Diabetic Retinopathy: A Proposed Multi-Task Deep Learning Model for Lesion Segmentation, Grading, and Localization

Raj Koticha
B.Tech CSE
Ahmedabad University
Ahmedabad, India
AU2240024

Hariohm Bhatt
B.Tech CSE
Ahmedabad University
Ahmedabad, India
AU2240179

Dhruv Premani
B.Tech CSE
Ahmedabad University
Ahmedabad, India
AU2240239

***Abstract*—Diabetic retinopathy (DR) is a complex disease that requires precise lesion segmentation, accurate disease grading, and reliable anatomical localization. In this report, we propose a novel multi-task deep learning model that integrates convolutional neural networks (CNNs), Transformer-based attention, and multi-task learning to jointly address segmentation, grading, and localization in fundus images. Our framework leverages advanced pretraining, data augmentation, and task-specific fusion modules to capture both local details and global context. We expect that this approach will achieve improvements in segmentation accuracy, grading performance, and localization precision compared to conventional single-task methods.**

***Index Terms*—Diabetic Retinopathy, Multi-Task Learning, Segmentation, Classification, Localization, CNN, Transformer, Deep Learning.**

I. Introduction

Diabetic retinopathy is a leading cause of blindness and requires early, accurate diagnosis to prevent irreversible vision loss. Traditional methods often treat lesion segmentation, disease grading, and anatomical localization as separate tasks. However, such separation neglects the interdependencies among these tasks. Recent studies have shown that integrating these tasks can enhance diagnostic performance by leveraging shared features [1]– [3].

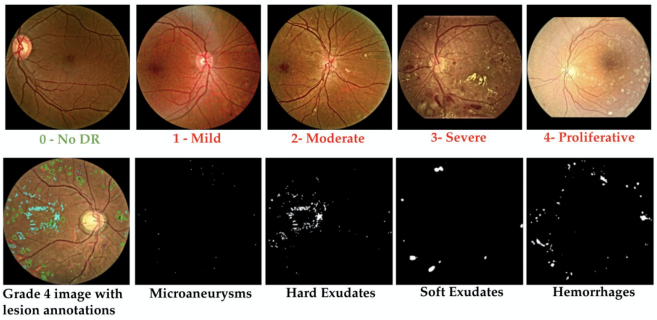

Fig. 1. Example images showing different DR grades (top row: 0–4) and annotated lesions (bottom row). From left to right: 0 (No DR), 1 (Mild), 2 (Moderate), 3 (Severe), and 4 (Proliferative). Bottom images illustrate microaneurysms, hard exudates, soft exudates, and hemorrhages.

In this report, we introduce our proposed model, a unified deep learning approach that combines segmentation, classification, and localization in a single end-to-end trainable network. By fusing CNN and Transformer-based components, the model captures both high-resolution details and global context, which is essential for recognizing subtle lesions and complex disease patterns.

II. Methodology

In seeking to surpass existing multi-task models such as DRG-Net, our proposed approach focuses on addressing several key limitations that hinder the efficiency and effectiveness of diabetic retinopathy detection. In the coming days, we will work on improving the following aspects:

A. Optimizing the Encoder

The encoder architecture in segmentation models often contains a high number of parameters, leading to increased computational costs and training inefficiencies. We aim to:

- Reduce the number of parameters in the encoder while ensuring that feature extraction remains effective for lesion segmentation.
- Maintain a balance between model complexity and segmentation accuracy to achieve an optimized trade-off.

B. Improving Segmentation Architecture

Traditional U-Net and its variants, while widely adopted, can be computationally intensive and inefficient for large-scale medical imaging tasks. To enhance efficiency, we will:

- Investigate alternative segmentation architectures that reduce computational overhead while preserving segmentation precision.
- Explore mechanisms that streamline the segmentation backbone without unnecessary depth or redundancy.

C. Enhancing Lesion Localization

Lesion detection in DR images presents challenges due to overlapping or clustered lesions, which standard Region Proposal Networks (RPNs) may struggle to address. To overcome these issues, we plan to:

- Identify methods to improve lesion localization, especially in cases of dense lesion distributions.
- Develop strategies that enable accurate differentiation between closely packed lesions without relying on rigid bounding-box proposals.

D. Multi-Task Learning Considerations

Beyond segmentation, our model must effectively handle DR grading and anatomical localization. Challenges in multi-task learning include:

- Ensuring that a shared encoder effectively captures global retinal features for all tasks without excessive parameter redundancy.
- Implementing augmentation strategies that enhance model robustness across different DR-related tasks.

Addressing these issues will allow us to refine the framework and develop a more efficient, accurate, and computationally optimized multi-task learning model for diabetic retinopathy detection.

III. Results

While the final implementation and training of our proposed model are still ongoing, we outline our planned evaluation protocol on the IDRiD dataset to demonstrate the potential effectiveness of our approach:

- **Segmentation:** We will measure the average Dice coefficient across multiple lesion types to assess the network's lesion delineation quality. Our hypothesis is that the lightweight encoder and attention-based modules will achieve high Dice scores comparable to or surpassing existing U-Net baselines.
- **Disease Grading:** We intend to evaluate overall DR classification accuracy as well as the quadratic weighted kappa (QWK) score. We anticipate that the synergy between segmentation and classification heads, combined with task-specific attention, will yield improved grading performance compared to single-task models.
- **Localization:** The model's ability to pinpoint the optic disc and fovea centers will be gauged by measuring the mean Euclidean distance error. We expect that integrating multi-scale features and spatial attention will support precise localization, even in challenging images with variable lighting or lesion distribution.

Upon completion of training, we plan to apply statistical significance tests (e.g., Wilcoxon signed-rank, paired t-tests) to verify that any observed improvements are robust and not due to chance. A comprehensive report on performance metrics and comparative analyses will be published once the model training and validation are finalized.

IV. Discussions

The revised segmentation encoder, combined with attention-based or lightweight convolutional modules, reduces the parameter footprint and computational overhead typical of U-Net designs. This not only speeds up training and inference but also makes the system more scalable across different imaging conditions. Moreover, by avoiding strict bounding-box proposals, we mitigate the RPN's limitations in detecting overlapping or clustered lesions—an essential requirement in advanced DR stages. Jointly learning disease severity and anatomical landmarks further benefits segmentation, as task interdependencies reinforce each other through shared features.

V. Conclusion

In this report, we presented our proposed multi-task deep learning model for diabetic retinopathy analysis that jointly performs lesion segmentation, disease grading, and anatomical localization. By focusing on reducing encoder parameters, moving beyond standard U-Net architectures, and handling overlapping lesions without conventional RPNs, our approach aims to advance the state of the art. Future work will extend this framework to multi-disease analysis and incorporate uncertainty estimation to further support clinical decision-making.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *arXiv preprint arXiv:1505.04597*, 2015.

[2] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 3431–3440.

[3] X. Yu, C. Zhang, Z. Huang, Y. Shi, and Y. Chen, "Cascaded convolutional neural network for retinal image segmentation and localization," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 1, pp. 169–178, 2018.

[4] H. Bui, S. Liu, S. La, and J. H. Lee, "A hybrid model for DR grading using segmentation-assisted feature extraction," *Comput. Methods Programs Biomed.*, vol. 191, pp. 105373, 2020.

[5] W. Zhou, D. Xu, and Y. Zhang, "RSG-Net: Retinopathy severity grading network for diabetic retinopathy classification," *IEEE Trans. Biomed. Eng.*, vol. 71, pp. 3476–3484, 2024.

[6] H. Md Tusfiqur, A. S. Abdullah, M. A. M. Azam, and S. A. Hassan, "DRG-Net: Interactive joint learning of multi-lesion segmentation and classification for diabetic retinopathy grading," *arXiv preprint arXiv:2212.14615*, 2022.