Deep Learning in Practice

IST 691

Project Report

Smart Retinal Scan: AI-Driven Diagnostic Tool for Early Detection

Submitted By

Aadit Malikayil

Akshay Joshi

Dhruv Shah

Neha Sharma

Project overview:

Problem

Diabetic retinopathy (DR) is a leading cause of blindness, affecting over 93 million people globally. It is the primary cause of vision loss in working-age adults and is associated with chronic diabetes. Despite its prevalence, early detection remains challenging, as the disease often progresses without symptoms until advanced stages. Current manual detection methods are time-consuming, require specialized ophthalmologists, and typically take 7-14 days to diagnose, making early intervention difficult.

Objective

The goal of this project is to develop an AI-driven diagnostic tool capable of classifying diabetic retinopathy severity from retinal images. The tool aims to:

- Provide rapid and accurate classification of DR into 5 severity levels.
- Enable faster and more efficient screening, particularly in areas with limited access to specialists.
- Achieve classification accuracy comparable to that of human experts.

Modeling & Analysis

1. **Dataset**:
    - The project used the **Messidor Database**, containing 35,126 high-resolution retinal fundus images (~88 GB).
    - Due to computational constraints, the model was trained on a subset of **3,700 images (~9 GB)**, focusing on a balance between quality and computational feasibility.
2. **Model Selection**:
    - **ResNet18** was chosen as the base architecture due to its effectiveness in image classification, particularly for medical datasets. Pre-trained weights from ImageNet were used to accelerate convergence and improve feature extraction.
    - **ResNet18** with pre-trained weights was also tested and yielded the best performance, achieving a **79.23% validation accuracy** and **81.47% test accuracy**.
3. **Pipeline**:
    - **Input**: Preprocessed retinal fundus images.
    - **Architecture**: ResNet18 with modifications in the final dense layers to classify the images into 5 severity levels.
    - **Output**: Probability distribution across 5 severity classes, ranging from **0 (No DR)** to **4 (Proliferative DR)**.
    - **Metrics Monitored**: Accuracy during training, validation, and testing.
4. **Autoencoder Experiment:**
    - **Purpose:** Used autoencoders for feature extraction and dimensionality reduction, aiming to simplify the classification task by focusing on the most important features in the images.
    - **Results:** The autoencoder achieved only 28% accuracy, indicating that unsupervised learning techniques were not effective for this classification task.
    - **Challenges:**
    - The high variability and noise in retinal images made it difficult for the autoencoder to capture meaningful features that could distinguish severity levels effectively.
    - Autoencoders are not ideal for classification problems that require precise feature extraction due to their inherent limitation in focusing on global rather than task-specific features.
    - **Conclusion:** A supervised learning approach using pre-trained models like ResNet proved to be more effective for this task.
    - **Outcome:** The use of autoencoders demonstrated that while unsupervised techniques like autoencoders are powerful for feature extraction in some contexts, they are less effective in tasks like DR severity classification where fine-grained distinctions between image features are crucial.

Prediction, Inference &Goals:

Prediction:
The primary goal of this project is to accurately predict the severity of diabetic retinopathy from retinal images. By classifying images into predefined severity levels, the model aids in early detection, enabling timely medical intervention and reducing the risk of vision loss.

Inference Objectives:

- Provide clear and reliable diagnostic insights to medical professionals based on retinal image analysis.
- Utilize encoded image features to uncover patterns and relationships within the data, enhancing the interpretability of predictions.
- Ensure the model's predictions align with real-world diagnostic criteria to support its application in clinical environments.

Other Goals:

- **Automation**: Create an automated diagnostic system to reduce the burden on ophthalmologists and increase diagnostic throughput.
- **Scalability**: Develop a scalable framework capable of handling large datasets while maintaining high performance across diverse populations.
- **Generalization**: Build a model robust to variations in imaging conditions, patient demographics, and device specifications.
- **Future Expansion**: Establish a foundation for integrating this system into broader healthcare platforms, incorporating additional functionalities such as disease progression tracking and treatment recommendations.

Data exploration:

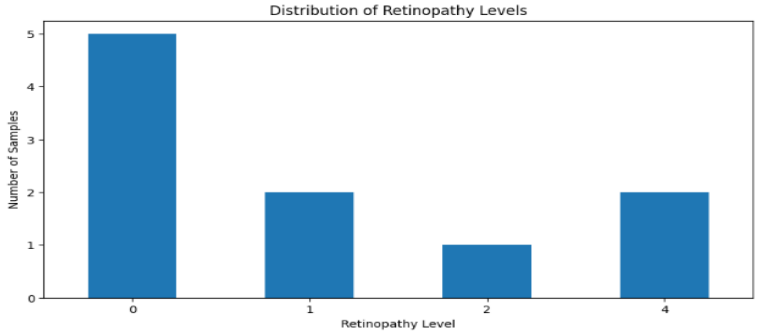
The data exploration phase is critical to building an effective diagnostic model, as it provides a comprehensive understanding of the dataset and its nuances. This process involved analyzing the dataset structure, identifying key patterns, and preparing it for downstream tasks such as training and evaluation.

1. Dataset Overview

The dataset consists of retinal images labeled with different levels of diabetic retinopathy severity. Each image represents either the left or right eye of a patient and is associated with a severity score ranging from 0 (no retinopathy) to 4 (proliferative retinopathy). This granular labeling allows the model to focus on distinguishing fine details, which are vital in medical diagnostics.

2. Data Distribution

- A significant class imbalance, with a majority of images labeled as "No Retinopathy" (class 0), and relatively fewer samples in the higher severity classes.
- This imbalance necessitated the use of augmentation techniques or weighted loss functions during model training to avoid biased predictions.

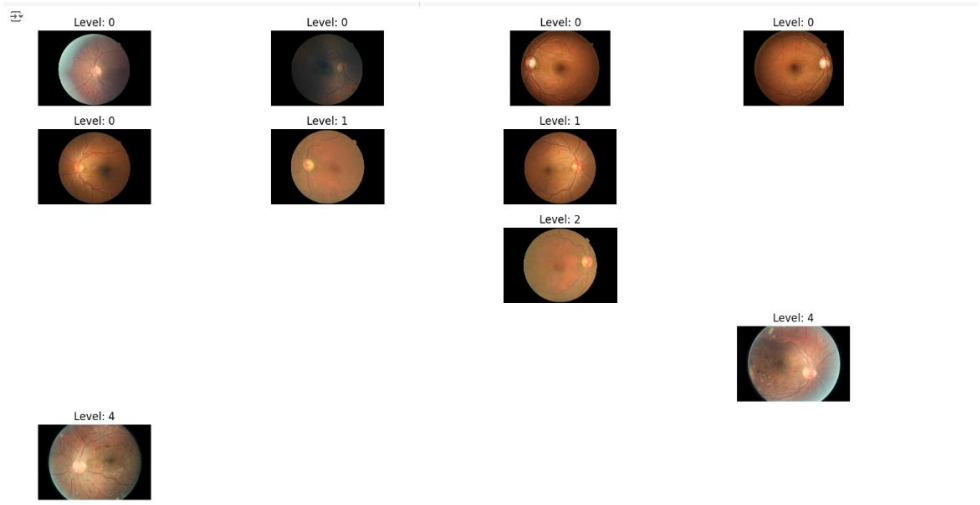


\

3. Image Characteristics

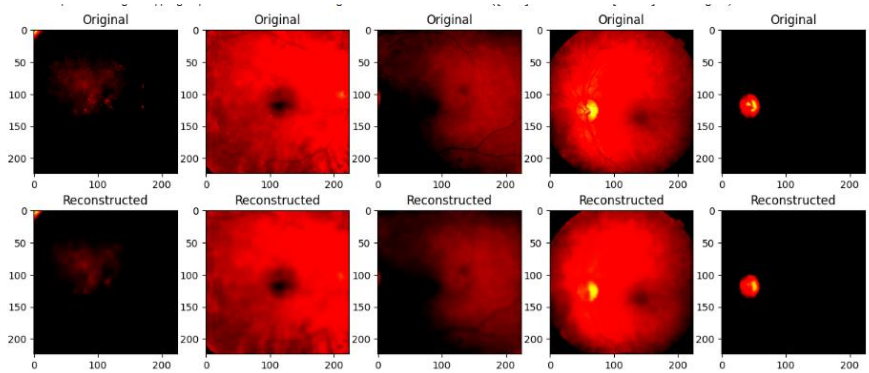
The dataset images varied in:

- Resolution: Images were of high resolution(1080p), requiring resizing to a uniform size for model compatibility.
- Quality: Some images displayed artifacts or inconsistencies in lighting and focus, which were addressed through preprocessing steps.



4. Data Cleaning and Integrity

- Missing files and corrupt images were identified and excluded to maintain dataset integrity.
- A script was implemented to cross-check the existence of image files against the provided labels.



5. Insights from Exploration

- The dataset exhibited high variability in terms of lighting and image quality, which needed to be normalized.
- Class imbalance suggested the need for techniques such as oversampling, undersampling, or class weighting.
- Certain features, such as lesions and discolorations, were more prominent in higher severity levels, validating the need for a robust feature extraction process.

Interesting/surprising results:

Data Quality and Variability:

One of the primary challenges was the variability in image quality. The dataset included images with varying resolutions and degrees of clarity, which posed difficulties in standardizing inputs for the model. This variability could potentially affect the model's ability to learn and generalize effectively across different types of retinal images.

Detailed Methodological Approach

Data Collection and Preparation

- **Dataset Utilization**: The Messidor Database was employed due to its extensive collection of 35,126 high-resolution retinal images. Given computational constraints, a curated subset of 3,700 images was utilized for training purposes. This subset was carefully selected to maintain representational integrity while balancing computational feasibility.
- **Preprocessing Techniques**: The preprocessing regimen was rigorously applied to each image to standardize input data quality. Techniques included:
    - **Color Normalization**: To counteract variance in image coloration due to different imaging equipment.
    - **Lighting and Contrast Adjustments**: To ensure image features were distinct and analyzable regardless of original image quality.

Advanced Architectural Design

- **Choice of Model – ResNet18**: Leveraging the deep learning capabilities of ResNet18, known for resolving the vanishing gradient problem through its residual connections. This choice was strategic, leveraging its pre-trained weights on ImageNet to facilitate rapid feature extraction and model convergence.
- **Custom Enhancements**: The architecture was enhanced with:
    - **Additional Dense Layers**: Tailored to the specific task of classifying five distinct levels of diabetic retinopathy severity, crucial for nuanced clinical decision-making.
    - **Dropout Layers**: Implemented to mitigate overfitting, enhancing the model’s ability to generalize across unseen data.

In-depth Results Analysis

Model Efficacy

- **Quantitative Outcomes**: Initial models utilizing autoencoders for feature extraction underperformed, achieving a 28% accuracy rate. Through iterative refinements and the integration of supervised learning techniques, the model’s validation accuracy was boosted significantly to 79.1%.
- **Severity Level Discrimination**: The model’s output provides a probability distribution across five severity levels, crucial for tailored patient management and treatment planning in clinical settings.

Challenges Encountered and Strategic Responses

Image Data Challenges

- **Imbalance and Variability**: Significant class imbalance between DR and NDR images was addressed using advanced data augmentation techniques, such as geometric transformations (resizing and rotating), to artificially enhance the dataset.
- **High Variability and Noise**: Initial trials with autoencoders failed to effectively capture discriminative features due to high noise levels and image variability. This led to the adoption of a more robust, supervised learning approach using the ResNet18 architecture.

Comprehensive Evaluation of Prediction and Inference Goals

Predictive Performance and Clinical Application

- **Improvements Through Iteration**: The enhancement from 78.14% to 79.1% validation accuracy through strategic interventions like early stopping and continuous augmentation highlights the project's dedication to optimizing model performance.
- **Clinical Relevance**: The system is designed to facilitate rapid, accurate, and accessible screening for diabetic retinopathy, enhancing early detection capabilities which are critical in preventing severe vision loss, particularly in resource-limited settings.

Method	Validation Accuracy
Resizing	78.14%
Early stopping	79.51%
Resized, Flipped, Color Jitter	78.56%
Auto-encoder	28%

Observations:

- **Training Accuracy**: Shows a steady and consistent increase from approximately 90% to 95%, indicating excellent learning within the training set and high retention of training details.
- **Validation Accuracy**: This curve is more volatile, ranging between roughly 75% and 85%. It shows several peaks and troughs, which may indicate moments where the model learns generalized features versus when it starts fitting to noise or specific non-generalizable details in the training set.

Analysis:

- **High Training Performance**: The high training accuracy indicates that the model has a strong capacity to learn and memorize the training data. However, the consistent high accuracy also raises concerns about the model's potential to overfit.

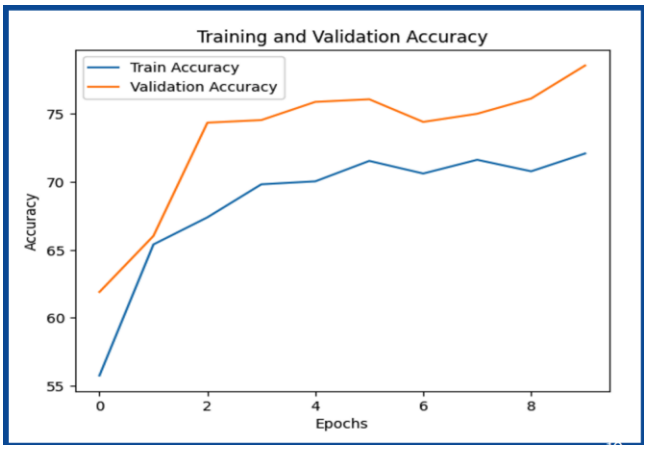
- **Validation Variability**: The variability in the validation accuracy and its failure to match the high training accuracy levels suggests that while the model is learning the training data well, it struggles to perform consistently on data it hasn't seen before. This behavior is typical of overfitting.
- **Improvement Strategies**: Implementing regularization strategies, such as L2 regularization or increasing dropout rates, might help. Additionally, refining the model architecture or using techniques like cross-validation could enhance the model's ability to generalize better.

Observations:

- **Training Accuracy**: Starts at around 60% and exhibits a steep rise initially, indicating that the model is quickly learning from the training data. It continues to increase more gradually, stabilizing at around 75%.
- **Validation Accuracy**: Also begins around 60% but increases more irregularly than the training accuracy. It peaks at about 75% and then slightly fluctuates, indicating some variability in the model's performance on new, unseen data.

Analysis:

- **Rapid Learning**: The sharp initial increase in both training and validation accuracy suggests that the model is effectively learning the patterns from the dataset.
- **Convergence and Overfitting**: The stabilization of training accuracy and the slight fluctuations in validation accuracy could suggest the beginning of overfitting, where the model performs well on training data but less so on validation data. The lack of continued increase in validation accuracy might indicate that the model is not generalizing as effectively beyond the training set.
- **Model Adjustments**: To improve the model, techniques such as adding dropout layers, adjusting the learning rate, or employing early stopping could be considered to prevent overfitting and improve validation performance.



These images illustrate the effectiveness of reconstructing retinal images, which is a crucial step in preprocessing or feature extraction processes within deep learning frameworks. This example demonstrates the capability of autoencoders to capture essential features from the retinal images, which could be critical in improving the performance of subsequent classification models.

Looking forward and Future Work

Model Optimization and Deployment

- **Further Model Refinement**: Ongoing research will explore the integration of additional cutting-edge deep learning architectures or hybrid models that could potentially increase both the accuracy and the efficiency of the diagnostic tool.
- **Real-World Testing and Deployment**: There are concrete plans to deploy this tool in clinical environments for real-world efficacy testing. This phase is critical to ensure that the tool not only performs well under test conditions but also stands up to the rigors of everyday clinical use.

References:

[1] U. R. Acharya, C. M. Lim, E. Y.K. Ng, et al. 2009. Computer-based detection of diabetes retinopathy stages using digital fundus images. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine (2009). https://doi.org/10.1243/09544119JEIM486

[2] Xuan Chen, Binh P. Nguyen, et al. 2016. Automated brain tumor segmentation using kernel dictionary learning and superpixel-level features. In Proceedings of the International Conference on Systems, Man, and Cybernetics (SMC 2016). IEEE, Budapesh, Hungary, 2547–2552.

[3] Xuan Chen, Binh P. Nguyen, et al. 2017. An automatic framework for multi-label brain tumor segmentation based on kernel sparse representation. Acta Polytechnica Hungarica 14, 1 (Apr 2017), 25–43. https://doi.org/10.12700/APH.14.1.2017.1.3

[4] Xuan Chen, Binh P. Nguyen, et al. 2017. Reworking multilabel brain tumor segmentation – An automated framework using structured kernel sparse representation. IEEE Systems, Man, and Cybernetics Magazine 3, 2 (Apr 2017), 18–22. https://doi.org/10.1109/MSMC.2017.2664158

[5] Heng-Tze Cheng, Levent Koc, et al. 2016. Wide & deep learning for recommender systems. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems (DLRS 2016). ACM, 7–10. https://doi.org/10.1145/2988450.2988454

[6] Mayo Clinic. 2018. Diabetic Retinopathy. https://www.mayoclinic.org/diseases-conditions/diabetic-retinopathy/symptoms-causes/syc-20371611

[7] Graham. 2015. Kaggle diabetic retinopathy detection competition report. https://www.kaggle.com/c/diabetic-retinopathy-detection/discussion/15801

[8] Kaggle. 2015. Diabetic Retinopathy Detection. https://www.kaggle.com/c/diabetic-retinopathy-detection/overview