THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center
Making Cancer History®

# Scoring Contour Agreement using a Beta Distribution Model

T. Y. LIM [1], X. WANG [1] and J. YANG [1]
[1] University of Texas MD Anderson Cancer Center, Houston, TX, USA.

JULY 12–16 2020 VIRTUAL
JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]

## INTRODUCTION

The metrics used to represent contour agreement are usually presented as a single value without an accompanying understanding of the reliability of that value. The lack of interpretation limits the utility of a value, such as in determining whether a resident's contouring skills truly improved after a rotation.

The beta distribution is a continuous distribution that allows for a host of flexible shapes while having a finite-length interval. Its standard form's interval of 0 to 1 may make it suited for modeling typical metrics used to represent contour agreement, such as the Jaccard coefficient that takes on a value between 0 to 1.

## AIM

To investigate the beta distribution as a method to comprehensively represent contour agreement, specifically for evaluating the effectiveness of training interventions.

## METHODS

1. Five attending physicians and seven radiation oncology residents contoured twenty-six head-and-neck structures on two patients.
2. For the residents, the contouring on the first patient was performed before an educational intervention while contouring on the second patient was performed after an educational intervention.
3. To generate consensus contours for comparison, the Simultaneous Truth and Performance Level Estimation (STAPLE) was applied to a collection of all the attending physician contours and each of the resident contour of each structure.
4. For each resident, Jaccard coefficients (contour agreement scores) were calculated between the consensus contour versus the resident contour of each structure and patient.
5. These 2184 Jaccard coefficients were then fitted to a beta distribution, stratified by structure, resident, and patient.
6. Lastly, plots and heat maps of the beta distributions' probability density functions (PDF) were generated to visualize the contour agreement.

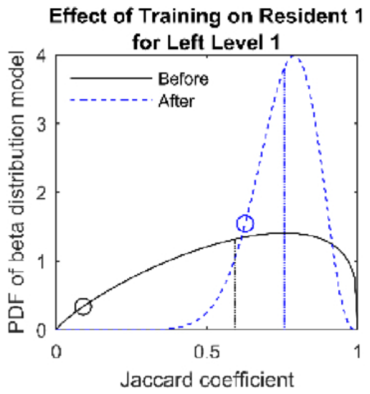
## RESULTS

Plots of the beta distributions' PDF could illustrate the confidence of the contour agreement.

Ideally, when comparing the beta distribution PDF before and after training, we expect to see a shift towards the right, a smaller range, and higher peaks (greater confidence in specified contour agreement).

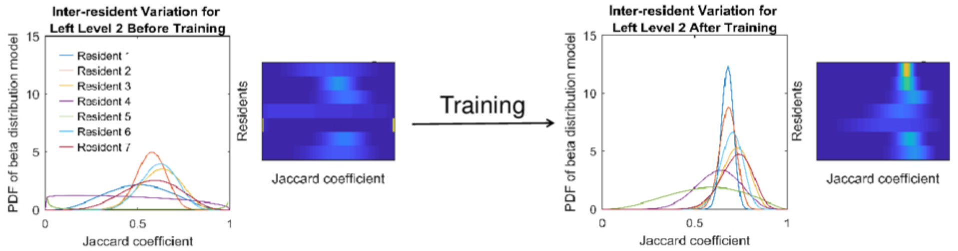
Shown on the right is an example of effective training. The training shifted the mean of the beta distribution from 0.59 to 0.76 (average agreement increased), the standard deviation from 0.24 to 0.10 (variability between the resident contour compared to attending contours reduced), and kurtosis from -0.87 to 0.1 (higher peak/confidence and less outliers), indicating the training was effective in bringing the resident contour in closer agreement with the attending contours.

The beta distribution could be used to evaluate which structures and residents were more susceptible to improvement with training compared to other structures and residents.

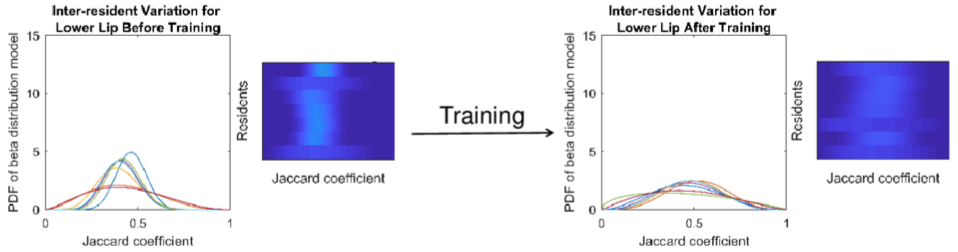



### Using a beta distribution model to evaluate structures: Which structures can be improved with training?

Structures that may be improved with training can be demonstrated by the beta distribution PDFs shifting to the right and having smaller spread (reflecting better contour agreement). For instance, the following figures show the beta distribution PDF curves (and in heatmap representations) for left level 2 lymph nodes before and after training. The training was useful in improving the residents' contouring of the left level 2 lymph nodes.

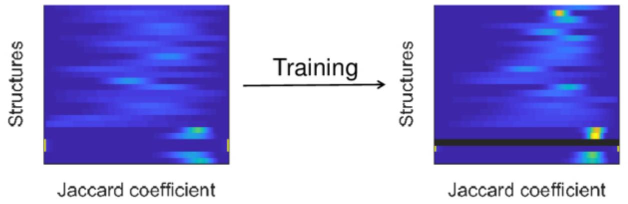



Structures that may **not** be easily improved with training can be seen in the beta distribution PDFs showing a wider spread and lower peaks, therefore the confidence is low even though the Jaccard coefficient values showed improvement. For instance, the following figures illustrate the beta distribution PDFs for lower lip before and after training. Training did not improve contour agreement for the lower lip, potentially due to the inherent subjectivity in contouring this structure, rendering the training ineffective.

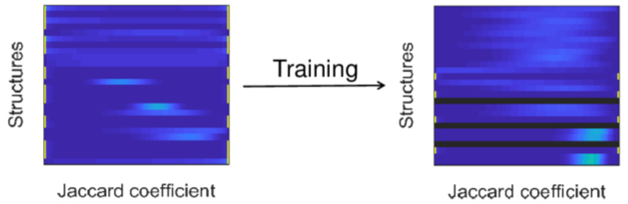



### Using a beta distribution model to evaluate individuals/algorithms: Which residents improved their contouring after training?

Residents who have improved with training can be seen from the beta distribution PDFs shifting towards the right, having reduced spread and higher peak values (bright yellow). For instance, the following figures show the beta distribution PDFs in heatmap representations for Resident 1 before and after training. Before training, some structures' peaks were reasonably defined and some other structures less so. After training, most structures had improved agreement. This suggests that with similar training, this resident may further improve their contour agreement with the attendings'.




Residents who may **not** have improved with training can be seen from the beta distribution not showing many peaks/confidence. For instance, the following figures show the beta distribution in heatmaps for Resident 5 before and after training. Before training, most contours had very poor agreement (shown by asymptotic yellow peaks at 0 and 1). After training, there were some improvements in a few structures but this can be said with low certainty (reflected by the lower color intensity). More intensive, or a different training intervention may be required for this resident.




## CONCLUSIONS

The beta distribution model comprehensively scored contour agreement in the context of each structure's inherent contour agreeability. It can represent the overall impact of training on a particular structure and illustrate the amenability of a particular structure to contouring training interventions given a contour's inherent interobserver variation.

Furthermore, the beta distribution model can represent the overall impact of interventions on a particular learner. This model could be used to evaluate the effectiveness of interventions, be it the training provided during rotations to residents or algorithm changes to automated contouring software. The scores may potentially be provided to residents periodically throughout their learning progress to enable residents to focus on improving their contouring of specific structures.

In summary, scoring contour agreement using a beta distribution model instead of using just a single parameter provides a more comprehensive assessment and puts the scores in the context of achievability given a specific structure.

## REFERENCES

1. **Devore JL**. Probability and Statistics for Engineering and the Sciences. 7th ed. Canada: Brooks/Cole, Cengage Learning, 2009. Chapter 4.5: The Beta Distribution; p. 167-168
2. **Warfield SK et al**. Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation. *IEEE Transactions on Medical Imaging 2004; 23, 7, 903-921*
3. **Awan M et al**. Prospective assessment of an atlas-based intervention combined with real-time software feedback in contouring lymph node levels and organs-at-risk in the head and neck: Quantitative assessment of conformance to expert delineation. *Practical Radiation Oncology 2013; 3, 186-193*
4. **Yang J et al**. Statistical Modeling Approach to Quantitative Analysis of Interobserver Variability in Breast Contouring. *International Journal of Radiation Oncology Biology Physics 2014; 89, 1, 214-221*

## CONTACT INFORMATION

tlim@mdanderson.org