CASEY EYE Institute
OHSU

Advanced Analytics in Ophthalmology: Expert Systems

Michael F. Chiang, MD
Knowles Professor of Ophthalmology & Medical Informatics and Clinical Epidemiology
Vice-Chair (Research), Department of Ophthalmology
Casey Eye Institute at Oregon Health & Science University

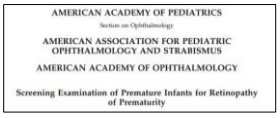
Disclosures & Collaborators

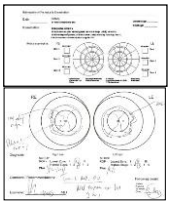
- Imaging & Informatics in ROP (i-ROP)
- NIH: R01EY19474, R21EY22387, P30EY10572
- NSF: SCH-1622679
- Research to Prevent Blindness
- Clarity Medical Systems (unpaid member of Scientific Advisory Board)
- Novartis (Steering Committee Member of RAINBOW study)
- Past Chair, AAO Medication Information Technology Committee
- Member, AAO Board of Trustees
- Chair, AAO IRIS Registry Data Analytics Committee

Example: Retinopathy of Prematurity

- Leading cause of childhood blindness
  - Treatable if early Dx (CRYO-ROP, ETROP, BEAT-ROP)
  - USA: 40,000 cases/year, 600 blind/year
  - Economic impact
- AAP-AAO guidelines (2001): "done with **indirect ophthalmoscopy**" in NICU
  - Documentation: **hand-drawn sketches**
- **Challenges** of dogma (practical):
  - Time-intensive: travel, coordination
  - Exam: Difficult, imprecise, subjective
  - More infants at risk (survival)
  - Medicolegal liability
  - **Limited access to care & training, especially in rural & underserved areas**

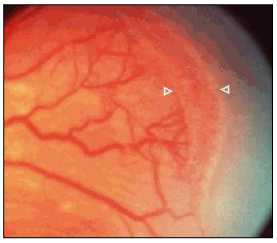
Fierson et al, Pediatrics 2001; 108:809-11

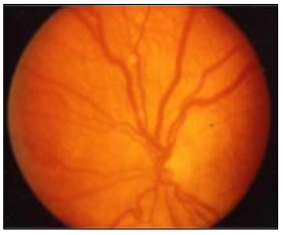

AMERICAN ACADEMY OF PEDIATRICS
Section on Ophthalmology
AMERICAN ASSOCIATION FOR PEDIATRIC OPHTHALMOLOGY AND STRABISMUS
AMERICAN ACADEMY OF OPHTHALMOLOGY
Screening Examination of Premature Infants for Retinopathy of Prematurity



Diagnosis: Gold Standards & ICROP

- Originally: descriptive, unstructured
- ICROP (1984):
  - International standard for clinical exams, **infrastructure for multicenter clinical trials**
  - Parameters: zone (I-III), stage (1-5), extent (clock hours), plus disease
  - Most fields don't have this standardized terminology…
  - CRYO-ROP, ETROP: **plus disease** is most critical parameter for severe treatment-requiring ROP → "arterial tortuosity & venous dilation" (**standard published photo**)


Stage 3


Plus disease

ICROP. Arch Ophthalmol 1984; 102:1130-4

Challenge: Diagnostic Accuracy

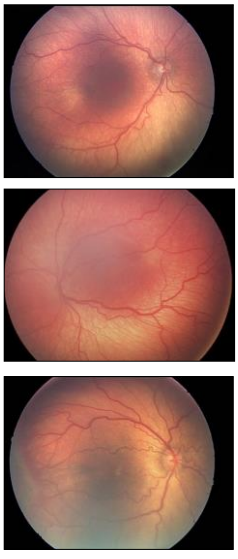
- 3 (14%) experts: "Plus"
- 18 (86%) experts: "Not Plus"

- 11 (52%) experts: "Plus"
- 10 (48%) experts: "Not Plus"

Chiang et al, Arch Ophthalmol 2007; 125: 875-80.

Science & Art of Medicine

- So what **is** plus disease:
  - Like pornography: "can't define, but know it when I see it"
  - Is "arterial tortuosity & venous dilation" in "central retina" an over-simplification?
  - Could this explain **variability**?
- Capture & encode detailed qualitative thoughts of 7 experts during plus disease diagnosis:
  - Videotaped while reviewing 7 images: (1) think-aloud protocol, (2) specific questions



Hewing et al, JAMA Ophthalmol 2013; 131:1026-32.

Challenge: Disagreement in Process

- Expert 1: Diagnosis **Plus Disease**

… looks like a very low gestational birth baby, it's taken quite a long time to get to this stage. There is a lot of **arterial tortuosity**, there is a little bit of **venous congestion** in the **superior temporal** and **superior nasal** quadrant, more in the **superior half** of the retina. By definition I think this has to be **plus**, because it's two quadrants at least, and even the other quadrants aren't normal…

… I don't know whether the **peripheral disease** is that bad, it may not be actually, could be…

- Expert 2: Diagnosis **Pre-Plus Disease**

… there is a lot of **tortuosity of the arteries**, the **veins are about 2 to 1**. This could either be a **pre-plus** eye or a **normal variant**, depending on a quick look at the periphery…

… curiously there is a lot of tortuosity **down here (inferior)**, it **looks like there is disease up there**…

… the fact that tortuosity is everywhere, you want to make sure if it's a congenital tortuosity kid.

… I would suspect **pre-plus**, could also be a normal variant.

- Expert 4: Diagnosis **Neither Pre-Plus nor Plus Disease**

…vessels seem to be **branching excessively** in that region (superonasal) and some **increased tortuosity** (superotemporal) as well, and this **vein looks too fat** (superotemporal)…

… if all the quadrants were like this quadrant (superotemporal) then it would be at least pre-plus and verging on plus, but since it's **only one quadrant** that's highly questionable…

… would not classify it as plus, I could see why some would call it pre-plus, I would not call it pre-plus, I would call it **no plus**.

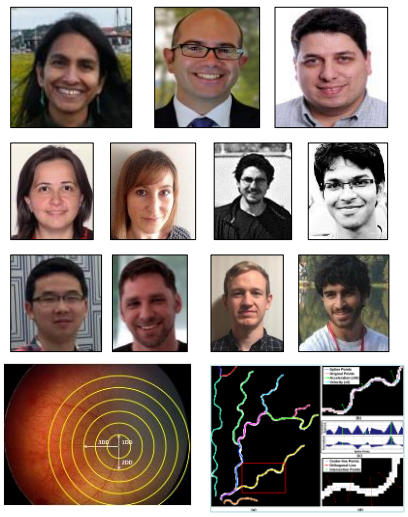
Features Mentioned by Experts

Feature	Number of Mentions
Arterial tortuosity	42/42
Arterial dilation	8/42
Venous tortuosity	10/42
Venous dilation	42/42
Central vessels	8/42
Peripheral vessels	14/42
Number of quadrants of abnormality	23/42
Vascular branching	8/42
Macular features	3/42
Other vascular features	7/42

Hewing et al, JAMA Ophthalmol 2013; 131:1026-32.

Approach: Retinal Image Analysis

- **Goal**: more accurate diagnosis by **quantifying** vascular parameters with image analysis
- Accurate **segmentation** of vessels from images
- Validation against robust **reference standard**
- Which **image features** (e.g. tortuosity, branching) are the key ones? How to quantify?
  - Strategy #1: Classic machine learning methods
  - Strategy #2: Convolutional neural networks ("deep learning")



Ryan MC et al, AMIA Proc Annu Symp, 2014; 1902-10
Campbell et al. JAMA Ophthalmol 2016; 134:651-7.

Machine Learning Approach

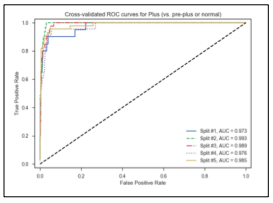
Classifier	Accuracy (vs. RSD)
Expert 1	64/73 (87%)
Expert 2	63/73 (86%)
Expert 3	58/73 (79%)
Expert 4	72/73 (99%)
Expert 5	64/73 (88%)
Expert 6	62/73 (85%)
Expert 7	68/73 (93%)
Expert 8	64/73 (88%)
Expert Consensus	71/73 (97%)
Computer System	69/73 (95%)

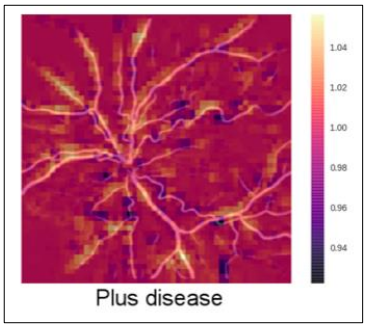
- **Manual image segmentation**
- Reference standard: combines image reading & ophthalmoscopic diagnosis
- Best performance with 6DD circular crop, **acceleration** metric
- Variable expert accuracy (79-99%)
- High computer system accuracy (95%)

Ataer-Cansizoglu et al, Trans Vis Sci Technol 2015; 4:5

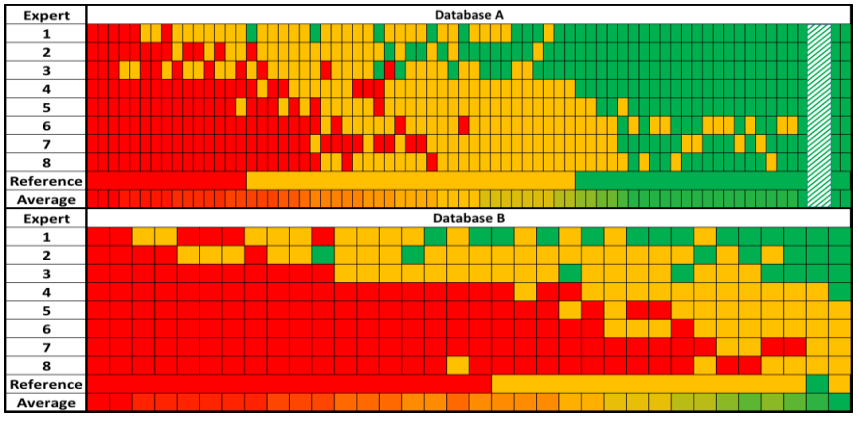
Deep Learning for ROP

- Used for diabetic retinopathy (JAMA), skin cancer (Nature), AMD
- Train **fully-automated** CNN for ROP → 6000 posterior pole images, each with reference standard (plus vs. pre-plus vs. normal)
  - AUC 0.98 to identify plus disease
- Independent test set: 91% accuracy (8 experts: mean 82% accuracy, range 77-94%)
- Occlusion analysis: what parts of image did experts use?





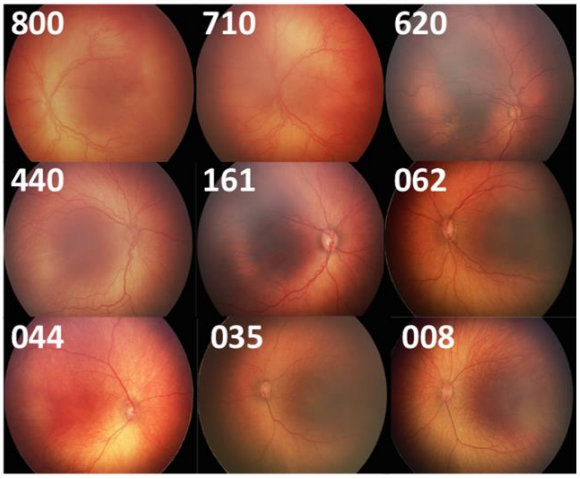
Inter-Expert Variability: Spectrum

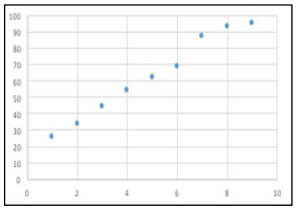


- Under-callers vs. over-callers (consistent across multiple data sets)
- **Continuous spectrum** of abnormality: over-simplified by categories
- Experts: **good at comparisons**, but **bad at labeling** (drawing lines)

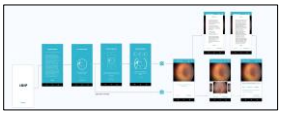
Campbell et al, Ophthalmology 2016;123:2338-44.

Continuous Spectrum of Abnormality





i-ROP Mobile App (Apple & Android)



Kalpathy-Cramer et al, Ophthalmology 2016;123:2345-51.

Key Points for FDA: Expert Systems

- Ophthalmic diagnosis is inherently subjective & qualitative: ROP (tortuous?), diabetic retinopathy (NV?)
  - Significant inconsistency, even among experts ("drawing the lines") → performance of "real-world" physicians may be worse, unclear impact of "clinical judgment"
  - **Potential role for expert systems to improve consistency**
  - **Bar for systems should be "human-like", not "perfection"**
  - **Validation: requires transparency, cannot use only a single human**
- Rapidly changing field: systems may undergo regular cycles of improvement (e.g. training with new data, better algorithms)
  - **Ideal to have efficient mechanism for "upgrades"**
- Does intended use of systems matter: advice to physicians ("decision support") vs. closed-loop system (e.g. screening for primary care)
  - Many real-world examples of the former outside FDA purview (e.g. EHRs)
  - **I hope FDA will consider different levels of oversight based on use**