Dorado, Abram C. **CMSC 199**

Annotated Bibliography of RRL

A Multi-Label Deep Learning Model with Interpretable Grad-CAM for Diabetic Retinopathy Classification

Summary:

- This paper introduces a deep learning-based multi-label classification model using Grad-CAM for detecting diabetic retinopathy (DR). The model, based on the ResNet architecture, classifies fundus images and identifies lesion areas automatically. By treating different lesions as multiple labels, it reduces annotation work while improving classification accuracy. The model achieved a sensitivity of 93.9% and a specificity of 94.4%. It provides interpretable visual feedback to assist clinical use, enhancing transparency and making it a reliable tool for DR detection and diagnosis.

Relation to Proposal:

- Upon reading this article, it is directly related to my proposal, as both focus on using deep learning models enhanced with Grad-CAM for explainable diabetic retinopathy (DR) detection. The multi-label approach described offers valuable insights into how I can structure my model to handle various lesion types and stages of DR, which is critical for improving the accuracy and reliability of my system. By treating different lesions as separate labels, this method simplifies the detection process and reduces the complexity of the task, an approach that I can incorporate to ensure my model is versatile and scalable. Moreover, the authors' emphasis on minimizing the labor required for annotating fundus images resonates with my objective to create an efficient, scalable system that can be applied in real-world clinical settings, especially where data is limited. The high sensitivity and specificity achieved by their model set a strong performance benchmark that my proposed system will aim to meet or exceed. The use of Grad-CAM to generate interpretable results aligns with my goal to increase clinician trust, making AI-driven diagnoses more transparent and trustworthy. This article provides both practical and theoretical guidance, helping me strike a balance between achieving high diagnostic accuracy and ensuring the model is interpretable and user-friendly in clinical environments.

Towards Explainable Deep Neural Networks for the Automatic Detection of Diabetic Retinopathy

Summary:

- Towards Explainable Deep Neural Networks for the Automatic Detection of Diabetic Retinopathy, explores the explainability of deep learning models used for detecting diabetic retinopathy (DR). The author evaluates several models, including VGG-16, and applies Grad-CAM to provide visual feedback on the model’s decisions. While the models demonstrated high accuracy, they often

struggled to highlight lesion-specific areas that corresponded with clinical diagnoses. This paper emphasizes the need for improved transparency in deep learning models for DR detection to foster trust and reliability in clinical settings, where explainability is critical for adoption.

Relation to Proposal:

- This study highlights key challenges in the development of explainable AI systems for DR detection, which is central to my research. It discusses the difficulty of aligning deep learning model predictions with lesion-specific regions, a problem I intend to address by refining the use of Grad-CAM in my project. The evaluation of VGG-16 as one of the models provides a useful comparison point for selecting an architecture that suits my objectives of interpretability and accuracy. Moreover, the paper's emphasis on explainability resonates with my goal to create a system clinicians can trust by providing transparent and interpretable results. The findings underscore the importance of creating models that not only achieve high diagnostic performance but also offer insights into the reasoning behind predictions, which is essential for gaining clinical acceptance. My proposal seeks to address the limitations pointed out in this paper by improving Grad-CAM's alignment with clinical outcomes and exploring new methods to enhance the interpretability of model predictions.

A novel approach for intelligent diagnosis and grading of diabetic retinopathy

Summary:

- The research study proposes the DRGCNN model, which combines EfficientNetV2-M with a Category Attention Module (CAM) for diagnosing and grading diabetic retinopathy (DR). The model addresses the issue of imbalanced data by balancing feature maps and incorporates bilateral eye images to improve grading accuracy. Experimental results on the EyePACS and Messidor-2 datasets show competitive kappa values of 86.62% and 86.16%. This approach improves diagnostic accuracy, reduces computational complexity, and enhances performance on real-world datasets, making it a valuable tool for automated DR diagnosis.

Relation to Proposal:

- This paper is highly relevant to my research, as it addresses common challenges in DR detection that I plan to confront, such as imbalanced data and the need for more comprehensive image analysis using bilateral eye images. The authors' use of EfficientNetV2 and the Category Attention Module (CAM) provides innovative techniques that I could adopt or adapt in my work. Specifically, the introduction of CAM helps in balancing the feature extraction process across multiple lesion types, an issue that is crucial for achieving high accuracy in diverse datasets. Furthermore, the study's use of publicly available datasets like

EyePACS and Messidor-2 offers a valuable benchmark for comparing the performance of my proposed system. The focus on computational efficiency, coupled with high diagnostic performance, aligns with my goal of developing a scalable, clinically applicable AI system for DR detection. This study also underscores the importance of robust data handling techniques, which will be a key consideration in the design and implementation of my project.

The adoption of deep learning interpretability techniques on diabetic retinopathy analysis: a review

Summary:

- This review focuses on deep learning methods used for analyzing diabetic retinopathy (DR), with a particular emphasis on model interpretability techniques such as Grad-CAM, occlusion, and sensitivity analysis. While these models have demonstrated high accuracy in diagnosing DR, the "black-box" nature of deep learning has hindered their widespread clinical adoption. The paper outlines both the strengths and weaknesses of various interpretability methods and suggests future research directions aimed at improving transparency. This is crucial for increasing clinician trust and ensuring the successful integration of AI-driven diagnostic tools into clinical practice for DR detection.

Relation to Proposal:

- The paper review is instrumental in framing the interpretability aspect of my proposal, as it discusses multiple techniques for enhancing transparency in deep learning models for DR detection. Grad-CAM, a core focus of my research, is thoroughly explored in this review, alongside other methods such as occlusion and sensitivity analysis, providing a broader context for my project. The authors highlight the importance of addressing the "black-box" problem in deep learning, which aligns with my objective to build a system that clinicians can understand and trust. The review also identifies gaps in current research, particularly in making AI-based models more interpretable for real-world clinical use, which my proposal aims to address. By focusing on improving the transparency of diagnostic models, I plan to contribute to overcoming the barriers highlighted in this paper, such as limited clinical adoption due to a lack of explainability. This comprehensive review reinforces the significance of my work in advancing the field of Explainable AI in medical diagnostics.

Diabetic retinopathy detection and diagnosis by means of robust and explainable convolutional neural networks

Summary

- Having read this article, it proposes two convolutional neural network (CNN) models for detecting diabetic retinopathy (DR) and distinguishing between non-proliferative and proliferative DR. The models achieved high accuracy, with 98% for the healthy vs. DR classification and 91% for differentiating between non-proliferative and proliferative DR. The authors used Grad-CAM and Score-CAM to provide visual explanations, improving the interpretability and trustworthiness of the models. The paper also introduces a similarity index to evaluate model robustness, making the system more reliable for clinical applications.

Relation to Proposal:

- The paper is highly relevant to my research, as it focuses on using CNNs with Grad-CAM for explainable DR detection, which aligns closely with the goals of my project. The models' success in providing accurate and interpretable results offers a strong foundation for my work, especially in using Grad-CAM to enhance clinician trust in AI-driven diagnoses. Additionally, the introduction of a similarity index to assess the robustness of model predictions provides a novel evaluation metric that I could incorporate into my system. This study's emphasis on explainability and accuracy in DR diagnosis directly supports my proposal's aim to make AI-based models more transparent and clinically applicable. Moreover, the paper's exploration of CNN architectures, including their limitations and performance across different stages of DR, provides valuable insights that will guide the selection and optimization of models for my project. The findings in this paper will serve as a benchmark for evaluating the success of my proposed system in terms of both interpretability and diagnostic performance.

References

Jiang, H., Xu, J., Shi, R., Yang, K., Zhang, D., Gao, M., ... & Qian, W. (2020, July). A multi-label deep learning model with interpretable Grad-CAM for diabetic retinopathy classification. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 1560-1563). IEEE. https://doi.org/10.1109/EMBC44109.2020.9175884

Alghamdi, H. S. (2022). Towards explainable deep neural networks for the automatic detection of diabetic retinopathy. Applied Sciences, 12(19), 9435. https://doi.org/10.3390/app12199435

Hai, Z., Zou, B., Xiao, X., Peng, Q., Yan, J., Zhang, W., & Yue, K. (2024). A novel approach for intelligent diagnosis and grading of diabetic retinopathy. Computers in Biology and Medicine, 172, 108246. https://doi.org/10.1016/j.compbiomed.2024.108246

Lim, W. X., Chen, Z., & Ahmed, A. (2022). The adoption of deep learning interpretability techniques on diabetic retinopathy analysis: a review. Medical & Biological Engineering & Computing, 60(3), 633-642. https://doi.org/10.1007/s11517-021-02487-8

Mercaldo, F., Di Giammarco, M., Apicella, A., Di Iadarola, G., Cesarelli, M., Martinelli, F., & Santone, A. (2023). Diabetic retinopathy detection and diagnosis by means of robust and explainable convolutional neural networks. Neural Computing and Applications, 35(23), 17429-17441.https://doi.org/10.1007/s00521-023-08608-8