High throughput lesion evaluation and quality control for incorporating quantitative imaging metrics into clinical practice

Anisha Keshavan PhD[1], Kelly Leyden MRes[1], Barbara Dappert MD[1], Michael Iv MD[1,2], Michael Becich MS[1], Michael Towbin BS[1], David Hughes BSN[1], Annalise Miner BS[3], Revere Kinkel MD[3], Jennifer Graves MD [3]

[1]Octave Bioscience, Menlo Park, CA, United States of America, [2]Stanford University, CA, United States of America, [3]University of California, San Diego, CA, United States of America

actrims AMERICAS COMMITTEE FOR TREATMENT AND RESEARCH IN MULTIPLE SCLEROSIS
ECTRIMS EUROPEAN COMMITTEE FOR TREATMENT AND RESEARCH IN MULTIPLE SCLEROSIS
MS VIRTUAL 2020
8th JOINT ACTRIMS-ECTRIMS MEETING SEPTEMBER 11-13

DISCLOSURES

K.L. and A.K. are employees of **Octave Bioscience**, Inc.

M.I.'s contribution to this study was as a paid consultant, and was not part of his Stanford University duties or responsibilities.

Dr. Kinkel is a scientific consult and/or educational speaker for Alexion, Bristol Myers Squibb, Biogen, Sanofi/Genzyme and Roche/Genentech

GOAL

MRI quantitative lesion metrics in the clinic

- MS lesion counts and volumes are poised to be **salient clinical biomarkers of disease progression**
- Why don't we have widespread adoption?

Lesion Volume
Lesion Counts
Whole Brain
Atrophy
Spinal cord area
etc...

Roadblock to clinical use: Quality Control bottleneck

- Automated segmentation methods have error rate
  - Algorithmic variability
  - Image quality variability
  - Low expert agreement
- Visual quality control (QC) is required before a clinical decision can be made.
- **Visual QC is fundamentally *time consuming***
  - Especially if segmentation errors need to be edited

OBJECTIVE

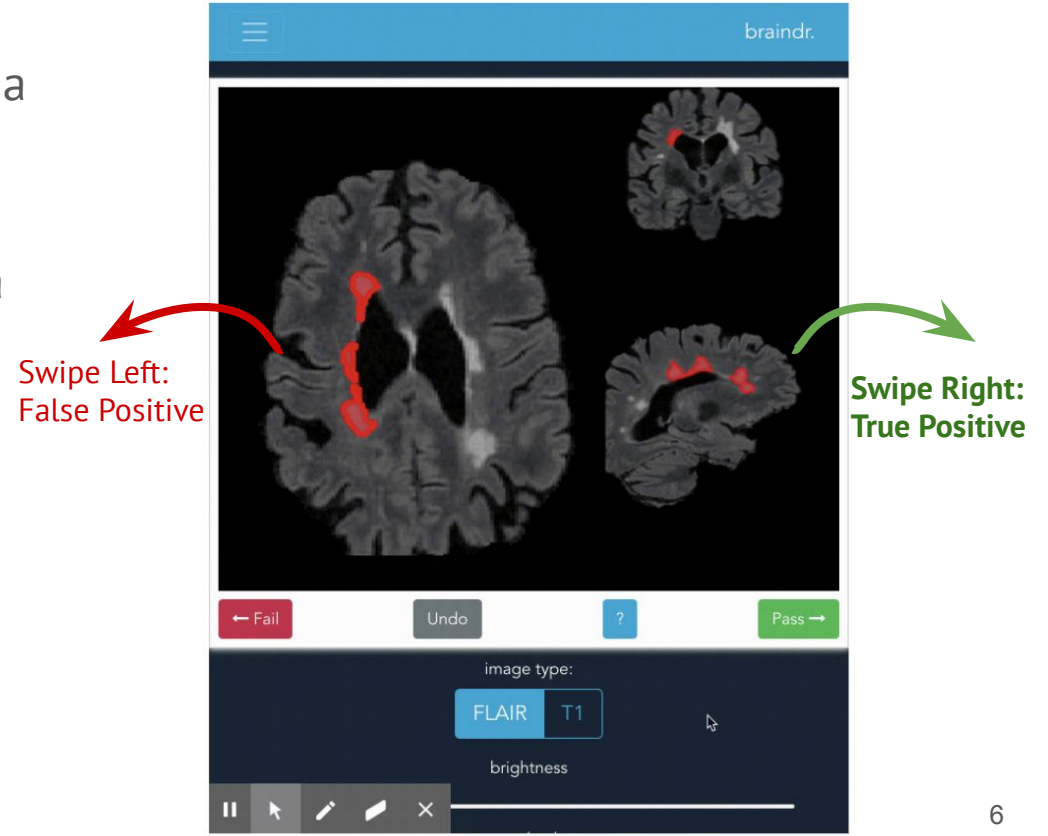
Time efficient QC and training

- Use an efficient web interface[1] to quickly:
    a. Evaluate experts and non-experts to understand and resolve disagreements
    b. QC the output of automated lesion segmentation methods.
- Evaluate feasibility
    a. number of lesions QC'd
    b. intra-rater reliability
    c. inter-rater reliability
    d. the types of lesions that are disagreed upon

[1] Keshavan A et al. *Frontiers in Neuroinformatics*, 2019

Braindr: App to quickly rate lesions

- Potential lesions (PLs) screened via binary “swipe” on app (right=pass, left=fail)
- 2D Triplanar view with lesion area blinking red → **quick decision making**
- Compatible with smartphones, tablets, and desktop
- Toggle between T2 FLAIR and T1 images
- Adjust brightness/contrast



Generate high volume of potential lesions to rate

- 3D T1 and T2 FLAIR images from 32 subjects were registered, N4 bias field corrected, and z-scored.
- Thresholded subtraction images (Z_FLAIR-Z_T1) at varying levels.
- Generated a triplanar image of each resulting segmentation (called a potential lesion, PL), resulting in over **80,000** individual PL's needing QC, which simulates a **high-throughput scenario with a high error rate**.
- Measured variability between and within raters by calculating the Intraclass correlation coefficients (ICC).

>14K images rated; reliable consensus

- Feasibility: **14,973** PLs were labelled by 5 raters.
- Inter-rater reliability for an average rating **ICC(2,k) = 0.92**, and individual ICC(2,1) = 0.74.
- Disagreements occurred more frequently on PL's in the **brainstem, cerebellum, hippocampus, and basal ganglia**.

Rater	ICC(1,1)
Neuroradiologist (MI)	0.97
Beginner (MB)	0.90
Tech 1 (AK)	0.87
Tech 2 (KL)	0.85
Neurologist (BD)	0.84

Intrarater Reliabilities

Feasible for high volume patient load in clinic

- We simultaneously evaluated raters, and QC'd lesions from an automated method using a quick, scalable, web application. This enables us to
  - 1) improve expert agreement on lesion identification
  - 2) develop better quality education materials for experts and non-experts alike
  - 3) train new raters quickly
  - 4) **ensure the quality of the measurements at scale**.

NEXT STEPS

Machine learning to triage lesion segmentation QC

- In a high-throughput scenario, we need a **triage system for segmentation QC**
- With the training dataset we've generated here, we can:
  - **Predict** which image segmentations will need more editing by raters
  - **Prioritize** the rating of potential lesions based on classification uncertainty to train models more quickly (Active Learning)
  - **Decide** which potential lesions should go to which raters (e.g. lesions in regions that are difficult to assess should go to more experienced raters)

Questions?

Contact us: akeshavan@octavebio.com, kleyden@octavebio.com, michaeliv@octavebio.com

Visit additional posters with Octave Bioscience

P0055 P0063 P0082 P0091 P0552 P0590