Robust Uncertainty-Informed Glaucoma Classification Under Data Shift

Homa Rashidisabet, PhD, Research Assistant Professor, University of Illinois Chicago
R. V. Paul Chan, MD, MSc, FACS; John H. Panton, MD; Thasarat Vajaranant, MD, MHA; Darvin Yi, PhD

Introduction

Glaucoma is one of the leading causes of irreversible blindness globally. Deep learning (DL) has emerged as a promising approach for the automated diagnosis of glaucoma. However, challenges persist in translating these advancements to clinical settings. Conventional DL classification methods often exhibit overconfidence and lack robustness when faced with a shift in training data distribution, posing challenges in out-of-distribution (OOD) scenarios. These issues raise concerns about the suitability of current glaucoma DL diagnostic algorithms for real-world clinical deployment, potentially impacting patient safety.

Hypothesis

Our proposed approach, centered around uncertainty quantification, aims to effectively identify OOD samples, thereby enhancing the reliability of glaucoma predictions.

Methods

As shown in Figure 1, we proposed a novel DL framework, the Dirichlet model, for joint binary glaucoma classification and OOD detection. Our method incorporates uncertainty quantification, allowing the model to express uncertainty in its predictions, thereby providing a more reliable glaucoma assessment and addressing the overconfidence commonly seen in standard DL models. We compare the OOD detection performance of the Dirichlet model to the standard softmax-based DL approach. Trained on 712 fundus images from the Illinois Eye and Ear Infirmary, we evaluate both glaucoma classification and OOD detection on the RIMONE-DL and O-RIGA fundus datasets, as well as non-medical CIFAR-10 and Fashion-MNIST datasets.

Results

The Dirichlet model consistently outperforms the softmax model in OOD detection by 9.5% to 27.5% across datasets. Table 1 shows Dirichlet achieving 82.1% and 78.7% AUC for detecting RIMONE-DL and O-RIGA glaucoma fundus datasets as OOD, with a strong performance of 100.0% AUC in detecting CIFAR-10 and Fashion-MNIST non-fundus datasets. Dirichlet maintains comparable glaucoma classification (AUC: 78.6% [78.2%, 79.1%], 54.3% [53.2%, 55.0%]) compared to softmax (AUC: 78.5% [78.2, 79.0], 60.2% [59.2%, 61.2%]) on RIMONE-DL and O-RIGA datasets.

Conclusion

The study demonstrates the effectiveness of our proposed uncertainty-aware Dirichlet model in OOD detection and glaucoma classification tasks across diverse domains, extending its utility beyond the initial training dataset. Furthermore, the incorporation of uncertainty scores in our model alerts users to instances where the model lacks sufficient information for a confident decision.

Figure(s)

Dataset	AUC% [95% confidence interval] on OOD data			
	OOD detection		Glaucoma classification	
	Dirichlet	Softmax	Dirichlet	Softmax
RIMONE-DL	**82.1 [81.6, 82.6]**	54.6 [53.6, 55.7]	54.3 [53.2, 55.0]	**60.2 [59.2, 61.2]**
O-RIGA	**78.7 [78.1, 79.3]**	69.2 [68.6, 69.9]	**78.6 [78.2, 79.1]**	78.5 [78.2, 79.0]
CIFAR10	**100.0[100.0,100.0]**	92.9 [92.8, 92.9]	-	-
Fashion-MNIST	**100.0[100.0,100.0]**	95.1 [95.0, 95.1]	-	-

Table 1. Comparison of performance between the proposed Dirichlet model and the standard softmax-based deep learning model across external out-of-distribution (OOD) test datasets. The best performance for each dataset is highlighted in bold.

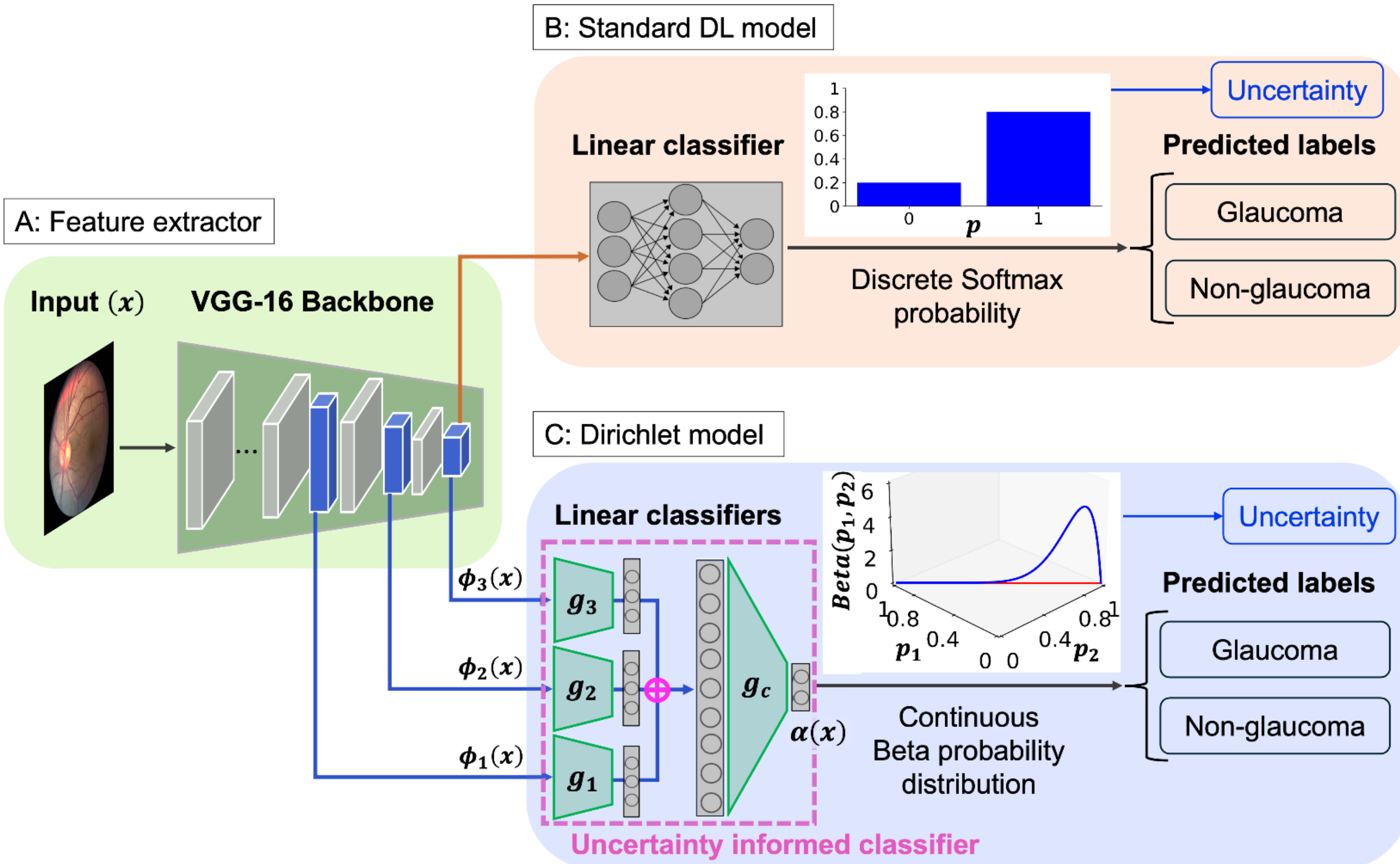

Figure 1. Glaucoma classification pipeline. (A) Feature extraction backbone using the VGG-16 architecture, (B) standard deep learning framework using the softmax function, and (C) our proposed Dirichlet model framework leveraging evidential deep learning for uncertainty quantification.

Keywords

Artificial Intelligence/Machine Learning