Deep Learning Method for Rapid Simultaneous Multi-structure Temporal Bone Segmentation

Caio Neves, MD, PhD[1,2]
Iruena Kessler, MD, PhD[2]
Nikolas Blevins, MD[1]

[1]Otolaryngology – Head and Neck Surgery, Stanford University School of Medicine, CA, USA

[2]Medical Sciences Graduate Program, University of Brasília, BRA

ABSTRACT

Introduction
This study presents a deep learning (DL) model for the rapid automated segmentation of nine key structures in the temporal bone from CT scans.

Methods
A total of 325 clinical temporal bone CT scans from a tertiary care hospital were used in this study. In each CT scan, nine key structures (Internal Carotid Artery, Chorda Tympani, Inner Ear, External Auditory Canal, Facial Nerve, Internal Auditory Canal, Ossicles, Otic Capsule, and Sigmoid Sinus) were manually segmented by an expert and used as multilabel ground truth. A state-of-the-art DL algorithm (SwinUNETR) was adapted and trained to build a prediction model using 80% of the sample. The remaining 20% of the sample was used as a test set to evaluate the model using metrics such as Dice, Balanced Accuracy (BAcc), Average Hausdorff Distance (AHD), and 95th Hausdorff Distance (HD95).

Results
On the test set, the metrics for all structures were as follows: Dice 0.87, BAcc 0.94, AHD 0.22 mm, and HD95 0.79 mm. The mean processing time was 9.1 seconds per study.

Conclusions A large training dataset and the use of optimized algorithms resulted in a robust model for the fast simultaneous segmentation of key structures in the temporal bone from clinical CT scans.

INTRODUCTION

Otologic surgery is a highly specialized discipline, requiring the surgeon to have in-depth knowledge of radiological and surgical anatomy, as well as intraoperative three-dimensional perception.

The small and delicate surgical field, which includes complex procedures such as cochlear implantation, tympanomastoidectomy and superior semicircular canal dehiscence repair, presents significant challenges due to the delicate interrelationships between the bones and the vital neurovascular structures in the temporal bone.

A method for the rapid and accurate generation of patient-specific high-fidelity 3D models for preoperative planning and intraoperative navigation could offer a variety of potential benefits for both the patient and the surgeon.

This study aims to describe the development and validation of a system for the segmentation of key structures in the temporal bone from clinical CT scans using deep learning algorithms.

MATERIALS AND METHODS

Nine key structures from 325 temporal bone CT scans were manually segmented by a specialist. The inclusion criterion was non-contrast temporal bone CT with a maximum spacing of 0.33mm. CT images that showed significant metallic artifacts or movements that prevented clear visualization of the structures were excluded.
In all CT scans, the Internal Carotid Artery, Chorda Tympani, Inner Ear, External Auditory Canal, Facial Nerve, Internal Auditory Canal, Ossicles, Otic Capsule and Sigmoid Sinus were segmented by a specialist using 3D Slicer and considered ground truth for the development of the DL model.
The dataset was randomly split into training set (n=260) and test set (n = 65).
A state-of-the-art algorithm (SwinUNETR) was adapted to train a DL segmentation model on the training set (n=260), using five-fold cross validation. The supervised training step was carried out for 1000 epochs, with network input of 96x96x96 pixels and initial learning rate of $2x10^{-4}$.
The final DL model was used to generate the segmentation of the test set (n= 65) and the DL derived segmentation was compared with the manually segmented structures, considered the ground truth.
The objective analysis of the accuracy of the model included the Dice coefficient, the balanced accuracy (BAcc) and the average (AHD) and the 95th percentile of the Hausdorff distance (HD95). The processing time for the DL segmentation was measured.

RESULTS

325 CT were selected and manually segmentation of the nine key structures was performed by a specialist. Clinical evaluation of the dataset showed that 69% of the dataset had normal anatomy, while 31% showed alterations caused by inflammation or previous surgery. The trained DL model showed an average Dice of 0.87 for all structures, BAcc of 0.94, AHD of 0.22 mm and HD95 of 0.79mm. The mean processing time was 9.1s per scan.

Structure		Dice	BAcc	AHD	HD95
Int. carotid artery	Mean	0.90	0.96	0.28	1.03
	SD	0.06	0.05	0.23	0.95
Chorda Tympani	Mean	0.59	0.83	0.41	1.06
	SD	0.18	0.10	0.89	1.75
Inner ear	Mean	0.95	0.98	0.06	0.32
	SD	0.02	0.02	0.05	0.41
Ext. auditory canal	Mean	0.90	0.94	0.29	1.05
	SD	0.10	0.06	0.43	1.05
Facial Nerve	Mean	0.83	0.95	0.14	0.46
	SD	0.06	0.03	0.05	0.22
Int. auditory canal	Mean	0.93	0.97	0.13	0.52
	SD	0.03	0.03	0.05	0.26
Ossicles	Mean	0.89	0.96	0.17	0.50
	SD	0.15	0.04	0.46	0.82
Otic capsule	Mean	0.95	0.98	0.07	0.26
	SD	0.03	0.03	0.03	0.03
Sigmoid sinus	Mean	0.86	0.92	0.47	1.91
	SD	0.07	0.06	0.34	1.47
Average	Mean	0.87	0.94	0.22	0.79
	SD	0.08	0.05	0.28	0.77

Table 1. Objective analysis of the DL model accuracy over the test dataset (n = 65)

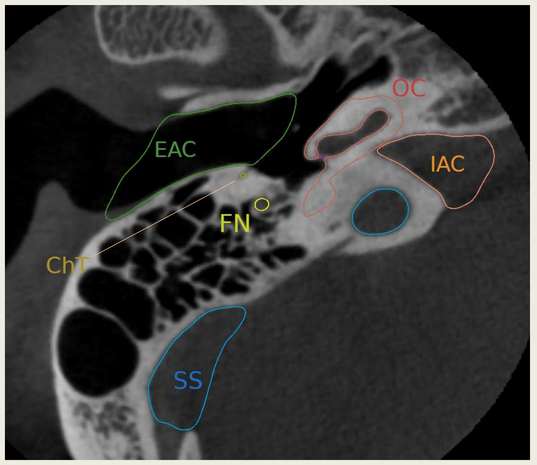

Figure 1. Axial scan with key structures segmented by the DL model. ChT: Chorda timpani; EAC: External auditory canal; IAC: Internal audit. canal; FN: Facial nerve; OC: Otic capsule; SS: Sigmoid sinus

DISCUSSION

This study describes the construction of a model for the automated segmentation of key structures in the temporal bone by analyzing CT scans using deep learning algorithms. The process aims to facilitate preoperative preparation, as well as leverage surgical simulation and mixed reality projects.
In this study we adapted a hybrid algorithm (SwinUNETR) which combines transformers with CNN, which can combine the benefits of the different architectures to build the model. These mechanisms can favor the training of a better-performing multi-structure model, as both the features of the structures are learned at various levels and the interrelationship between the structures are incorporated into the model.
When compared with the results of the current literature on the subject, we can see an advantage of the model developed in this work in terms of the number of structures analyzed simultaneously and, above all, the size of the dataset. With a larger dataset, we can expect a greater variety of radiological findings when training the algorithm and assume greater robustness and generalization of the model.

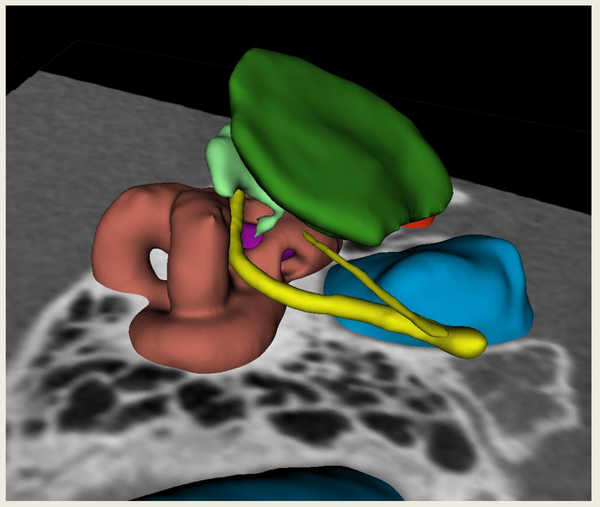

Figure 2. 3D rendering of the key structures as segmented by the DL model

CONCLUSIONS

This work involved the development and validation of a system for anatomical segmentation of temporal bone structures in CT scans using deep learning techniques. The automated segmentation model produced showed a high degree of accuracy in the test set and robustness for generalization to new data.

The results obtained highlight the potential of using deep learning techniques as a relevant tool for anatomical segmentation of temporal bone CT scans, for personalized preoperative planning and the development of improved intraoperative navigation systems.

REFERENCES

1. Fauser J, Stenin I, Bauer M, et al. Toward an automatic preoperative pipeline for image-guided temporal bone surgery. *Int J Comput Assist Radiol Surg*. 2019;14(6):967-976.
2. Ding AS, Lu A, Li Z, et al. A Self-Configuring Deep Learning Network for Segmentation of Temporal Bone Anatomy in Cone-Beam CT Imaging [published online ahead of print, 2023 Mar 8]. *Otolaryngol Head Neck Surg*. 2023
3. Ke J, Lv Y, Ma F, et al. Deep learning-based approach for the automatic segmentation of adult and pediatric temporal bone computed tomography images. *Quant Imaging Med Surg*. 2023;13(3):1577-1591.
4. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. *Lect Notes Comput Sci*. 2022;12962