SIIM CMIMI24
Oct 21-22 | Boston, MA
Conference On Machine Intelligence In Medical Imaging

Deep Learning for Incidental Parotid Tumors on CT: Optimal Methods for Screening and Segmentation

Wei Shao, PhD Student, UC Irvine; Shirin Salehi; Chanon Chantaduly; Hayden Troutt, MPH; Peter Chang, MD

Introduction/Background

Parotid gland tumors (PGT) are the most common salivary gland tumors. With increasing imaging utilization, most PGTs are detected incidentally on CT, however many are overlooked by radiologists prioritizing acute pathology. This study presents a deep learning (DL) solution for opportunistic PGT detection on CT with a focus on optimizing complimentary objectives for tumor screening and segmentation.

Methods/Intervention

A retrospective cohort of 11,449 consecutive non-contrast head CT exams were aggregated from two academic centers. PGTs, defined as a parotid mass >10 mm, were identified from radiology or histopathology reports and annotated with a mask by an expert neuroradiologist. In total, 219 PGTs were identified (N=112 hospital A, N=107 hospital B). A multistage DL pipeline was developed for PGT detection (Fig. 1). First, an initial model localizes each parotid gland. Subsequently, a single 3D U-Net simultaneously implements the segmentation (per-voxel spatial overlap) and screening (per-exam tumor detection) tasks. To convert segmentation outputs into binary screening results, thresholds for positive voxel predictions were calibrated for optimal accuracy. Given complimentary objectives of the segmentation and screening tasks, various loss functions (binary cross-entropy, focal loss, soft Dice) and training cohorts (full cohort, positive only) were evaluated. Performance was assessed using five-fold cross-validation.

Results/Outcome

Table 1 summarizes results for the various experiments. Overall, the best screening model achieved a per-exam specificity, sensitivity, PPV, NPV and accuracy of 0.947, 0.719, 0.858, 0.878, 0.872, while the best segmentation model achieved a Dice score of 0.71. Of the positive predictions, six tumors were missed by the original interpreting physician. In general, cross-entropy (CE) outperformed focal loss (FL) for segmentation, while FL outperformed CE for screening due to improved specificity and lower false positives. Soft Dice (SD) tended to improve both tasks. The use of negative training examples significantly decreased tumor Dice score while reducing false positives for the screening task.

Conclusion

By combining a first-pass screening model with a subsequent focused segmentation model, a unified DL framework can identify and delineate PGTs on routine CT with high accuracy.

Statement of Impact

A DL model can identify incidental PGTs on routine CT imaging with high accuracy including tumors missed in a realistic clinical workflow.

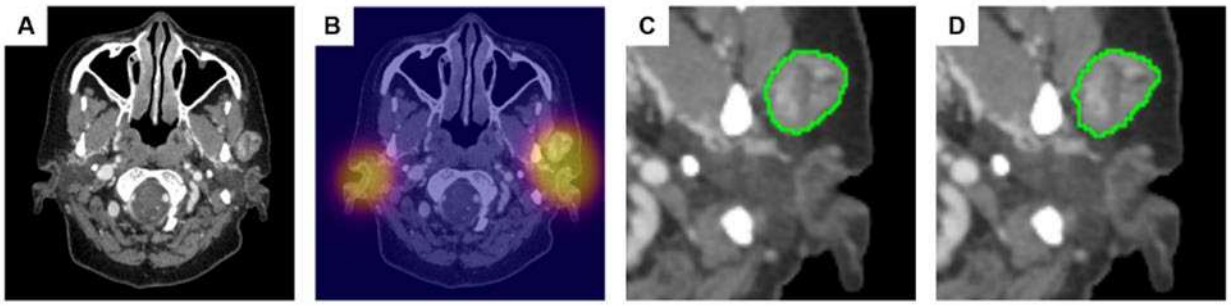

Fig. 1. Overview of two-step deep learning algorithm for parotid mass detection and segmentation. (A) Original full resolution CT exam is used by initial deep learning localization algorithm to generate prediction heatmaps (B) isolating the right and left parotid glands. The initial localization algorithm outputs are used to generate cropped volumes of each individual parotid gland, after which a second segmentation algorithm is used to identify parotid masses. (C) Final algorithm output, and (D) corresponding ground-truth annotation show high consensus.

Table 1

Dataset	Loss function	Spec, Sen, PPV, NPV, Acc	Dice score
Full cohort	CE	0.911, 0.760, 0.802, 0.890, 0.863	0.618
	CE, SD	0.937, 0.717, 0.841, 0.875, 0.866	0.663
	FL	0.934, 0.738, 0.832, 0.884, 0.870	0.592
	FL, SD	0.947, 0.719, 0.858, 0.878, 0.872	0.576
Positive only	CE,	N/A	0.652
	CE, SD	N/A	0.685
	FL	N/A	0.71
	CL, SD	N/A	0.706

Table 1. Screening and segmentation performance of experimented training methods. CE: cross-entropy; SD: soft dice; FL: focal loss; spec: specificity; sens: sensitivity; PPV: positive predictive value; NPV: negative predictive value; acc: accuracy

Keywords
Deep Learning; Parotid Tumor; Screening; Segmentation; Optimal methods