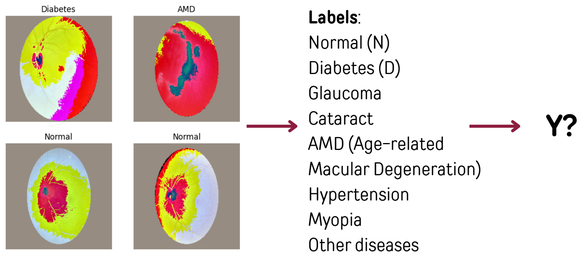
Interpretable AI for Trustworthy Ocular Disease Diagnosis

Kannak Sharma, Robotics and Autonomous Systems in Artificial Intelligence
MORE Mentor: Dr. Chris Bryan, Assistant Prof. (SCAI, ASU)
CS Advisor: Dr. Vivek Gupta, Assistant Prof. (SCAI, ASU)
Ophthalmology Advisor: Dr. Brian S. McKay, Prof.(UoA)

How can Concept Bottleneck Models (CBMs) combined with multi-modal Question Answering improve interpretability & accuracy in ocular disease classification?

Introduction

This project aims to improve ocular disease diagnosis through an **explainable**, **trustworthy interpretable** model that uses Concept Bottleneck Models (CBMs) combined with GPT-4 and domain expert-generated Q&A pairs.



What is Concept Bottleneck Model (CBM)?

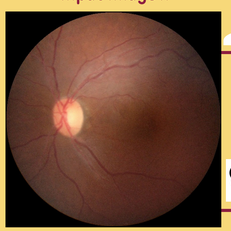


Fig. Fundus scan from ODIR-5K

Human Designed Concepts

- **Vessel Abnormalities: "hemorrhages," "blood vessels"**
- **Optic Disc Issues: "optic disc cupping," "pallor"**
- **Macula Issues: "macular edema," "retinal layers"**
- **Retinal Issues: "retinal detachment," "pigmentation"**
- **Tumor/Neovascularization: "choroidal tumor"**

LLM Generated Concepts (Q&A)

- **Are there dark spots indicating potential hemorrhage?**
- **Is the optic disc grayish or pale, suggesting optic neuropathy?**
- **Are there irregular blood vessels in the retinal region?**

Predicted Label: Diabetic Retinopathy

Prompt: Describe markers of diabetic retinopathy in a Fundus scan

References

- N. M. Selvaraj, X. Guo, A. W.-K. Kong, and A. Kot, "Improving Concept Alignment in Vision-Language Concept Bottleneck Models," Nanyang Technological University, Singapore, 2024.
- Y. Yang, A. Panagopoulou, S. Zhou, D. Jin, C. Callison-Burch, and M. Yatskar, "Language in a Bottle: Language Model Guided Concept Bottlenecks for Interpretable Image Classification," University of Pennsylvania, CVPR 2024.

Methodology

Human-Aligned Post-Hoc Explanation:

- Align with how a clinician might reason through a diagnosis
- Post-hoc explanation would highlight features as the key markers that influenced the decision.

Example:

```
if diagnosis == "Diabetic Retinopathy":
    features = ["presence of microaneurysms", "retinal hemorrhages"]
```

Trustworthy human-interpretable model

- Providing transparency in how it reaches its conclusions
- Users can comprehend the steps or logic behind the model's predictions
- Mechanisms, such as decision trees, feature importance scores, or visual prototype paths

QnA Dataset

Examples: Diabetes (D),Do the retinal vessels show signs of abnormal growth (neovascularization)?,"Yes"
Cataract (C),Is there a yellowing of the lens that may affect the quality of the image?,"Yes, the lens shows signs of yellowing."

Model Design

Data Generation and Concept Creation

Q&A pairs → GPT - 4 + Domain Expert

Bottleneck Layer Creation

Concepts evaluated using GPT-4 responses

Uniqueness | Relevancy

Challenge with Decision Trees

non-exclusive symptoms (e.g., a symptom may appear in multiple diseases). Like: Retinal Hemorrhages occur in Diabetes & Hypertension

Prioritize Features in Decision Path Construction

Initial Question: "Are there microaneurysms present?" (Highly specific to diabetic retinopathy);
Next: "Is there drusen in the retina?" (Specific to AMD)

Baseline Model Results

ResNet50 model achieved an overall accuracy of 57% and ROC AUC of 83.44, showing moderate discriminatory ability across eight eye disease.

Class	Precision	Recall	F1-Score	Support
Normal	57	8	65	1108
Diabetes	53	49	51	650
Glaucoma	52	31	39	114
Cataract	68	59	63	124
AMD	69	29	41	100
Hypertension	100	7	13	55
Myopia	86	83	84	100
Other diseases	34	10	16	306

Fig. Classification Report Table

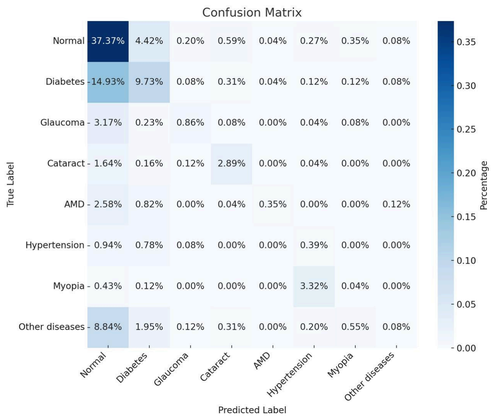


Fig. Normalised Confusion Matrix

-Myopia: Best performance (F1-score 84).
-Glaucoma and AMD: Moderate precision, low recall (F1-scores 39 and 41).
-Hypertension: High precision (100), very low recall (07), F1-score 13.
-Other Diseases: Poor results (F1-score 16), indicating classification challenges.

Results

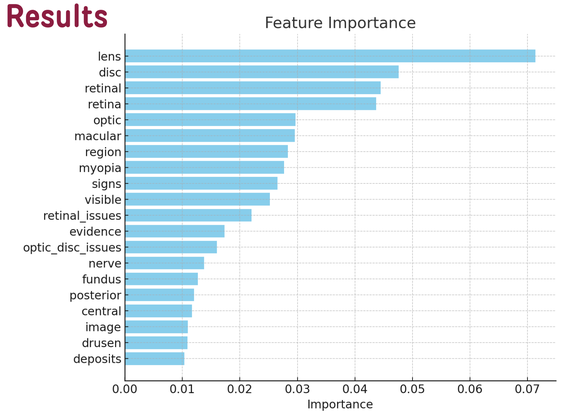


Fig. Feature Importance for Ocular Disease Detection

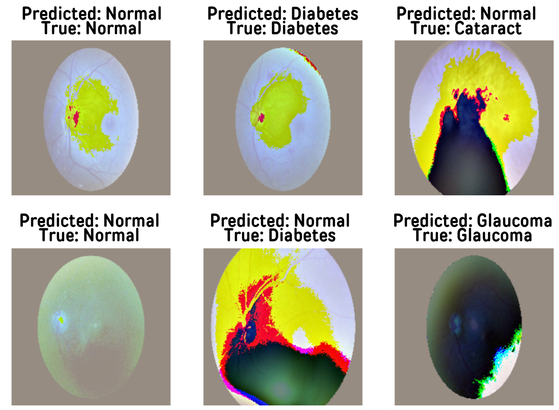


Fig. Improved Model Prediction

MORE
Masters Opportunity for Research in Engineering

Ira A. Fulton Schools of Engineering
Arizona State University