2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Automatic Segmentation of Intracochlear Anatomy in MR Images Using a Weighted Active Shape Model

Yubo Fan, *Student Member, IEEE,* Rueben A. Banalagay, Nathan D. Cass, Jack H. Noble, *Member, IEEE,* Kareem O. Tawfik, Robert F. Labadie, and Benoit M. Dawant, *Fellow, IEEE*

***Abstract*—There is evidence that cochlear MR signal intensity may be useful in prognosticating the risk of hearing loss after middle cranial fossa (MCF) resection of acoustic neuroma (AN), but the manual segmentation of this structure is difficult and prone to error. This hampers both large-scale retrospective studies and routine clinical use of this information. To address this issue, we present a fully automatic method that permits the segmentation of the intra-cochlear anatomy in MR images, which uses a weighted active shape model we have developed and validated to segment the intra-cochlear anatomy in CT images. We take advantage of a dataset for which both CT and MR images are available to validate our method on 132 ears in 66 high-resolution T2-weighted MR images. Using the CT segmentation as ground truth, we achieve a mean Dice (DSC) value of 0.81 and 0.79 for the scala tympani (ST) and the scala vestibuli (SV), which are the two main intracochlear structures.**

***Clinical Relevance*— The proposed method is accurate and fully automated for MR image segmentation. It can be used to support large retrospective studies that explore relations between MR signal in preoperative images and outcomes. It can also facilitate the routine and clinical use of this information.**

I. Introduction

The cochlea is an essential part of the human inner ear that is responsible for hearing. It is a spiral-shaped bony structure that contains three cavities: the scala vestibuli (ST), the scala tympani (SV), and the scala media (SM). The ST and the SV are filled with perilymph. They are separated by the osseous spiral lamina and meet at the helicotrema, which is the cochlear apex. The SM is located between the ST and the SV and separated by the basilar membrane and Reissner's membrane, respectively. It only occupies a small portion of the cochlea and is filled with endolymph.

Because the cochlea is a fluid-filled structure surrounded by bone, MR and CT images provide complementary information. In CT images, the surrounding bone is visible while in the MR images it is the intracochlear fluid that produces the signal [1], [2]. We have developed and evaluated automated methods for the segmentation of all inner structures in CT images and we have applied them to the segmentation of images acquired before and after cochlear implant procedures [3], [4]. Here, we evaluate these methods for the segmentation of T2-weighted MR images.

Segmentation of MR images is important because recently it has been shown that cochlear MR signal intensity has a clear relationship with hearing loss in untreated acoustic neuroma (AN) patients [5], [6] and predicts hearing outcomes after microsurgical resection [7] and stereotactic radiosurgery [8]. The biomolecular processes underlying degraded cochlear T2 signal remain unclear, but it is thought that they reflect increased protein concentration in the cochlear fluids [9], which has been observed in perilymph samples from cochleae of ears affected by AN [10]. Whether cochlear MR signal intensity at the time of AN diagnosis can predict long-term hearing outcomes in patients with untreated AN remains unknown, and it is unclear whether cochlear MR signal intensity precedes, coincides with, or follows observed deterioration in hearing over time.

Automated methods would permit large-scale retrospective studies and routine computation of these quantities, thus potentially facilitating prognostication of hearing outcomes in AN. Cochlear MR signal is also used to detect cochlear obliteration after AN surgery to evaluate patients for subsequent cochlear implantation [11], [12], [13]. Moreover, MR images can serve as a radiation-free alternative to preoperative CT images for planning cochlear implantation surgeries.

We note that previous work addressing the segmentation of the inner ear in MR images does not separate the intracochlear anatomy (ICA) from the entire labyrinth. Recent work on the topic includes Zhu et al. [14] who segment the labyrinth in MR images using level sets and a statistical shape model as prior and Vaidyanathan et al. [15] who develop a 3D U-Net-based method to segment the labyrinth. Segmentation of the ICA is however necessary to conduct studies that relate cochlear signal to outcomes and to the best of our knowledge this has not been reported.

II. Methods

The method we propose for the segmentation of the ICA in MR images is adapted from our previously developed weighted active shape model (wASM) method developed to segment the same structures in CT images [4]. We modify this method as discussed in the next following subsections to make it applicable to the T2-weighted images included in this study.

A. Shape Model Creation

Briefly, more details can be found in [4], we use a series of microCT image volumes (a typical isotropic voxel dimension of 0.036 mm) in which the intracochlear anatomy is visible to build the model. The ST and SV are manually delineated in each of the microCT image volumes to create a surface for each structure while maintaining point-to-point

Y. Fan, R. A. Banalagay, J. H. Noble, B. M. Dawant, are with the Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37235 USA. (Corresponding author: Y. Fan; email: yubo.fan@vanderbilt.edu)

N. D. Cass, K. O. Tawfik, and R. F. Labadie are with the Department of Otolaryngology – Head & Neck Surgery, Vanderbilt University Medical Center, Nashville, TN 37232 USA.

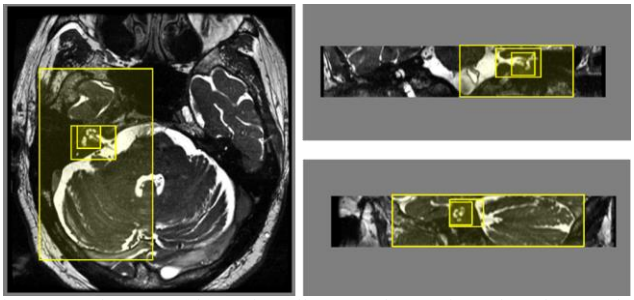

Figure 1. The ROIs (shown in yellow) used to register the atlas to other volumes. Axial view (left), coronal view (top right), and sagittal view (bottom right).

correspondence between volumes. The covariance matrix of the vertices is created and its eigenvectors are computed as proposed by Cootes et al. [16] to produce the eigenmodes of deformation.

B. Segmentation Using the Weighted Active Shape Model

After the shape model is built, the segmentation can be performed by 1) placing the initial shape in the target image, i.e., the image to be segmented; 2) iteratively fitting the wASM to the target image; 3) after the fitting converges, use the final shape as the segmentation result. The whole process is fully automatic, and we detail it in the following subsections.

1) Initialization: The aim of initialization is to localize the cochlea and place initial model points in the target image, which is done by registering an MR atlas image to the target image. Without loss of generality, we assume that the cochlea to be segmented is in the left ear. If it is in the right, we begin the process by mirroring the target image. The MR atlas image is acquired with a FIESTA sequence on a 3 T scanner, with voxel size 0.3125 mm × 0.3125 mm × 0.4 mm. To obtain the intracochlear anatomy model points in this MR atlas, we first perform the wASM segmentation on its corresponding CT image using the method in [4], then align the CT and MR images with a rigid-body registration, and finally project the model points from the CT volume to the MR atlas image.

The registration process between the atlas image and the target image consists of an affine [17] followed by a nonrigid registration [18]. Because all the high-resolution T2-weighted images (including the atlas image, see Fig. 1) were obtained with an acquisition protocol that covers only a small part of the head in the superior-inferior direction (usually less than 30mm), we follow a four-step process to improve convergence and registration accuracy in the cochlear region. We first register the whole images and then three regions of interest (ROIs) that are empirically chosen around the cochlea and have enough content to permit registration. We call ROI#1, ROI#2, and ROI#3 the three large- to small-sized ROIs shown in Figure 1. ROI#1 has a size of 65 mm × 107 mm × 28 mm and is chosen to cover the left-half of the brain. ROI#2 contains the whole labyrinth and the inner auditory canal of the left ear. The strong T2-weighted signals of the perilymph, endolymph, and cerebrospinal fluid make it easy to distinguish from the surrounding non-fluid anatomy. ROI#3 is smaller than ROI#2 but still covers the cochlea. It is selected to produce a very accurate registration of the cochlea. After the above affine transformations are computed, a

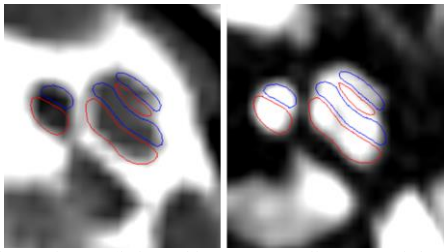

Figure 2. The CT image (left) and its corresponding T2-weighted MR image (right). The red contour shows the ST segmentation, and the blue contour shows the SV segmentation.

nonrigid registration is performed between the cochlear ROIs (ROI#3) of the atlas image and the target image. The position of the initial model points on the target image can then be obtained by projecting the points from the atlas image using a concatenation of the affine and nonrigid transformations. Finally, the shape model is fitted to the initial point-set in a weighted-least-square sense (see the following section for the detail of the fitting process) to initialize the iterative search.

2) Iterative search: Following the wASM approach put forth in [4], the model starts from a set of initial model points, and the optimal solution is computed in the target image iteratively until the shape converges. Two sets of model points were pre-defined in the wASM approach proposed to segment CT images: "edge" points and "nonedge" points. The edge points are located on the cochlear external walls and have strong image gradients. The nonedge points are the remaining points without salient image features. They were treated differently in the candidate point adjustment step and given different weights (1 for edge points and 0.01 for nonedge points) in the wASM fitting process. For the MR images used in this study (i.e., high-resolution T2-weighted MR images), although the image contrast is provided by the fluid signal, the points located close to the cochlear external walls have strong image gradients. An example of registered CT and MR images showing the cochlea is shown in Fig. 2. Also, note that even though the separation between the ST and the SV within the cochlea can be discernable in high-resolution T2-weighted MR images [19], we observe that the image gradients at these locations are very weak compared to the cochlear external walls and sometimes even nonexistent. As a result, we followed the approach described in [3] and used the same model point subsets and weights to fit the ASM to the images.

Specifically, at each iteration, every model point $\vec{y}_i$ from the last wASM fitting is adjusted to its new candidate position. If $\vec{y}_i$ is an edge point, a search is performed along the surface normal of that point. The adjusted candidate point $\vec{y}_i{}'$ is chosen to be the point with the largest gradient magnitude along the surface normal over the range of -1 mm to 1 mm from $\vec{y}_i$. If $\vec{y}_i$ is a nonedge point, then its initial position, which is the position of this corresponding point projected from the atlas image using the initial registration transformation, is used as the new candidate point $\vec{y}_i{}'$. The next step within this iteration is to fit the shape model to the candidate points in the weighted-least-squares sense. A 7 degree-of-freedom weighted point registration between the candidate shape and the mean shape $\bar{v}$ is performed to get the transformation T. The residuals are computed as

$$\vec{d} = T(\vec{y}') - \bar{v}. \quad (1)$$

The weighted-least-square fit is solved as

$$\vec{b} = (U^T W^T W U)^{-1} U^T W^T W \vec{d}, \quad (2)$$

where U is the matrix of eigenvectors, W is the diagonal matrix point weights. The coefficient $\vec{b}$ is constrained such that the Mahalanobis distance between the fitted shape and the mean shape is not greater than 3, i.e.,

$$\sqrt{\sum_{j=1}^{N-1} \frac{b_j^2}{\lambda_j}} \leq 3. \quad (3)$$

The estimated shape after this wASM fitting is then given by

$$\vec{y} = T^{-1}(\bar{v} + U\vec{b}). \quad (4)$$

The process of candidate point searching and wASM fitting is iterated until convergence, and the final shape is the segmentation result.

C. Validation

Because our previously developed wASM method to segment the ICA in CT images has been applied to various clinical applications and shown to be robust and accurate [3], [4], [20], [21], [22] and because contouring the cochlea in 132 ears would be impractical, we utilize the wASM method to create the ground truth. Specifically, we first segment the paired CT and MR images individually (note that the wASM methods in both the CT and MR share the same shape model), then rigidly register the CT image to the MR image using the same mutual information-based registration technique as in Section II.B. Finally, we project the CT segmentation result onto the MR image to provide the ground truth. Dice similarity coefficient (DSC) [23] and the average surface distance (ASD) are calculated between the wASM segmentation of the MR images and the ground truth. For DSC, which measures the volumetric overlap, we denote the binary mask of each segmented ICA structures in CT and MR images B_{CT} and B_{MR}. $|X|$ represents the number of voxels in the binary mask X. DSC is computed as

$$DSC(B_{\mathrm{MR}}, B_{\mathrm{CT}}) = 2\frac{|B_{\mathrm{MR}} \cap B_{\mathrm{CT}}|}{|B_{\mathrm{MR}}| + |B_{\mathrm{CT}}|}. \quad (5)$$

For ASD, which measures the average symmetric distance between the surface meshes, we define M_{CT} as the segmented surface mesh in the CT image and M_{MR} as the segmented mesh in the MR image. The ASD is then computed as

$$ASD(M_{\mathrm{MR}}, M_{\mathrm{CT}}) = \frac{D(M_{\mathrm{MR}}, M_{\mathrm{CT}}) + D(M_{\mathrm{CT}}, M_{\mathrm{MR}})}{2}, \quad (6)$$

where $D(M_{\mathrm{MR}}, M_{\mathrm{CT}})$ is the average distance from every point on M_{MR} to the surface of M_{CT} and vice versa.

The evaluation using DSC and ASD requires an accurate registration between the paired CT and MR images, but we visually observed that small registration errors can remain after automated registration. To factor this error out, we obtain the rigid transformation between the CT and MR images by performing a point-based registration between the point-sets of the CT wASM result and the MR wASM result. We subsequently calculate the DSC and ASD between the transformed CT segmentation and the MR segmentation. We consider the evaluation results obtained in this way to be the lower bound because this registration process minimizes the point-to-point distance between the two segmentations (point-sets) before calculating the evaluation metrics.

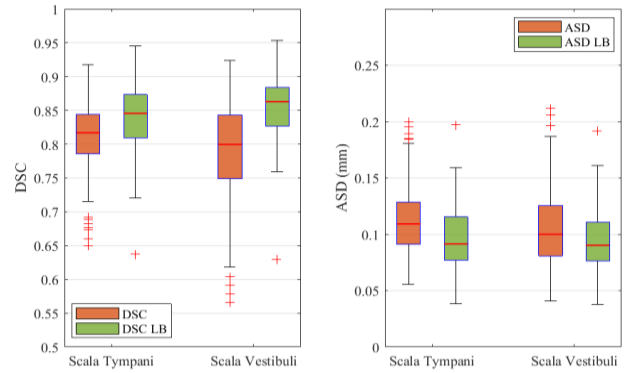

Figure 3. DSC and ASD results.

III. Experiments and Results

A. Imaging Data

We retrospectively collected preoperative images of 66 cochlear implant recipients treated at the Vanderbilt University Medical Center. Each patient had undergone preoperative CT and MR imaging of the temporal bone. The CT images were acquired with a Revolution EVO (GE Healthcare) scanner. For these images, a typical voxel dimension is 0.47 mm × 0.47 mm × 0.1 mm. The MR images were acquired with 3 T MR scanners from different vendors (GE Healthcare, Philips Healthcare, and Siemens Healthcare). A number of high-resolution T2-weighted MR sequences, including FIESTA, bFFE, CISS, DRIVE, SPACE were used to scan the patients and were included in the study. Images acquired with the FIESTA sequence made up 82% of the MR images. For these, a typical voxel dimension is 0.3125 mm × 0.3125 mm × 0.4 mm.

B. Results

The proposed segmentation method is tested on 132 ears of 66 subjects. The registration and wASM segmentation for each ear takes about 2 minutes. The process is fully automated but failed for 12/132 ears. These cases required a manual alignment between the atlas MR image and the target MR image to localize the cochlea before the wASM segmentation because of imaging artifacts or pathologies that affected the registration process.

Fig. 3 shows DSC and ASD of the ST and the SV for the 132 cochleae. Metrics with "LB" (lower bound) indicate that the evaluation is performed after the point registration. We report mean DSC for the ST and the SV equal to 0.81 and 0.79, respectively. The mean ASD for the ST and the SV are both 0.11 mm, which is far smaller than the typical voxel dimension of the MR images.

IV. Discussion and Conclusions

In this work, we propose a wASM-based method to segment the ICA in T2-weighted MR images. In the fully automated pipeline we have developed, the cochlea is first localized using a series of registrations, and then the wASM is fitted to the cochlea iteratively in the target image. To evaluate the results, we use the same shape model to segment the corresponding CT images and calculate the DSC and ASD between the two segmentation results, achieving a mean DSC of 0.81 and 0.79 for the two ICA structures. These results are promising and show that this automated segmentation method could potentially be used to conduct large-scale studies to, for

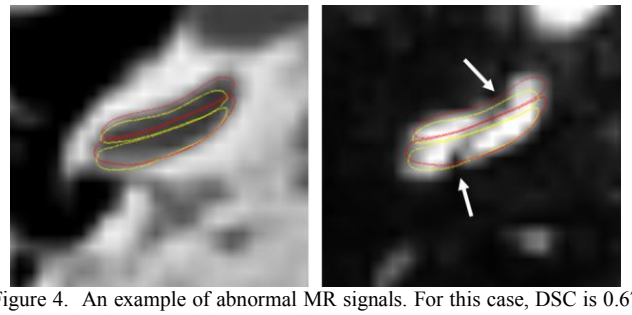

Figure 4. An example of abnormal MR signals. For this case, DSC is 0.67 and 0.59 for the ST and the SV. (yellow contour: segmentation in MR; Red contour: segmentation in CT; white arrows: abnormal MR signals)

instance, find correlations between cochlear MR signal and hearing outcomes over time in AN patients; we have initiated such study. It may also enable the routine and clinical use of this cochlear signal information.

We observe that abnormal MR signal occurs in several cases, e.g., local hypointensity within the cochlea. This may be caused by cochlear pathologies and it affects the segmentation results. Fig. 4 shows such an example but we note that the segmentation in this MR image remains reasonable because of the robust nature of wASM. One possible improvement is to adaptively downweight the outlier points during the fitting. In addition, since the image registration-based cochlear localization fails on 12/132 of the cases, we will explore alternative machine learning-based methods as we have done to register an atlas and other volumes to segment the cochlea in CT images [24].

ACKNOWLEDGMENT

We would like to thank William Rodriguez and Dr. Bob Dwyer for the effort of transferring a large number of images for this study. This research is supported by the National Institutes of Health (NIH) grant R01DC014037, R01DC008408, and R01DC014462 from the National Institute of Deafness and Other Communications Disorders. This work has been supported in part by the NIH, National Institute of Biomedical Imaging and Bioengineering Training Grant No. T32EB021937.

REFERENCES

[1] P. Digge, "Imaging Modality of Choice for Pre-Operative Cochlear Imaging: HRCT vs. MRI Temporal Bone," *J. Clin. Diagn. Res.*, 2016, doi: 10.7860/JCDR/2016/18033.8592.

[2] V. M. Joshi, S. K. Navlekar, G. R. Kishore, K. J. Reddy, and E. C. V. Kumar, "CT and MR Imaging of the Inner Ear and Brain in Children with Congenital Sensorineural Hearing Loss," *RadioGraphics*, vol. 32, no. 3, pp. 683–698, May 2012, doi: 10.1148/rg.323115073.

[3] J. H. Noble, R. F. Labadie, O. Majdani, and B. M. Dawant, "Automatic Segmentation of Intracochlear Anatomy in Conventional CT," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, pp. 2625–2632, Sep. 2011, doi: 10.1109/TBME.2011.2160262.

[4] J. H. Noble, R. F. Labadie, R. H. Gifford, and B. M. Dawant, "Image-Guidance Enables New Methods for Customizing Cochlear Implant Stimulation Strategies," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 5, pp. 820–829, Sep. 2013, doi: 10.1109/TNSRE.2013.2253333.

[5] A. J. Bowen, M. L. Carlson, and J. I. Lane, "Inner Ear Enhancement With Delayed 3D-FLAIR MRI Imaging in Vestibular Schwannoma," *Otol. Neurotol.*, vol. 41, no. 9, pp. 1274–1279, Oct. 2020, doi: 10.1097/MAO.0000000000002768.

[6] M. E. Miller *et al.*, "Hearing Preservation and Vestibular Schwannoma: Intracochlear FLAIR Signal Relates to Hearing Level," *Otol. Neurotol.*, vol. 35, no. 2, pp. 348–352, Feb. 2014, doi: 10.1097/MAO.0000000000000191.

[7] K. O. Tawfik, M. McDonald, Y. Ren, O. Moshtaghi, M. S. Schwartz, and R. A. Friedman, "Cochlear T2 Signal May Predict Hearing Outcomes After Resection of Acoustic Neuroma," Otology & Neurotology, Jul. 2021, doi: 10.1097/MAO.0000000000003228.

[8] V. Prabhu *et al.*, "Preserved Cochlear CISS Signal is a Predictor for Hearing Preservation in Patients Treated for Vestibular Schwannoma With Stereotactic Radiosurgery," *Otol. Neurotol.*, vol. 39, no. 5, pp. 628–631, Jun. 2018, doi: 10.1097/MAO.0000000000001762.

[9] J. Haneda, K. Ishikawa, and K. Okamoto, "Better continuity of the facial nerve demonstrated in the temporal bone on three-dimensional T1-weighted imaging with volume isotropic turbo spin echo acquisition than that with fast field echo at 3.0 tesla MRI," *J. Med. Imaging Radiat. Oncol.*, vol. 63, no. 6, pp. 745–750, Dec. 2019, doi: 10.1111/1754-9485.12962.

[10] O. Af, F. Mw, and M. Aw, "Perilymph total protein levels associated with cerebellopontine angle lesions.," *Am. J. Otol.*, vol. 2, no. 3, pp. 193–195, Jan. 1981.

[11] Y. Feng, J. I. Lane, C. M. Lohse, and M. L. Carlson, "Pattern of cochlear obliteration after vestibular Schwannoma resection according to surgical approach," *The Laryngoscope*, vol. 130, no. 2, pp. 474–481, 2020, doi: https://doi.org/10.1002/lary.27945.

[12] N. West, H. C. R. Sass, M. N. Møller, and P. Cayé-Thomasen, "Cochlear MRI Signal Change Following Vestibular Schwannoma Resection Depends on Surgical Approach," *Otol. Neurotol.*, vol. 40, no. 10, p. e999, Dec. 2019, doi: 10.1097/MAO.0000000000002361.

[13] F. C. E. Hill, A. Grenness, S. Withers, C. Iseli, and R. Briggs, "Cochlear Patency After Translabyrinthine Vestibular Schwannoma Surgery," *Otol. Neurotol.*, vol. 39, no. 7, p. e575, Aug. 2018, doi: 10.1097/MAO.0000000000001858.

[14] S. Zhu, W. Gao, Y. Zhang, J. Zheng, Z. Liu, and G. Yuan, "3D automatic MRI level set segmentation of inner ear based on statistical shape models prior," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Oct. 2017, pp. 1–6, doi: 10.1109/CISP-BMEI.2017.8301973.

[15] A. Vaidyanathan *et al.*, "Deep learning for the fully automated segmentation of the inner ear on MRI," *Sci. Rep.*, vol. 11, no. 1, Art. no. 1, Feb. 2021, doi: 10.1038/s41598-021-82289-y.

[16] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models-Their Training and Application," *Comput. Vis. Image Underst.*, vol. 61, no. 1, pp. 38–59, Jan. 1995, doi: 10.1006/cviu.1995.1004.

[17] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imaging*, vol. 16, no. 2, pp. 187–198, Apr. 1997, doi: 10.1109/42.563664.

[18] G. K. Rohde, A. Aldroubi, and B. M. Dawant, "The adaptive bases algorithm for intensity-based nonrigid image registration," *IEEE Trans. Med. Imaging*, vol. 22, no. 11, pp. 1470–1479, Nov. 2003, doi: 10.1109/TMI.2003.819299.

[19] J. C. Benson, M. L. Carlson, and J. I. Lane, "MRI of the Internal Auditory Canal, Labyrinth, and Middle Ear: How We Do It," *Radiology*, vol. 297, no. 2, pp. 252–265, Sep. 2020, doi: 10.1148/radiol.2020201767.

[20] R. F. Labadie and J. H. Noble, "Preliminary Results With Image-guided Cochlear Implant Insertion Techniques," *Otol. Neurotol.*, vol. 39, no. 7, p. 922, Aug. 2018, doi: 10.1097/MAO.0000000000001850.

[21] A. Rivas *et al.*, "Automatic Cochlear Duct Length Estimation for Selection of Cochlear Implant Electrode Arrays," *Otol. Neurotol.*, vol. 38, no. 3, p. 339, Mar. 2017, doi: 10.1097/MAO.0000000000001329.

[22] R. Banalagay, R. F. L. M.d, and J. Noble, "Validation of active shape model techniques for intra-cochlear anatomy segmentation in CT images," in *Medical Imaging 2021: Image Processing*, Feb. 2021, vol. 11596, p. 115961M, doi: 10.1117/12.2582096.

[23] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945, doi: 10.2307/1932409.

[24] D. Zhang, J. Wang, J. H. Noble, and B. M. Dawant, "HeadLocNet: Deep convolutional neural networks for accurate classification and multi-landmark localization of head CTs," *Med. Image Anal.*, vol. 61, p. 101659, Apr. 2020, doi: 10.1016/j.media.2020.101659.