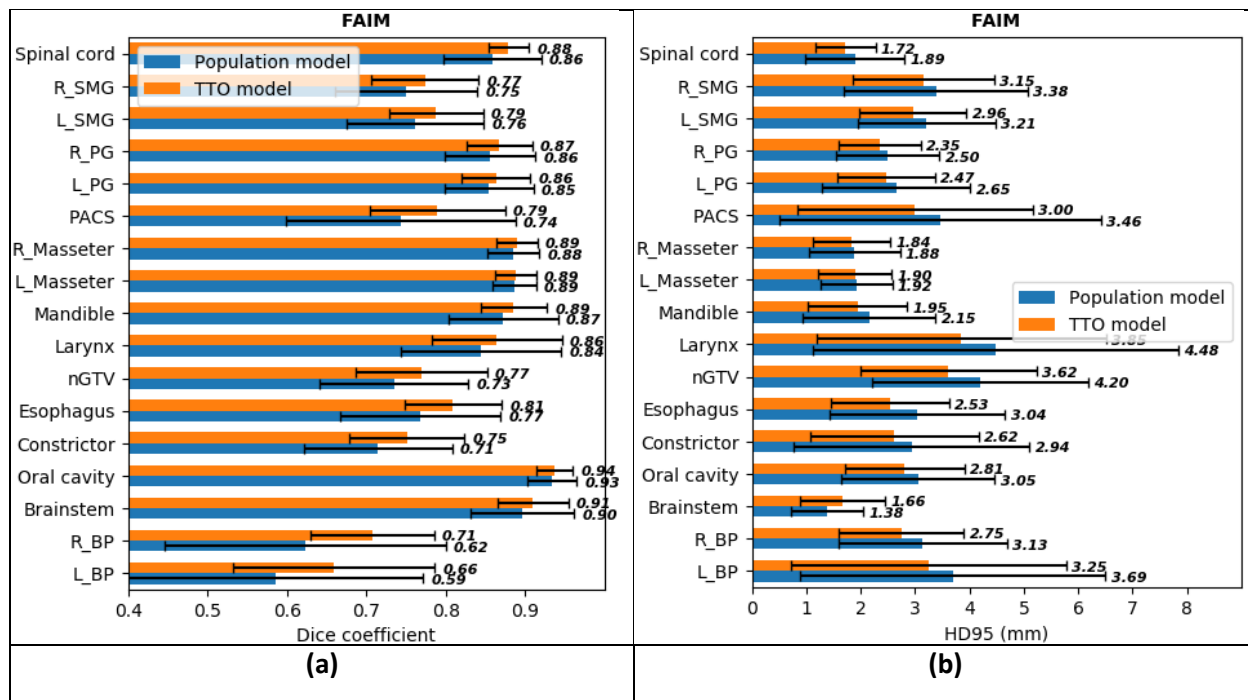
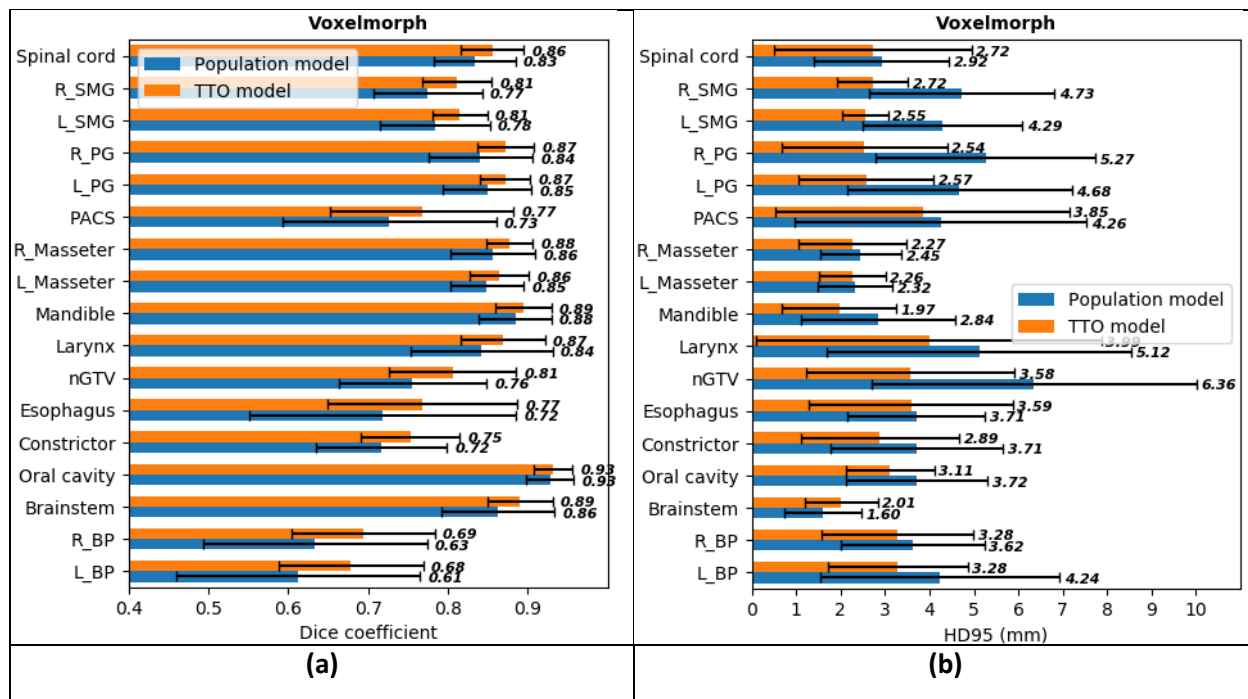


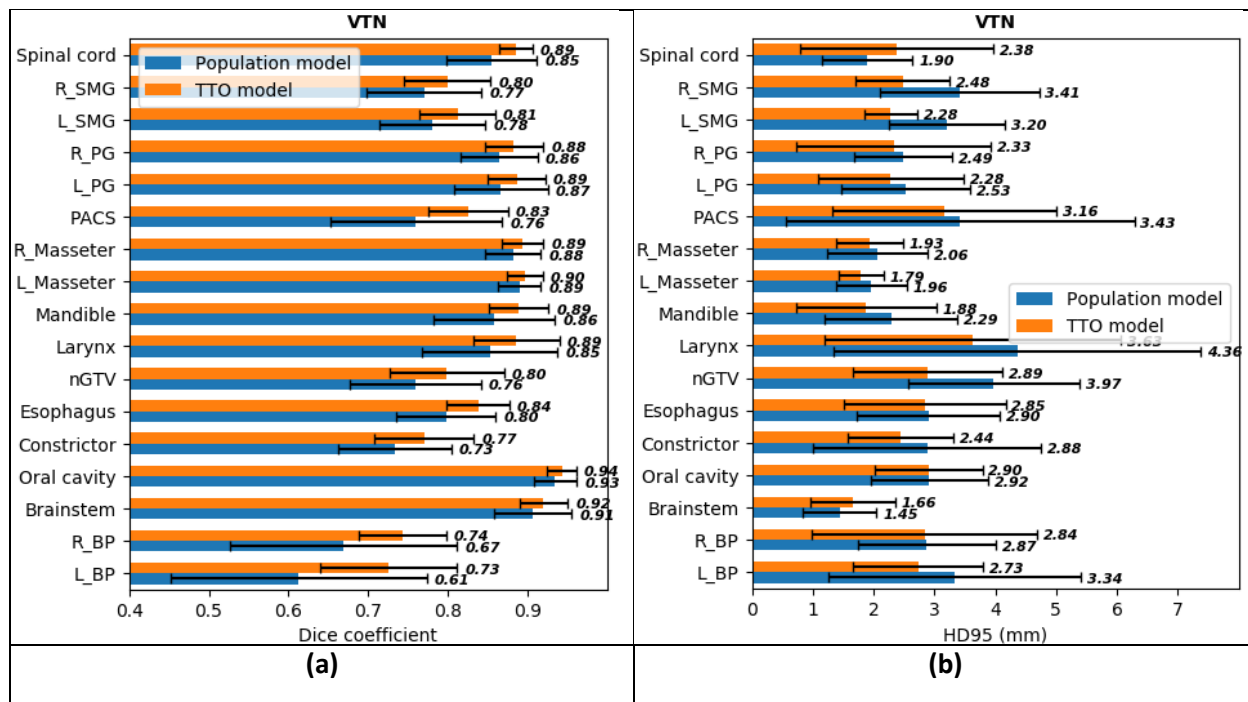
**Supplementary figure 1. Population model vs. individualized model for CNN architecture.** DSC and HD95 were calculated against the ground truth for each of the 17 selected structures from the 39 test patients. The bar plots represent mean values and the error bars represent standard deviation.



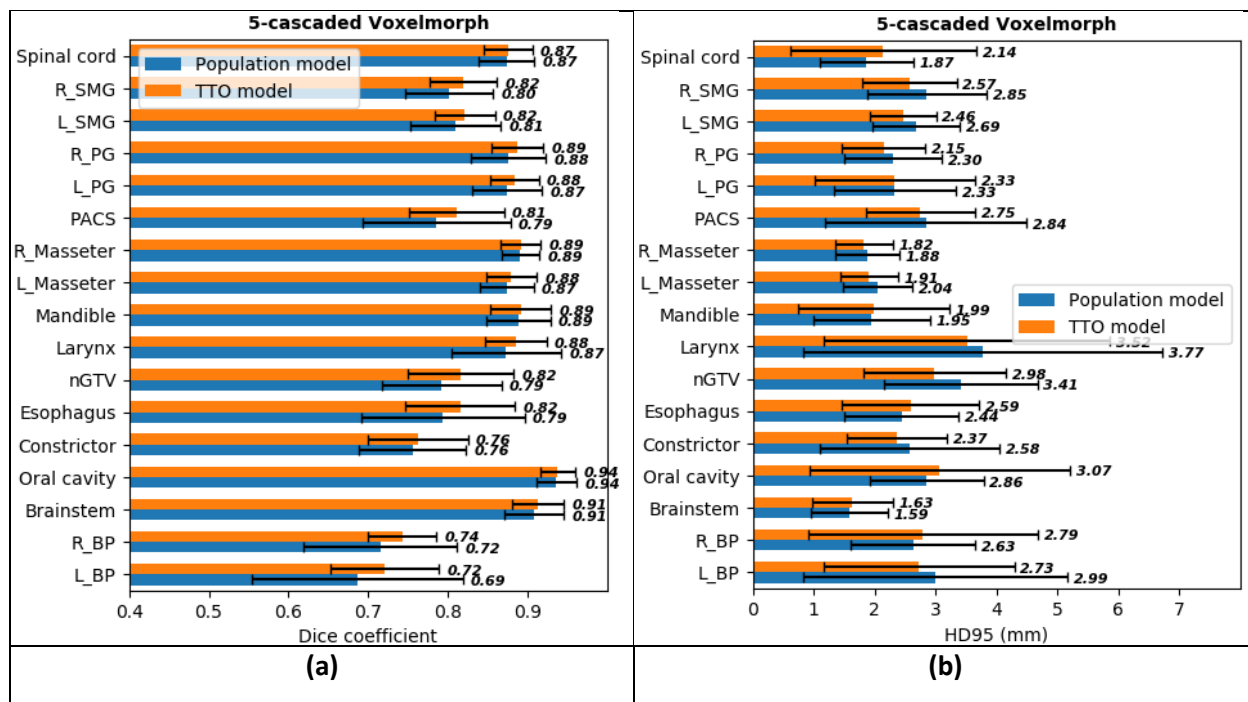
**Supplementary figure 2. Population model vs. individualized model for FAIM architecture.** DSC and HD95 were calculated against the ground truth for each of the 17 selected structures from the 39 test patients. The bar plots represent mean values and the error bars represent standard deviation.



**Supplementary figure 3. Population model vs. individualized model for Voxelmorph architecture.** DSC and HD95 were calculated against the ground truth for each of the 17 selected structures from the 39 test patients. The bar plots represent mean values and the error bars represent standard deviation.

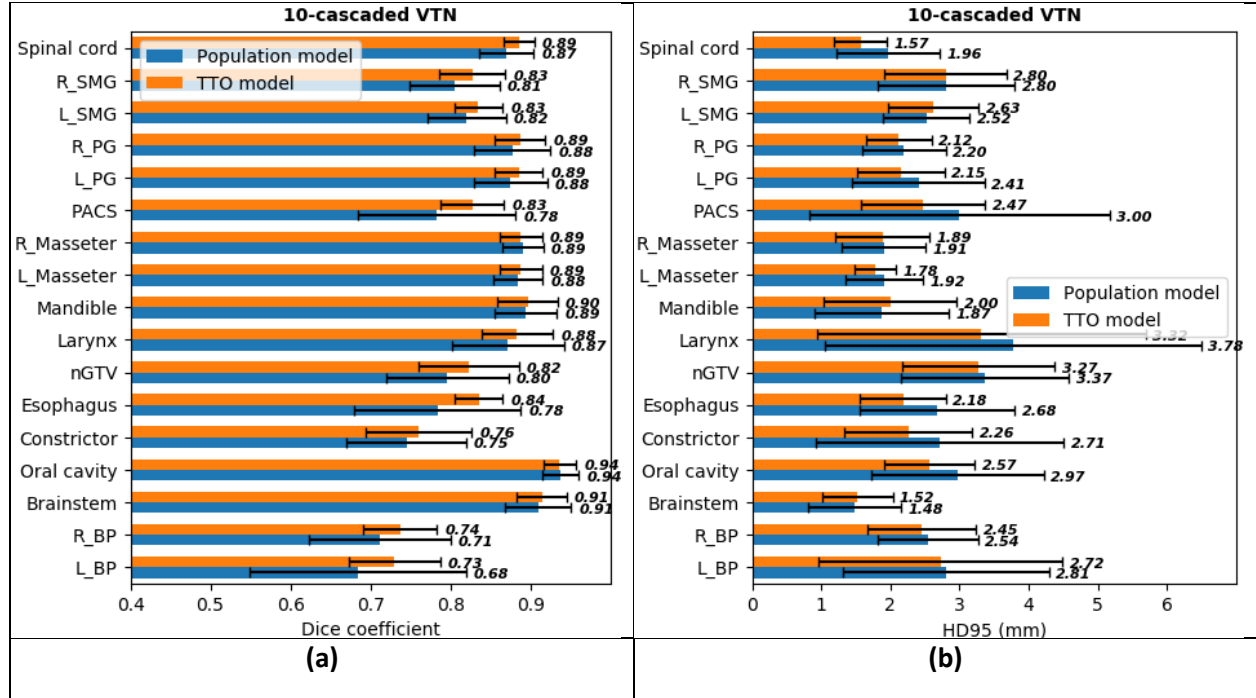


**Supplementary figure 4. Population model vs. individualized model for VTN architecture.** DSC and HD95 were calculated against the ground truth for each of the 17 selected structures from the 39 test patients. The bar plots represent mean values and the error bars represent standard deviation.

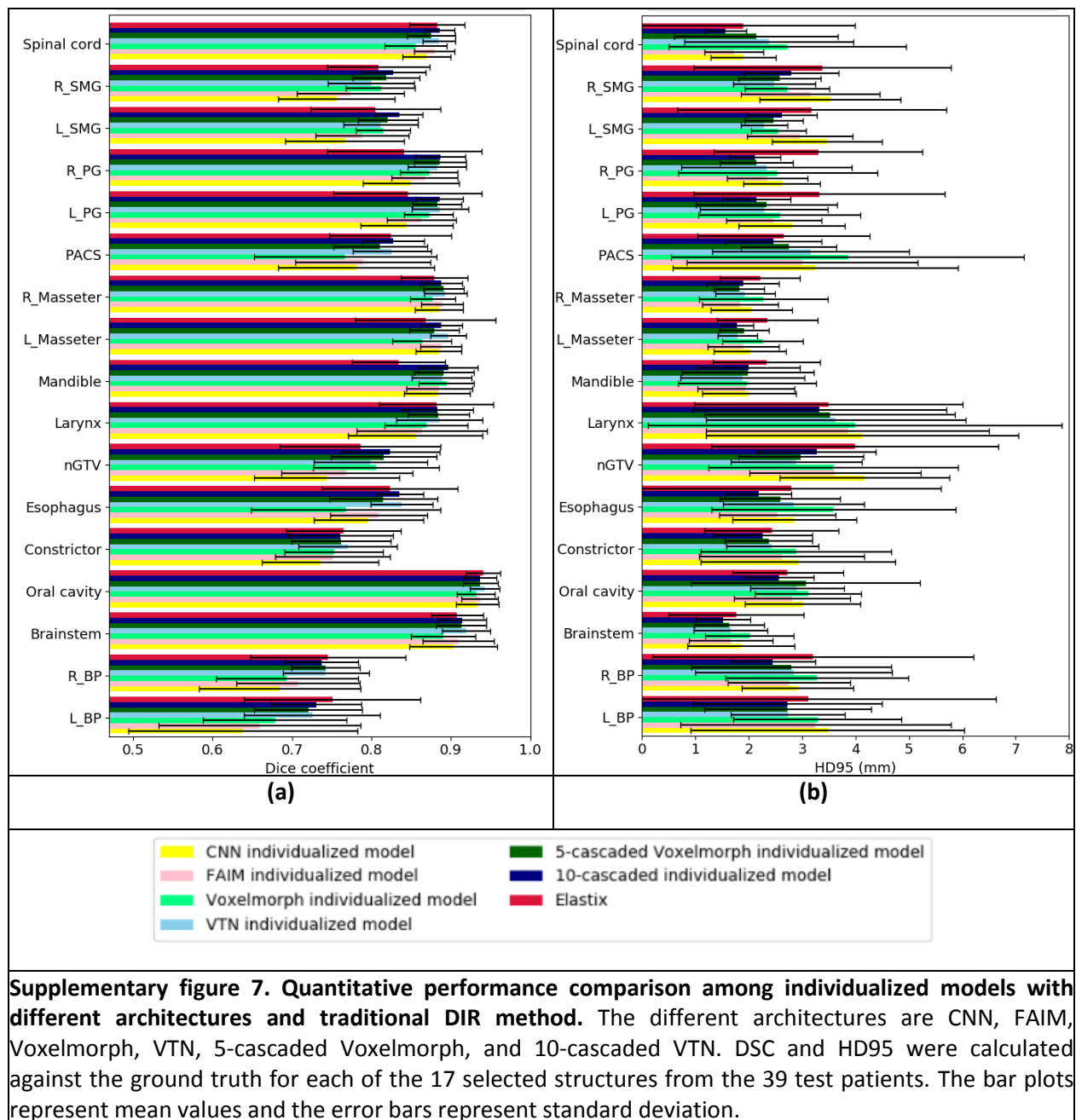


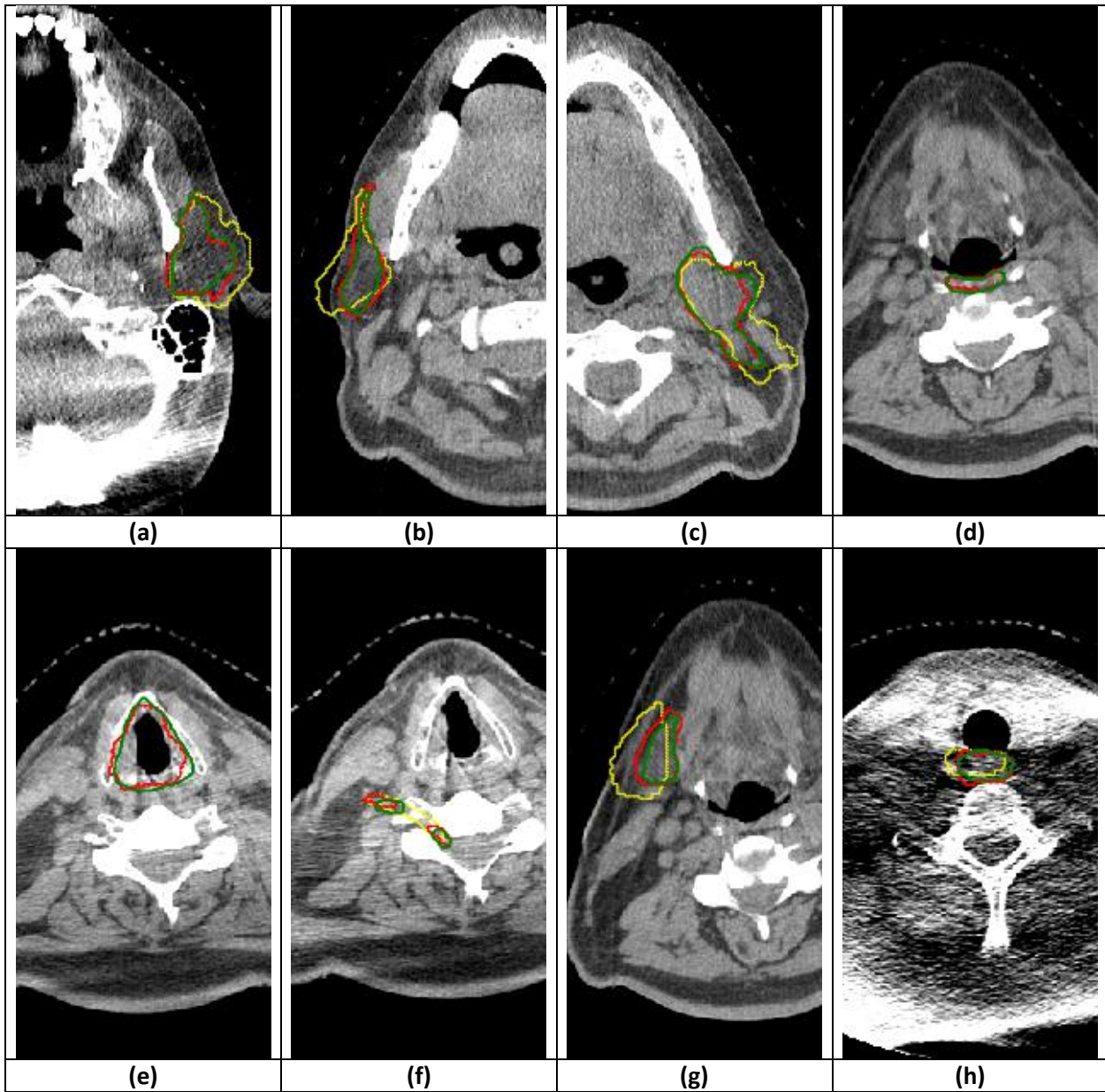
**Supplementary figure 5. Population model vs. individualized model for 5-cascaded Voxelmorph architecture.** DSC and HD95 were calculated against the ground truth for each of the 17 selected

structures from the 39 test patients. The bar plots represent mean values and the error bars represent standard deviation.



**Supplementary figure 6. Population model vs. individualized model for 10-cascaded VTN architecture.** DSC and HD95 were calculated against the ground truth for each of the 17 selected structures from the 39 test patients. The bar plots represent mean values and the error bars represent standard deviation.





**Supplementary figure 8. Illustration of generalization problem a population model has, while a TTO model can avoid this problem.** The segmentations from (a) to (h) are: (a) L\_PG, (b) R\_PG, (c) nGTV, (d) Constrictor, (e) Larynx, (f) R\_BP, (g) R\_SMR, and (h) Esophagus. Green, red, and yellow contours represent ground truth, TTO model segmentation and population model segmentation, respectively. The architecture used in this experiment is Voxelmorph.