

# Analysis of agreement of clinical experts to B-scan of interest algorithm



Poster #3544502

Mahsa Darvishzadeh-Varcheie, PhD<sup>1</sup>; Katherine Makedonsky, OD<sup>1</sup>; Gary C. Lee, PhD<sup>1</sup>; Vidya Kulkarni, BS<sup>1</sup>; Hugang Ren, PhD<sup>1</sup>; Aaron Lech, OD<sup>2</sup>; Laura Tracewell, OD<sup>1</sup>; Rishi P. Singh, MD<sup>3</sup>; Katherine Talcott, MD<sup>3</sup>; Niranchana Manivannan, PhD<sup>1</sup>; Mary K. Durbin, PhD<sup>1</sup>

<sup>1</sup>Carl Zeiss Meditec, Inc., Dublin, CA, USA; <sup>2</sup>ClearVue Eye Care, Roseville, CA, USA; <sup>3</sup>Center for Ophthalmic Bioinformatics, Cleveland Clinic Cole Eye Institute, Cleveland, OH, USA

## PURPOSE

- Optical coherence tomography (OCT) scans are important for clinical evaluation of the retina, but manual evaluation of each B-scan is time consuming and requires significant expertise.
- To improve the workflow, a B-scan of interest tool (BSOI) was created using deep learning to indicate B-scans of interest as well as B-scans of poor image quality (IQ).
- The purpose of this study was to determine if the BSOI tool agreed with doctors' clinical judgement.

## METHODS

- Study description:
  - Participants: 47 subjects with various pathologies including macular degeneration, diabetic eye disease, epiretinal membranes, macular hole and central serous retinopathy.
  - Imaging device: CIRRUS™ 6000 (ZEISS, Dublin, CA).
  - Scan type: Macular Cube 512x128 scans.
  - The BSOI tool first runs an IQ algorithm which flags B-scans with poor IQ then further flags good IQ scans if it finds one of the retinal pathologies as referenced in [1-2].
  - The prototype tool was integrated into the macular thickness analysis report, flagging B-scans of interest with a red indicator and poor IQ scans with a yellow indicator.
- Grading process:
  - Graders: total of 5 independent experts: 2 retina specialists (RS) and 3 optometrists (OD).
  - Graders evaluated the tool results and used a feedback feature to indicate if they agreed (thumbs up) or disagreed (thumbs down) (see Figure 2).
  - Graders also answered a survey rating the clinical utility and loading time of the tool.
  - Intraclass correlation coefficient (ICC) and pair-wise Cohen's kappa between graders were used to assess inter-grader consensus.

The B-scan of interest tool successfully flags B-scans that may have otherwise been missed by doctors.

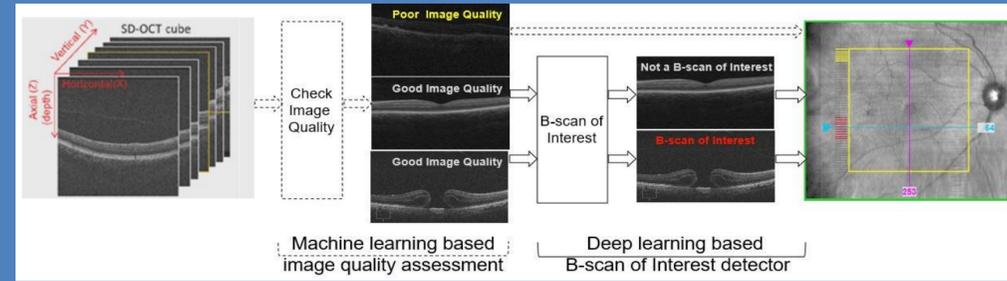


Figure 1. The tool checks OCT B-scans for image quality and flags for potential clinical interest.

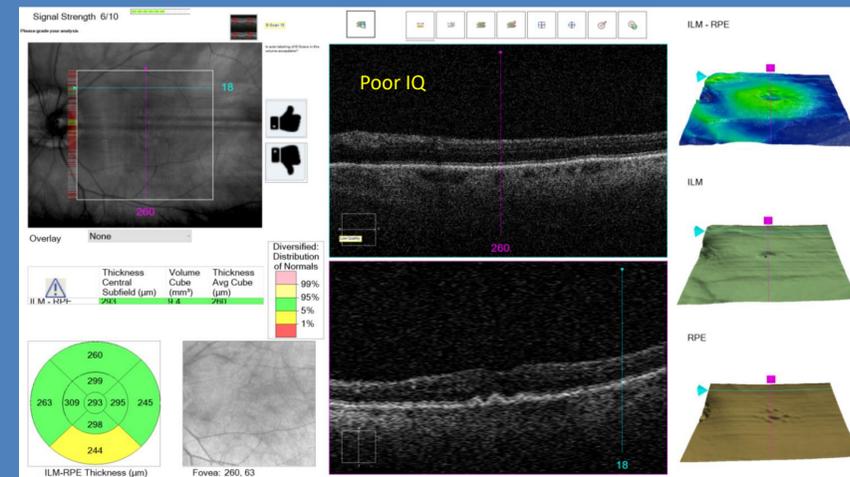
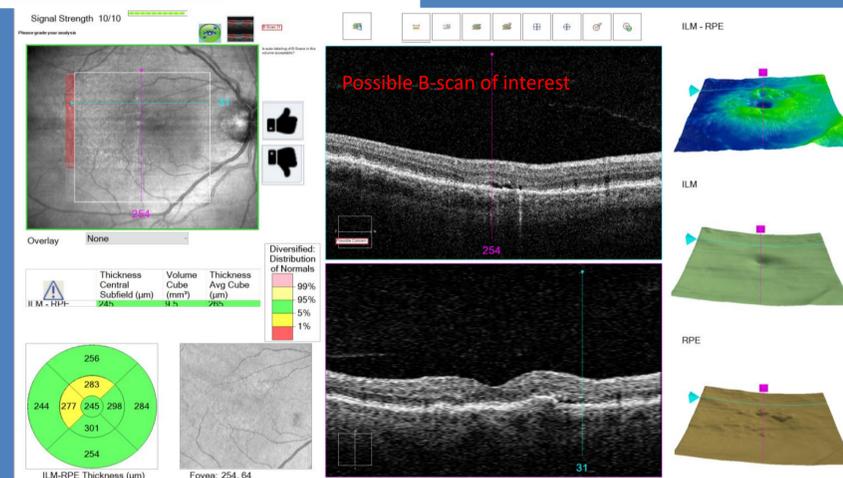


Figure 2. B-scan of interest tool grading interface with examples: poor image quality and flagged B-scans (left), and flagged B-scans (bottom)



Email: mahsa.darvishzadeh@zeiss.com

Disclosures: MD (E), KM (E), GL (E), VK (E), HR (E), AL (C), LT (C), RS (C), KT (F), NM (E), MD (E) – Carl Zeiss Meditec, Inc.

## RESULTS

- Mean agreement of an individual grader with the tool was 79% and majority vote agreement between graders and the tool was 81%.
- The ICC with 95% confidence interval was 0.44 (0.32, 0.60) indicating poor reliability of agreement between all doctors.
- Kappas (0.25 to 0.67) showed fair to moderate agreement between pairwise doctors.

Grader Agreement						
Grader	1	2	3	4	5	Grader Consensus
Designation	OD1	RS1	RS2	OD2	OD3	
Observed Agreement (Thumbs up)	0.83 (0.70, 0.93)	0.83 (0.69, 0.92)	0.79 (0.64, 0.90)	0.79 (0.64, 0.90)	0.7 (0.55, 0.83)	0.81 (0.67, 0.91)

Table 1. Observed agreement between graders and tool (top) and survey results (bottom)

Survey	
Does the tool increase your workflow efficiency? (Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree)	<b>Answers: 80% Agree, 20% Strongly Agree</b>
Will the tool be useful in aiding your diagnosis? (Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree)	<b>Answers: 80% Agree, 20% Neutral</b>
In your practice, would you expect your technicians to use the tool or the doctor only? (Tech, Doctor, Both)	<b>Answers: 60% Doctors, 40% Both</b>
Is the load time of the tool appropriate? (Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree)	<b>Answers: 80% Strongly Agree, 20% Agree</b>
What improvements would you like to see in the tool?	<b>Answers: Three grader liked to see the reason B-scan was flagged, one grader liked to have the region of "possible issue" be identified. One grader liked to see a better bracketing.</b>
What are the most valuable aspects of the feature? Who would benefit from it and why?	<b>Answers: Four graders felt that the most valuable aspects of the feature were related to time saving with the most benefit for optometrists or comprehensive ophthalmologists. One grader felt that the low-quality element was meaningful and provides the doctor with a sense of how trustworthy the analysis is.</b>

## CONCLUSIONS

- All graders agreed that the tool increases efficiency with appropriate loading time, can be useful for diagnoses and can be used by both technicians as well as clinicians.
- When evaluating agreement between graders on individual cases, differences in inter-clinician's grades were observed, even between retina specialists.
- The mean agreement between the B-scan of interest algorithm and the graders was considered clinically acceptable.

## References

- <sup>1</sup>Yu et al. IOVS 2020; 61(9): Abstract PB0085.
- <sup>2</sup> Elezaby et al. IOVS 2020; 61(9); Abstract PB0090