망막 이미지에서의 질병 진단: 교차 데이터셋 연구

Van-Nguyen Pham[1], Sun Xiaoying[1], 추현승 [2]
[1] 성균관대학교 전자전기컴퓨터공학과 박사과정
[2] 성균관대학교 전자전기컴퓨터공학과 교수

nguyenpv195@g.skku.edu, alvy.sun@g.skku.edu, choo@skku.edu

Disease Diagnosis on Fundus Images: A Cross-Dataset Study

Van-Nguyen Pham[1], Sun Xiaoying[1], Hyunseung Choo[1]
[1]Dept. of Electrical and Computer Engineering, Sungkyunkwan University

요약

This paper presents a comparative study of five deep learning models—ResNet50, DenseNet121, Vision Transformer (ViT), Swin Transformer (SwinT), and CoatNet—on the task of multi-label classification of fundus images for ocular diseases. The models were trained on the Ocular Disease Recognition (ODIR) dataset and validated on the Retinal Fundus Multi-disease Image Dataset (RFMiD), with a focus on five disease classes: diabetic retinopathy, glaucoma, cataract, age-related macular degeneration, and myopia. The performance was evaluated using the area under the receiver operating characteristic curve (AUC-ROC) score for each class. CoatNet achieved the best AUC-ROC scores for diabetic retinopathy, glaucoma, cataract, and myopia, while ViT outperformed CoatNet for age-related macular degeneration. Overall, CoatNet exhibited the highest average performance across all classes, highlighting the effectiveness of hybrid architectures in medical image classification. These findings suggest that CoatNet may be a promising model for multi-label classification of fundus images in cross-dataset scenarios.

1. Introduction

Fundus imaging plays a crucial role in the early detection and diagnosis of retinal diseases such as diabetic retinopathy, glaucoma, and age-related macular degeneration. These conditions, if left untreated, can lead to irreversible vision loss. With advancements in machine learning, automated analysis of fundus images has gained significant attention, particularly for multi-label classification tasks where multiple disease conditions may coexist in a single image.

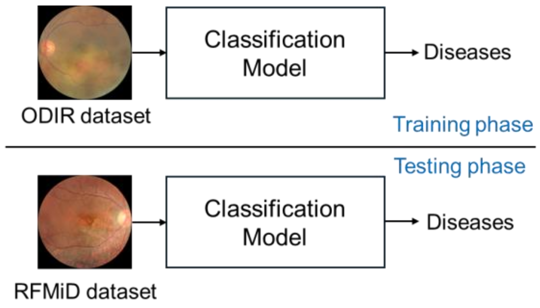
Multi-label classification is a more complex task compared to single-label classification, as it requires the model to accurately predict the presence of multiple disease conditions. Given the inherent challenges of fundus image classification, such as variations in image quality and disease presentation, the need for robust models is paramount.

In this study, we aim to compare the performance of various machine learning models trained on the **ODIR** dataset, which consists of fundus images labeled with multiple ocular diseases, and validated on the **RFMiD** which contains a diverse set of retinal images from different patients. By training on ODIR and validating on RFMiD, we can assess the generalization capability of these models, particularly in handling cross-dataset variations.

The primary objective of this work is to explore how well models trained on the ODIR dataset can perform on a different yet related dataset, RFMiD, in the context of multi-label classification. Through this comparison, we explore the strengths and weaknesses of various models in generalizing across datasets, which is crucial for developing reliable automated diagnostic systems.

2. Methodology

This study uses the ODIR dataset for training and the RFMiD for testing, as shown in Figure 1. The ODIR dataset contains 10,000 fundus images from multiple centers, annotated with 8 disease labels. The RFMiD dataset consists of 3,200 images with 5 labels in common with the ODIR dataset, providing a suitable dataset for cross-dataset validation. All images were resized to 224x224 pixels, normalized using mean and standard deviation from the ImageNet dataset, and augmented with random horizontal and rotation to improve generalization.



(Figure 1) Overall process of this study.

We compare the performance of multiple models, including CNN-based architectures like ResNet50 [1] and Densenet121 [2], Transformer-based models such as Vision Transformer [3] and Swin Transformer [4], and the hybrid CNN-Transformer-based model like CoatNet [5]. Training was conducted for 20 epochs with a batch size of 8, using the AdamW optimizer with a learning rate of 1e-5 for Transformer-based and hybrid CNN-Transformer models, and 1e-4 for CNN models. Binary cross-entropy was used as the loss function to handle the multi-label classification task.

3. Experiment Result

Model performance was evaluated using the area under the receiver operating characteristic curve (AUC) score for each label. AUC provides a robust measure of the models' ability to distinguish between disease and non-disease categories across different thresholds. There are 5 diseases in common between the two datasets: diabetic retinopathy (D), glaucoma (G), cataract (C), age-related macular degeneration (A), and myopia (M). The performance of each class is reported in Table 1. Among the models, CoatNet consistently demonstrated superior performance, achieving the highest AUC scores for diabetic retinopathy, glaucoma, cataract, and myopia. In contrast, for AMD, ViT outperformed CoatNet, achieving the best AUC score in this category. Despite this, CoatNet still had the best average performance across all classes. Traditional CNN-based models like ResNet50 and DenseNet121, while competitive, were generally outperformed by hybrid models such as CoatNet, showcasing the advantage of combining convolutional and Transformer-based architectures for fundus image classification.

<Table 1> Performance comparison of different models in terms of AUC score (%)

Model	D	G	C	A	M	Average
Resnet50	94.06	74.68	87.59	92.25	90.72	87.86
Densenet121	93.81	72.36	85.02	90.41	85.59	85.44
ViT	94.85	79.97	85.56	94.34	93.65	89.67
SwinT	95.62	76.41	83.47	93.09	91.75	88.07
CoAtNet	94.90	83.60	88.40	92.96	94.18	90.81

4. Conclusion

In this study, we evaluated the performance of five deep learning models—ResNet50, DenseNet121, ViT, SwinT, and CoatNet—for multi-label classification of fundus images, focusing on five ocular disease classes: diabetic retinopathy, glaucoma, cataract, age-related macular degeneration, and myopia. The models were trained on the ODIR dataset and validated on the RFMiD dataset, using the AUC score as the primary evaluation metric. Among the models, CoatNet achieved the highest AUC scores for diabetic retinopathy, glaucoma, cataract, and myopia, demonstrating the advantages of hybrid architectures that combine convolutional and Transformer-based approaches. These results suggest that CoatNet is highly effective for multi-label classification of fundus images, especially in cross-dataset validation scenarios. Future work may explore fine-tuning these models or integrating additional datasets to further improve the generalization capabilities of deep learning models in ophthalmology.

Acknowledgements

This work was supported in part by the BK21 FOUR Project (50%) and the Korea government (MSIT), IITP, Korea, under the ICT Creative Consilience program (RS-2020-II201821, 25%), Development of Brain Disease (Stroke) (RS-2024-00459512, 25%).

참고문헌

[1] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[2] Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. "Densely connected convolutional networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708. 2017.

[3] Dosovitskiy, Alexey. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).

[4] Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. "Swin transformer: Hierarchical vision transformer using shifted windows." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012-10022. 2021.

[5] Dai, Zihang, Hanxiao Liu, Quoc V. Le, and Mingxing Tan. "Coatnet: Marrying convolution and attention for all data sizes." Advances in neural information processing systems 34 (2021): 3965-3977.