

IMAGE SEQUENCE ANALYSIS VIA GRU AND ATTENTION FOR TRACHOMATOUS TRICHIASIS CLASSIFICATION Presented by: Juan C. Prieto¹

Collaborators: Hina Shah¹, Kasey Jones², Robert F. Chew², Hashiya M., Kana¹, Jerusha Weaver¹, Rebecca M. Flueckiger⁴, Scott McPherson⁴, Emily W. Gower^{3,4}

Department of Psychiatry, UNC Chapel Hill, NC¹

Center for Data Science, RTI International, Research Triangle Park, NC^2

Department of Epidemiology, UNC Chapel Hill, NC³

Global Health Division, RTI International, Research Triangle Park^4



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL



MOTIVATION

- Trachomatus Trichiasis (TT) is a condition caused by chlamydia trachomatous (an infectious ocular condition) where the eyelid turns inward leading to eventual blindness

- Need to bridge the gap between community screeners and experts in assessing TT in rural areas.



APPROACH Х_{t-1} х_{t+1} 3024 x 3024 3024 x 3024 512 x 512 512 x 512 FE FE FE N*2048 Eyelid's UNET polynomial RNN fitting GRU_{6,512} b_h.512 GRU, GRU, 32x384x384 GRU, 512 GRU, GRU, f_h.512 Xht-1,1024 X_{ht+1} .N*1024 Polynomial fit ttAttNet with eyelid Attention, (b, x, Attention (f, x, a, 1024 **a**_b.1024 segmentation ,3072 Dense,2 TT a Generate Sequence concatenate

RESULTS

DATA

- Images collected as part of Maximizing Trichiasis Surgery Success Trial in Ethiopia
- Images captured on commercial mobile phones
- Segmented and graded for TT by expert ophthalmologist
- Segmentation training: 805 images
- Classification training set: 1,398 images
- Hold out test set (for both classification and segmentation): 308 images

Label	Precision	Recall	F1-score	Dice			
Background	0.99	0.98	0.99	0.98			
Sclera	0.85	0.80	0.83	0.82			
Cornea	0.85	0.91	0.98	0.88			
Upper Eyelid	0.70	0.73	0.71	0.71			

Segmentation scores









Classification scores on various architectures

NN	input	class	precision	recall	f1-score	accuracy
ttVGG19 ₅₁₂	16x16x512	normal	0.84	0.72	0.78	0.83
		tt	0.81	0.91	0.86	
$ttResNet50_{512}$	16x16x2048	normal	0.90	0.73	0.81	0.85
		tt	0.82	0.94	0.87	
ttMobileNetV 2_{512}	16x16x1280	normal	0.87	0.70	0.78	0.82
		tt	0.79	0.92	0.85	
$\operatorname{ttAttNet}_{VGG19}$	32x512	normal	0.78	0.84	0.81	0.87
		tt	0.92	0.88	0.90	
$\mathrm{ttAttNet}_{ResNet50}$	32x2048	normal	0.87	0.86	0.87	0.91
		tt	0.92	0.93	0.93	
$ttAttNet_{MobileNetV2}$	32x1280	normal	0.83	0.81	0.82	0.87
		tt	0.89	0.91	0.90	

Sub-images with top attention weights



CONCLUSIONS

- Approach to analyze high resolution images of eyes captured by commercial phones and identify TT
- Different architectures are evaluated using feature extraction/pre-trained networks
 - Resampled images 512x512
 - Cropped concatenated image patches and sequence analysis
 - Segmentation and attention weights provide feedback and insight about the predictions
- High accuracies are achieved

Feature extraction algorithm	Classification accuracy	Initial estimate of computation time for full pipeline (Samsung Galaxy S8)
VGG19	0.87	22.2 s
Resnet50	0.91	14.71 s
MobilenetV2	0.87	8 s

