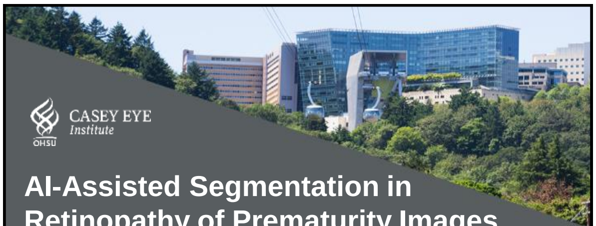


CASEY EYE Institute

AI-Assisted Segmentation in Retinopathy of Prematurity Images

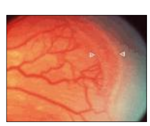
Michael F. Chiang, MD
Knowles Professor of Ophthalmology & Medical Informatics and Clinical Epidemiology
Associate Director, Casey Eye Institute
Oregon Health & Science University

Disclosures & Collaborators

- Imaging & Informatics in ROP (i-ROP)
- NIH (R01EY19474, R21EY22387, P30EY105072, K12EY27720, T32EY23211), NSF (SCH-1622679), Research to Prevent Blindness
- Clarity Medical Systems (unpaid member of Scientific Advisory Board), Novartis (RAINBOW Steering Committee), Inteleretina (member)
- AAO Board of Trustees, Telemedicine Task Force, IRIS Registry Data Analytics Task Force (Chair), IRIS Registry Executive Committee

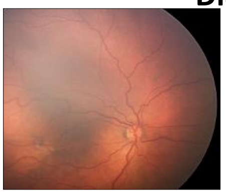
Retinopathy of Prematurity (ROP)

- Leading cause of childhood blindness
  - Bedside ophthalmsocopy in NICU
  - Very limited access to care
- ICROP (1984):
  - International standard for clinical exams, **infrastructure for multicenter trials**
  - Parameters: zone (I-III), stage (1-5), extent (clock hours), plus disease
  - Many fields don't have this standardized terminology…
  - Clinical trials: **plus disease** is most critical parameter for treatment-requiring ROP → "arterial tortuosity & venous dilation" (**standard published photo**)



ICROP. Arch Ophthalmol 1984; 102:1130-4

Challenge: Disagreement in Diagnosis



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

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

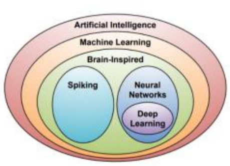
Chiang et al, Arch Ophthalmol 2007; 125: 875-80.

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**.

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

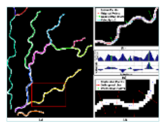
Approach: Artificial Intelligence



Sze *et al.* (2017)

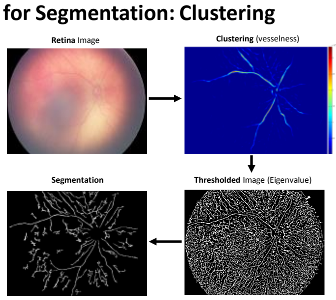
Machine Learning Overview

- **Segmentation**
- Feature extraction
  - Examples: vascular curvature, branching, dilation
- Feature representation
  - Combine image features (e.g. mean, two largest values, Gaussian mixed models)
- Classification
  - Examples: support vector machine, K-nearest neighbors



Machine Learning for Segmentation: Clustering

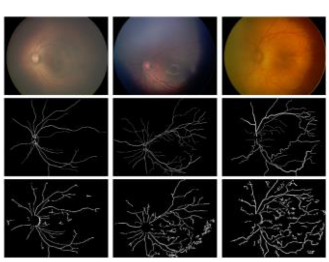
- Original retinal image
  - Pre-processing to emphasize vessels
- Clustering algorithm
  - Vessels vs. not vessel (e.g. Gaussian mixed model, Frangi filter)
- Thresholding
  - Foreground vs. background (B&W)
- Post-processing
  - Remove spurious areas



Ataer-Cansizoglu, Pattern Recognition Letters 2012; 46: 1140-50.

Machine Learning Segmentation: Results

- **Reference standard:** manual segmentation by experts (100 images)
- **Performance:**
  - Accuracy: 0.94 ± 0.02
  - Sensitivity: 0.64 ± 0.05
  - Specificity: 0.95 ± 0.02



Original image

Manual segmentation

Automated segmentation

Ataer-Cansizoglu, Pattern Recognition Letters 2012; 46: 1140-50.

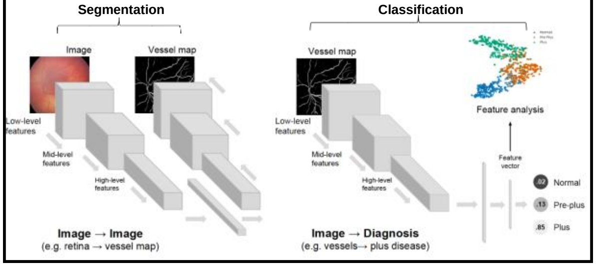
Machine Learning: Diagnostic Classification

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%)

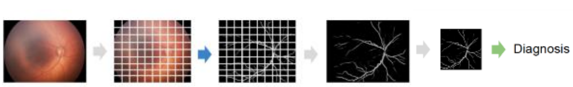
- **Manual image segmentation**
- Reference standard: combines image reading & ophthalmoscopic diagnosis
- Best performance with 6DD circular crop, **acceleration** feature
- Combination of features using GMM approach, SVM classifier

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

Deep Learning Overview



Deep Learning for Segmentation



Vessel segmentation

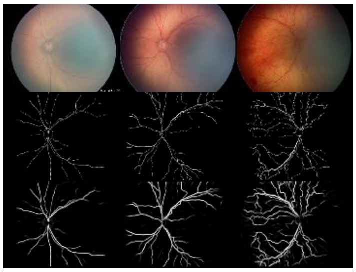
- 200 retinal photographs
  - 640 x 480 color images
  - Annotated in Adobe Photoshop
- Broken up into 48x48 patches
- U-net architecture

Plus disease diagnosis

- 5,511 retinal photographs
  - Reference standard diagnosis (3)
  - Pre-processed using U-net
- Resized and cropped to 224x224
- Inception v1 architecture

Brown et al. JAMA Ophthalmology 2018; 136:803-10.

Deep Learning for Segmentation: Examples

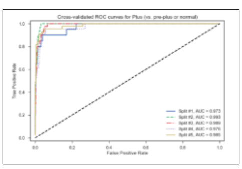


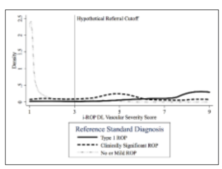
Original image

Manual segmentation

Automated segmentation

Deep Learning: Diagnostic Classification

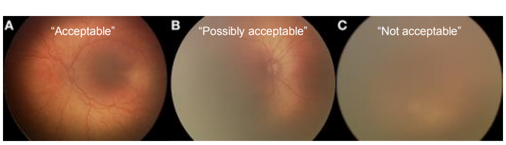




- Fully-automated CNN: **AUC 0.98** for diagnosis of plus disease (5-fold cross-validation)
- Independent test set (100 images): **91% accuracy** (8 experts: mean 82% accuracy, range 77-94%, outperformed 7/8 experts)
- **Quantitative severity score**: potential for disease screening & prediction

Brown et al. JAMA Ophthalmology 2018; 136:803-10.
Redd et al. Br J Ophthalmol 2019. In press.

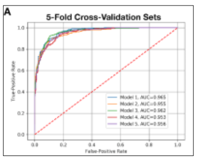
What About Image Quality?

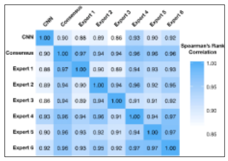


- Varying levels of real-world image quality → 6,139 posterior images graded by 3 experts, including quality metric
- Train CNN (Inception V3, weights initialized after training with ImageNet) to identify Acceptable quality images, 5-fold cross-validation on 4,000 images (remainder as independent test set)

Coyner, et al. Ophthalmol Retina 2019. In press.

Deep Learning for Image Quality





- AUC 0.959 for Adequate quality images (5-fold cross validation), 0.965 (test set)
- 30 images rank ordered from lowest to highest quality (6 experts): Spearman correlation coefficient = 0.90 compared to overall consensus rank ordering

Coyner, et al. Ophthalmol Retina 2019. In press.

Summary

- Ophthalmic diagnosis is subjective & qualitative
  - Significant inconsistency in both diagnostic **classification and process**
  - Potential role of artificial intelligence to improve consistency
  - Bar for systems should be "human-like", and validation requires multiple experts
- Role of artificial intelligence in image segmentation & image quality
  - Significantly better performance of deep learning methods for vessel segmentation
  - But critical importance of **explainability** (what it means to "look bad"), and evidence that feature extraction is still extremely important
- Diagnostic **classification vs. screening**
  - Importance of differing levels of FDA oversight based on intended use