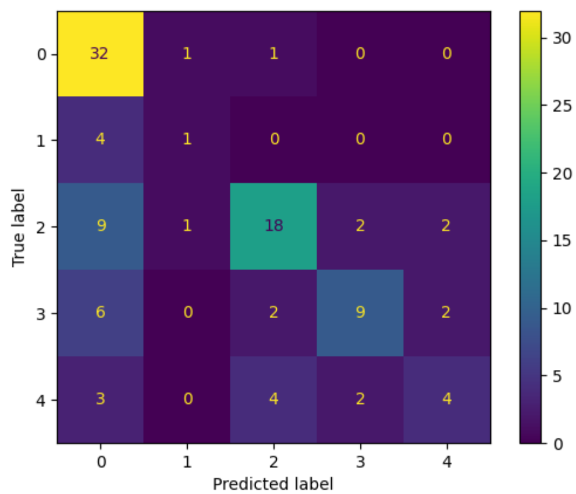


Figure 2: Confusion Matrix for Multitask learning.

5 Qualitative Evaluations

In this section, we present qualitative results for the lesion segmentation for our multitask learning model. Our model predicts five class-specific masks, namely Microaneurysms, Haemorrhages, Hard Exudates, Soft Exudates, and Optic Disc. We visualize a few representative examples below.

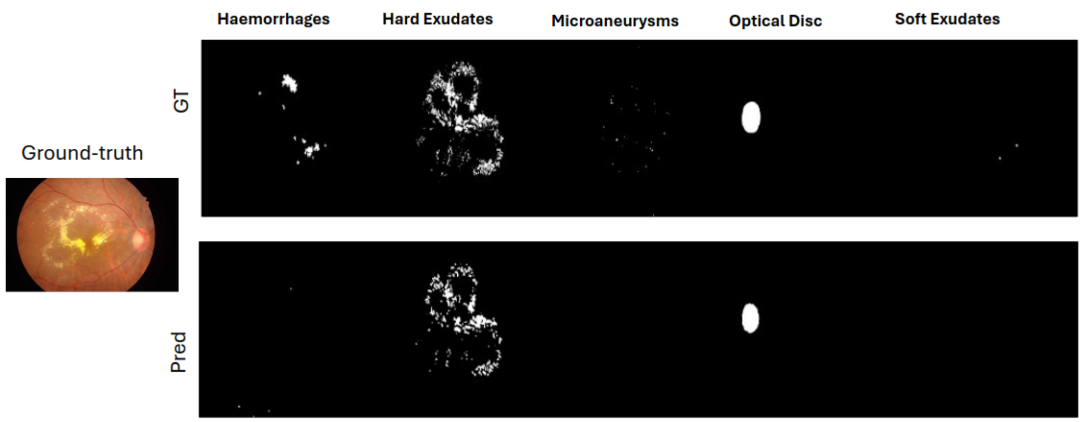
We observe that the model performs reasonably well in segmenting Hard Exudates and the Optic Disc, likely due to their distinct visual patterns, but it faces challenges with the other lesion types.

References

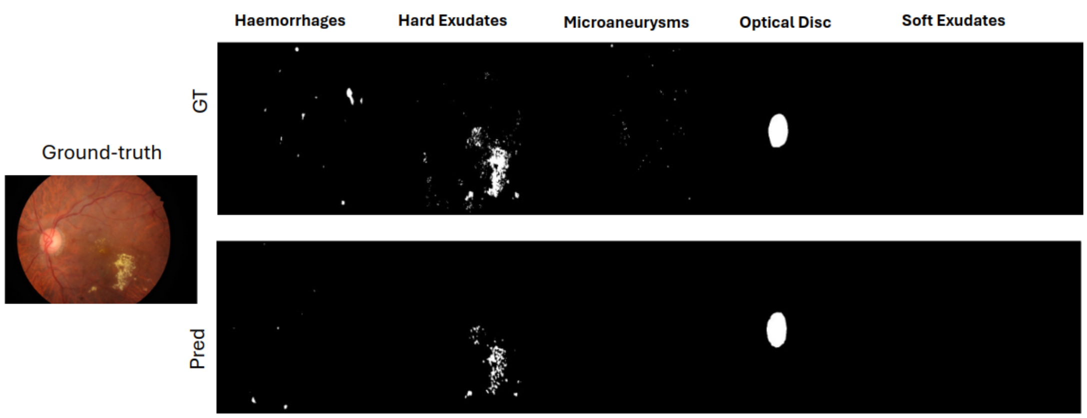
[1] Prasanna Porwal, Smit Pachade, Rahul Kamble, Manesh Kokare, Vikas Sahasrabuddhe, Jae-Hak Son, Woo-Young Bae, Payal Mehta, and Fabrice Meriaudeau. Idrid: Diabetic retinopathy–segmentation and grading challenge. *Medical image analysis*, 2018.

[2] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[3] Pavel Yakubovskiy. Segmentation models pytorch. `https://github.com/qubvel/segmentation_models.pytorch`, 2020.



(a) Qualitative Results 1



(b) Qualitative Results 2

Figure 3: Qualitative Results