Analyzing clinical variables indicative of uveal melanoma to determine how they affect decisions made from an artificial intelligence classifier

Emily A. Laycock;[1] Ezekiel Weis, MD, MPH; [1] Trafford Crump, PhD [1,2]
1. University of Calgary Cumming School of Medicine; 2. University of Calgary Schulich School of Engineering

UNIVERSITY OF CALGARY

Introduction

Figure 1: Fundus image of UM

- Choroidal nevus (CN): an intraocular melanocytic lesion with malignant potential
- Uveal melanoma (UM): can develop from a CN, most common primary intraocular cancer in adults
- 45% mortality rate within 15 years of UM diagnosis[1]

Diagnostic ambiguity + Shortage of specialists = Delay in diagnosis ➡ Poor prognosis

- Early detection → earlier referral for treatment
- Fundus images → used to train artificial intelligence (AI) model to detect presence of lesion → mechanize skill set of ocular oncologists → faster detection, better prognosis

Problem: AI can generate false negative (FN) diagnoses, particularly from fundus images

FN: to not detect the presence of a lesion when one is present in the image → problematic

Objective

To determine if there are certain features associated with the lesion that cause the AI model misclassify an image as FN.

Methods

Model
- Transfer learning pre-trained model
- Test on eye lesions

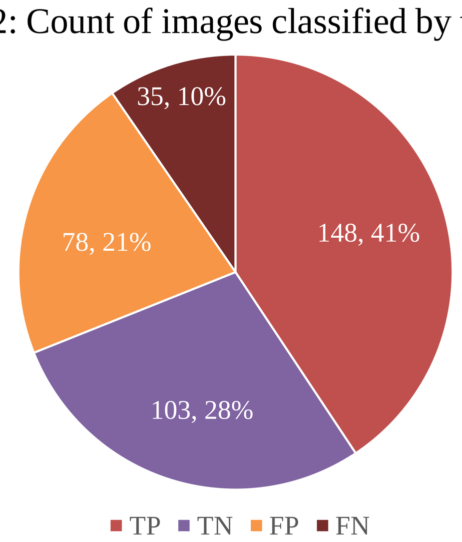
Dataset
- Fundus images labeled "lesion present", "lesion absent"
- Collected from Alberta Ocular Brachytherapy Program in Edmonton, AB
- Abstracted charts from patient EMR and fundus images

Statistical Analysis

To determine if there are any statistically significant relationships between variables and the outcome of the AI classification.
- Univariate Logistic Regression - determine the individual effect of each variable on image classification
- Multivariate Logistic Regression - determine the combined effect of variables on image classification

Results

Figure 2: Count of images classified by the model



Demographics			**Clinical features of lesion**		
Mean **age** (SD) in years	62.5 (14.6)		**Localization** of epicentre	Macula: 65 (35.5%)	Peripheral: 118 (64.5%)
Sex	Male: 58 (31.7%)	Female: 125 (68.3%)	Largest **diameter** (SD) in mm	3.7 (2.4)	
Study **eye**	Right: 84 (45.9%)	Left: 99 (54.1%)	**Thickness** (SD) in mm	1.6 (0.2)	
Visual **acuity** (SD)	30 (16.8)		Presence of **orange pigment**	Yes: 14 (7.7%)	No: 169 (92.3%)
			Presence of **subretinal fluid**	Yes: 10 (5.5%)	No: 173 (94.5%)
			Presence of **drusen**	Yes: 90 (49.2%)	No: 93 (51.8%)
			Hollow	Yes: 4 (2.2%)	No: 179 (97.8%)
			If lesion is **100% visible** in image	Yes: 179 (97.8%)	No: 179 (97.8%)
			Fully **pigmented**	Yes: 147 (80%)	No: 36 (20%)
			AI classification	TP: 148 (81%)	FN: 35 (19%)

Table 1: Descriptive statistics of variables collected from the dataset (n = 183)

Variable	Estimate (SE)	P-value
Sex	0.015 (0.405)	0.97
Study eye	0.009 (0.377)	0.98
Age	0.026 (0.015)	0.084
Location	0.677 (0.381)	0.076
Diameter	-0.227 (0.0999)	0.023
Thickness	-0.724 (1.049)	0.49
Orange pigment	-0.376 (0.788)	0.634
Subretinal fluid	1.116 (0.625)	0.098
Drusen	-0.916 (0.310)	0.022
Hollowness	-15.152 (1199.77)	0.99
100% visibility	15.15 (1199.77)	0.99
Pigmentation	2.398 (0.432)	>0.001

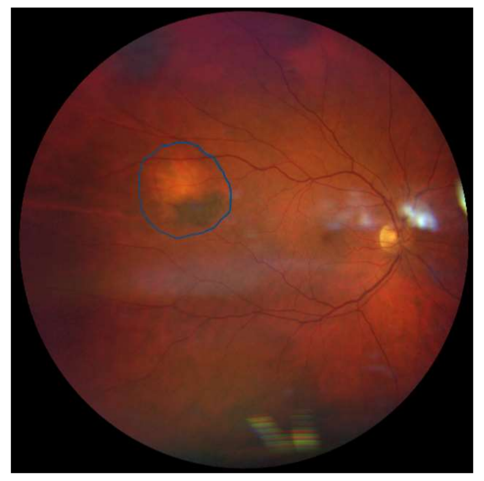
Table 2: Univariate logistic regression of AI classification and collected variables. Variables that scored a p-value <0.1 are in blue and variables that scored a p-value <0.05 are in red.

Variable	Estimate (SE)	P-value
Age	0.029 (0.018)	0.098
Location	0.021 (0.474)	0.965
Diameter	**-0.248 (0.124)**	**0.047**
Subretinal fluid	1.531 (0.856)	0.074
Drusen	**-1.245 (0.521)**	**0.017**
Pigmentation	**2.827 (0.526)**	**>0.001**

Table 3: Multivariate logistic regression of AI classification and collected variables. Variables that scored a p-value <0.1 from the univariate logistic regression were included. Variables that scored a p-value <0.05 are in red.

Discussion

Non- or variable pigmentation → FN



If non-pigmented: lacks melanin → blends into background, harder for model to locate

Smaller diameter → FN

Smaller lesion → harder for model to locate

Absence of drusen → FN

Drusen: protein deposits from degraded photoreceptors, manifests as discrete white build-up → useful factor for model to identify a lesion

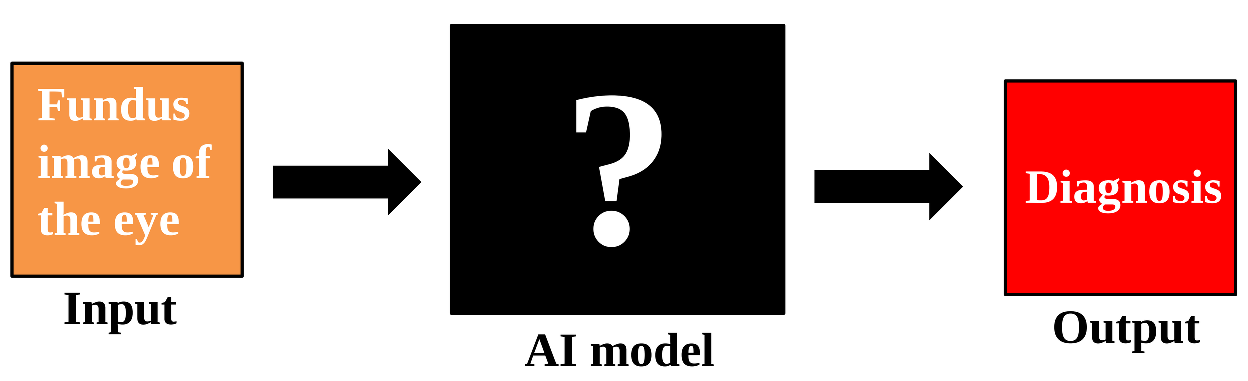
Implications of AI image diagnosis usage in clinic

- Physicians' prediction accuracy improves when working with classification model[2] → future iterations of our model could be a helpful tool to streamline diagnostic process
- Sensitivity may decrease[2] → physicians over-rely on AI predictions → training on proper use of AI in clinic needed

Contribution to AI interpretability

- Gives insight on how model is making decisions, black box → glass box
  - Which variables associated with lesion contribute to misdiagnosis
- Increases patient/physician trust in AI
- Provides information on how to improve model → training sets for future iterations must include small lesions, non-pigmented lesions, lesions without drusen

Black box phenomenon



Considerations

Limitations
- Small sample size
- Uses early iteration of model

Strengths
- Uses real-world data → generalizability
- Contributes to AI interpretability → patient/physician trust in AI

Future Directions

- Increase sample size and test on future iterations of model
- Test of fundus images from other eye care centers → account for other methods of taking fundus images
- Detecting lesions → differentiating between CN and UM

References

1. Weis E, Salopek TG, McKinnon JG, Larocque MP, Temple-Oberle C, Cheng T, et al. Management of uveal melanoma: a consensus-based provincial clinical practice guideline. Curr Oncol. 2016 Feb;23(1):e57-64.
2. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep Learning in Medical Image Analysis. Adv Exp Med Biol. 2020;1213:3-21.

For further information contact Emily Laycock at emily.laycock1@ucalgary.ca
Funding provided by VPR Catalyst Grant