Deep Learning-Based Classification for Eye Diseases using OCT Images

TECHNION Israel Institute of Technology

Master Thesis: Roni Davidov
Supervisor: Prof. Anath Fischer, Ronit Schneor

CAD Laboratory
Mechanical Engineering Faculty

Abstract

Optical Coherence Tomography (OCT) is widely used for diagnosing and monitoring retinal diseases and structural biomarkers such as Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and drusen. However, manual interpretation of OCT scans is time-consuming, subjective, and sensitive to inter-observer variability, particularly in large clinical datasets and in cases involving subtle pathological changes. This study presents a deep learning–based framework for automated classification of retinal diseases and biomarkers from OCT images. The approach progresses from single-slice (2D) analysis using CNN models to patient-level multi-slice (3D) modelling. A hierarchical classification strategy is introduced to separate initial screening from fine-grained diagnosis, reflecting clinical decision pathways and improving sensitivity in screening tasks. Multi-slice information is incorporated using hybrid CNN architectures combined with BiLSTM or Transformer-based aggregation. Experimental results demonstrate that transfer learning improves training stability, while multi-slice modelling enhances patient-level performance. Grad-CAM visualizations indicate that the model focuses on clinically relevant retinal structures.

Problem Statement

Optical Coherence Tomography (OCT) is widely used for diagnosing and monitoring retinal diseases, generating large volumes of high-resolution imaging data in routine clinical practice. Interpreting these scans requires significant clinical expertise and time, and manual analysis remains subjective and prone to inter-observer variability. Subtle pathological biomarkers, especially in early disease stages, may therefore be overlooked during large-scale screening.

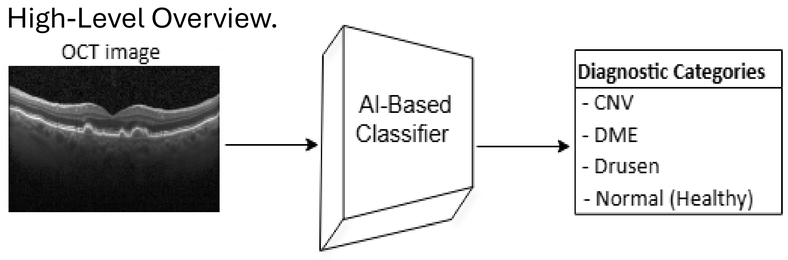
These challenges highlight the need for automated analysis tools that can support clinicians by analysing OCT data consistently, assisting in the identification of suspicious findings, and improving diagnostic reliability while preserving the clinician's role in final decision-making.

Future Work

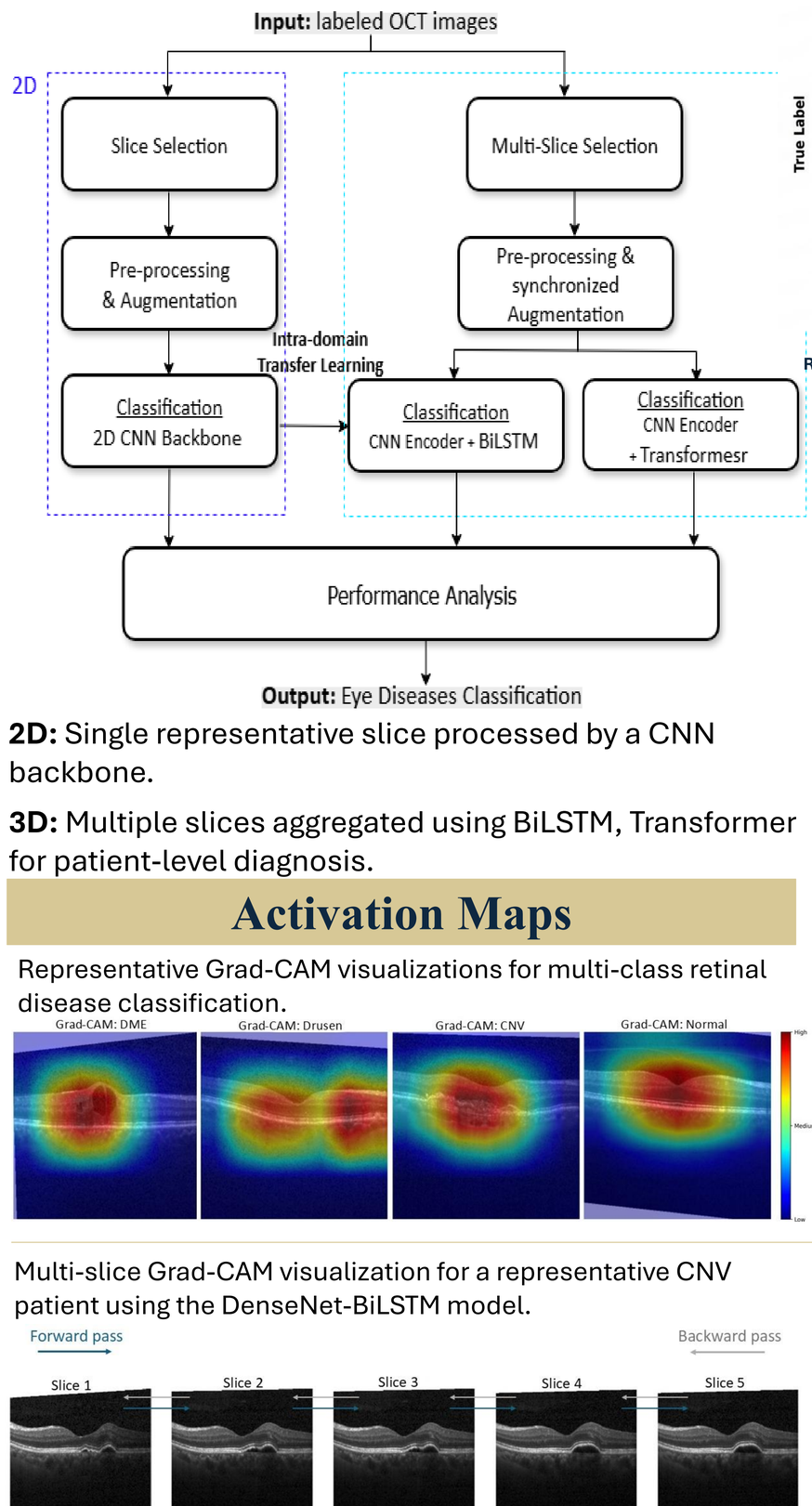
Future work will extend the proposed framework toward longitudinal OCT analysis using sequential scans from the same patient to model disease progression over time. Expert evaluation of Grad-CAM activation maps by ophthalmologists may further validate whether model decisions rely on clinically meaningful retinal structures. Future studies should also evaluate the framework on larger multi-institutional datasets to improve generalization and support clinical deployment.

Approach

High-Level Overview.



Overview of the proposed single-slice and multi-slice OCT classification framework.



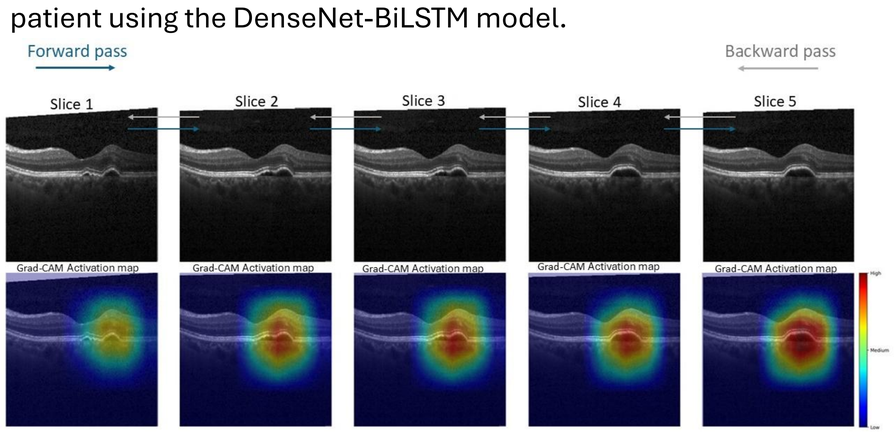
2D: Single representative slice processed by a CNN backbone.

3D: Multiple slices aggregated using BiLSTM, Transformer for patient-level diagnosis.

Activation Maps

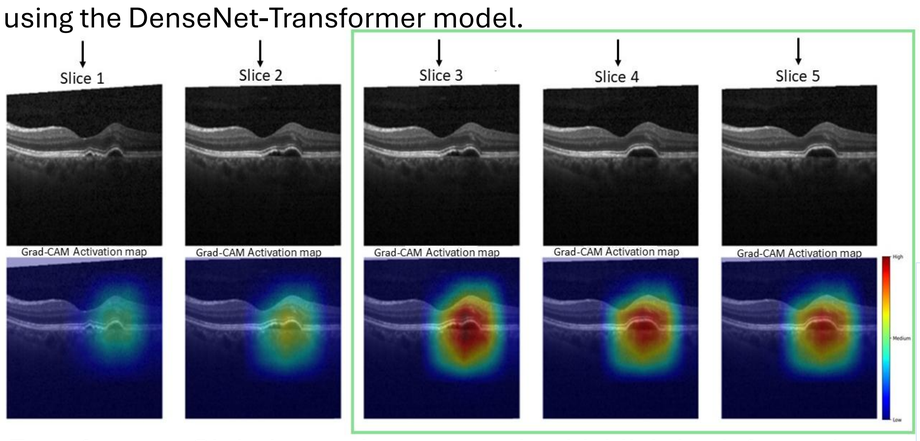
Representative Grad-CAM visualizations for multi-class retinal disease classification.

Multi-slice Grad-CAM visualization for a representative CNV patient using the DenseNet-BiLSTM model.



BiLSTM - Sequential slice aggregation.

Multi-slice Grad-CAM visualization for a representative CNV patient using the DenseNet-Transformer model.

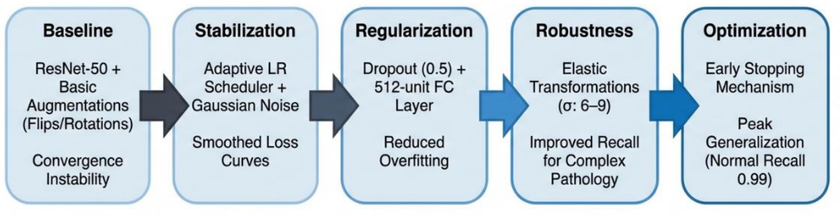


Transformer - Global attention aggregation highlighting informative slices (green box).

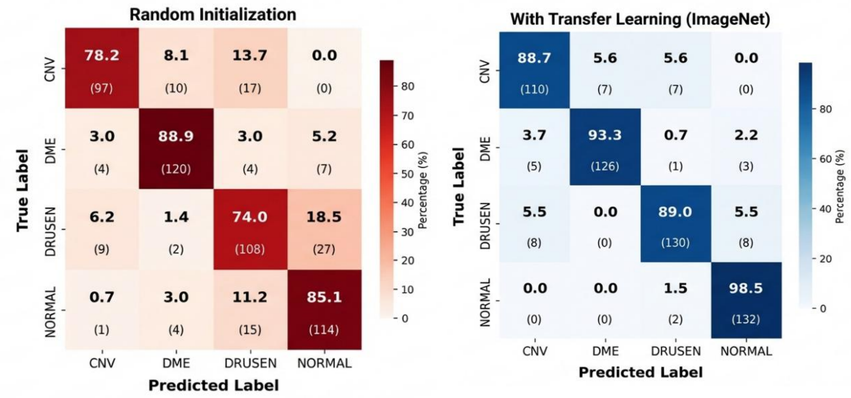
Results

Results 2D Classification - Baseline Optimization

Resnet-50 Model
Overall Accuracy: 92%



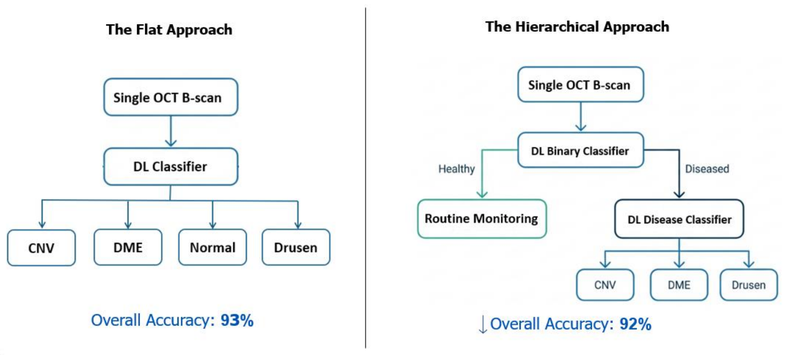
Results 2D Classification - Transfer Learning



Results 2D Classification - Architecture Comparison

ResNet-50	DenseNet-121	MobileNetV3-L
CV Accuracy: 92.8%	CV Accuracy: **93.2% ± 0.012**	CV Accuracy: 89.6%
Holdout Accuracy: 92%	Holdout Accuracy: 93%	Holdout Accuracy: 87%
Speed: 34.64 ms/slice	Speed: 37.19 ms/slice	Speed: **11.20 ms/slice**
Status: Baseline Reference	Strength: Dense feature reuse leads to highest generalization and stable feature learning.	Weakness: Accuracy drop suggests susceptibility to distribution shifts.

Results 2D Classification - Hierarchical Classification



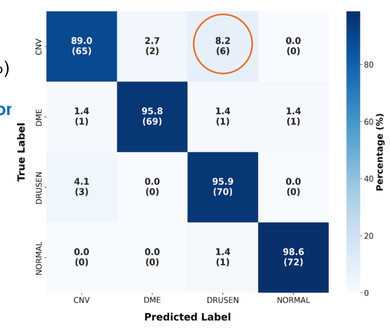
Hierarchical Classification Binary Classification

97% - Overall Accuracy

High recall for both screening classes:
- Diseased Recall: **97%** of pathological cases are successfully flagged.
- Healthy Recall: **96%** reliable filtering of normal cases.

Results 3D Classification - DenseNet-BiLSTM Architecture

- **Overall Accuracy:** 94.85%
- **High Recall:** DME (95.8%), Drusen (95.9%)
- Maintain **near-perfect Normal identification (~ 99%)**



Results 3D Classification - DenseNet-Transformer architecture

- **Overall Accuracy:** 95.53% (best)
- **High Recall:** DME (97.2%), Drusen (97.3%)
- **Improved CNV Precision:** 94% to 97%
- Maintain **near-perfect Normal identification (~ 99%)**

Summary -From Single-Slice to Multi-Slice Analysis

Model	Accuracy	Key Characteristic
2D DenseNet-121	**93.0%**	High baseline, but lacks volumetric context.
3D DenseNet-BiLSTM	**94.85%**	Strong sequential modeling; prone to CNV/Drusen confusion.
3D DenseNet-Transformer	**95.53%**	Best Performance. Global attention captures non-linear slice relationships.

References: