Cochlear Implant Fold Detection in Intra-operative CT Using Weakly Supervised Multi-task Deep Learning

Mohammad M. R. Khan[(✉)], Yubo Fan, Benoit M. Dawant, and Jack H. Noble

Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN 37235, USA
mohammad.mahmudur.rahman.khan@vanderbilt.edu

Abstract. In cochlear implant (CI) procedures, an electrode array is surgically inserted into the cochlea. The electrodes are used to stimulate the auditory nerve and restore hearing sensation for the recipient. If the array folds inside the cochlea during the insertion procedure, it can lead to trauma, damage to the residual hearing, and poor hearing restoration. Intraoperative detection of such a case can allow a surgeon to perform reimplantation. However, this intraoperative detection requires experience and electrophysiological tests sometimes fail to detect an array folding. Due to the low incidence of array folding, we generated a dataset of CT images with folded synthetic electrode arrays with realistic metal artifact. The dataset was used to train a multitask custom 3D-UNet model for array fold detection. We tested the trained model on real post-operative CTs (7 with folded arrays and 200 without). Our model could correctly classify all the fold-over cases while misclassifying only 3 non fold-over cases. Therefore, the model is a promising option for array fold detection.

Keywords: Tip fold-over · cochlear implant · synthetic CT

1 Introduction

Cochlear implantations (CIs) are considered to be a standard treatment in case of individuals with severe-to-profound hearing loss [1]. During CI surgical procedure an electrode array (EA) is implanted in the cochlea for stimulating the auditory nerve. Along the cochlear duct length, the neural pathways are arranged in a tonotopic manner by decreasing frequency [2]. Naturally, these pathways get activated according to their characteristic frequencies present in the incoming sound. After implantation, the electrode arrays are used to stimulate the nerve pathways and induce hearing sensation [3].

During the insertion process, one complication surgeons aim to avoid is called "tip fold-over," where the tip of the array curls in an irregular manner inside the cochlear volume resulting in array folding [4]. This can occur when the tip of the electrode array gets stuck within an intracochlear cavity as the surgeon threads the array into the cochlea, largely blind to the intra-cochlear path of the array and with little tactile

feedback available to indicate the tip of the array is folding [5]. Therefore, further pushing the base of the EA into the cochlea results in folding of the array (shown in Fig. 1) [5]. Tip fold-over can result in many complications which include trauma, damage to residual hearing and poor positioning of the EA inside cochlea, which ultimately leads to poor hearing restoration for the patient. If intra-operative detection of a tip fold-over is possible, the surgeon can address it through re-implantation of the electrode array [6]. In addition, post-operative detection of minor fold-over can help audiologists to deactivate the affected electrodes to attempt to reach a more satisfactory hearing outcome [4].

Although most CI centers do not currently attempt to detect tip foldovers, the current standard approach among sites that do is visual inspection of intraoperative fluoroscopy. However, these identification methods require experience to align the view optimally and limit the radiation exposure during the fluoroscopy [7]. Various studies (McJunkin et al., Sabban et al., Dirr et al., Timm et al., Sipari et al., Gabrielpillai et al., Jia et al., Garaycochea et al.) have reported on other approaches for detecting tip fold-over through CT imaging, NRT (Neural Response Telemetry) and EcochG (Electrocochleography) [7–9]. In some studies, it is reported that the intraoperative electrophysiological measures, such as NRT or EcochG, sometimes fail to identify the tip fold-over cases [9, 10]. Pile et al. [11] developed a robotic system for tip fold-over detection where the support vector machine classifier is used on the EA insertion force profile. This approach is associated with robot assisted insertion techniques. This broad body of work emphasizes the need for an accurate approach for intra and/or post operative fold-over detection. Therefore, the goal of this study is to develop an approach to detect tip fold-overs in cone beam or conventional CT images using state-of-the-art deep neural network-based image analysis.

As tip fold-over cases are reported to be rare, it would be difficult to acquire a substantial number of cases for fully supervised training of any data-driven method. Zuniga et al. [4] studied CI surgeries in 303 ears and reported 6 tip fold-over cases (less than 2%). Dhanasingh et al. [5] investigated 3177 CI recipients' cases from 13 studies and reported 50 tip fold-over cases (1.57%). Only 0.87% (15 cases) tip fold-over cases were reported among 1722 CI recipients according to Gabrielpillai et al. [12]. Dhanasingh et al. [5] analyzed 38 peer reviewed publications and reported that the rate of tip fold-over with certain types of arrays might be as high as 4.7%. Data scarcity thus makes it difficult to curate a balanced training dataset of tip fold-over cases.
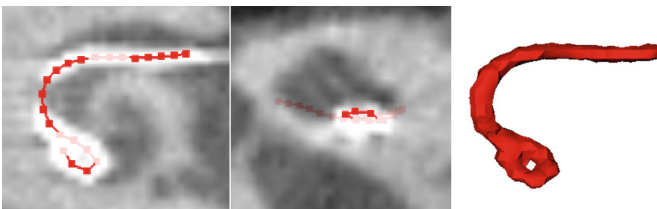

Fig. 1. Coronal view (left) and axial view (middle) of electrode array tip fold-over in a postoperative CT. 3D rendering (right) of a folded over electrode array.

Deep learning methods are the current state-of-the-art in medical image analysis, including image classification tasks. Numerous approaches have been proposed for using 3D networks to solve image classification tasks, e.g. [13–15]. Multi-tasking neural networks have also been used for simultaneous medical image segmentation and classification tasks. These networks have shared layers from the input side which branch into multiple paths for multiple outputs, e.g., [16, 17]. Along with choosing the appropriate network for the CT image classification task one of the typical concerns is the class balance of the training dataset [17]. As tip foldovers are rare, augmentation approaches are needed to reduce the effect of data imbalance.

Therefore, in this work we design a dataset of CT images with folded synthetic arrays with realistic metal artifact to assist with training. We propose a multi-task neural network based on the U-Net [18], train it using the synthetic CT dataset, and then test it on a small dataset of real tip fold-over cases. Our results indicate that the model performs well in detecting fold-over cases in real CT images, and therefore, the trained model could help the intra-operative detection of tip fold-over in CI surgery.

2 Methodology

2.1 Dataset

In this study, we utilize CT images from 312 CI patients acquired under local IRB protocols. This included 192 post-implantation CTs (185 normal and 7 tip-fold), acquired either intra-operatively (cone beam) or post-operatively (conventional or cone beam), and 120 pre-implantation CTs used to create synthetic post-implantation CT. As image acquisition parameters (dimensionality, resolution and voxel intensity) varies among the images, we preprocessed all the CT images to homogenize these parameters. First, the intracochlear structures (e.g., Scala Tympani (ST) and Scala Vestibuli (SV)) were segmented from the CT image using previously developed automatic segmentation techniques [19–21]. Using the ST segmentation, a region-of-interest CT image was cropped from the full-sized CT image keeping the ST at the center of the cropped image. Then the cropped CT resolution was resampled to an isotropic voxel size of 0.3 mm with a $32 \times 32 \times 32$ grid. As the final step of the preprocessing, the voxel intensity of the cropped image was normalized to ensure comparable intensity distribution among all the CT images.

2.2 Synthetic CT Generation

Our synthetic post-operative CT generation approach is inspired by the process of synthetic preoperative CT generation by Khan et al. [22]. First, a random but realistic location for the electrodes is estimated in a real pre-implantation CT image. This was done by considering some constraints, such as the relative location of the ST, electrode spacing, active array length, relative smoothness of the electrode curve, whether a fold exists, and if so, the location of the fold. Randomized variability within plausible margins ensured generating a unique and realistic electrode array shape for each case.

Once we estimated the probable locations of the electrodes, we placed high intensity (around 3–4 times the bone intensity) cubic blocks with a dimensionality of $4 \times 4 \times$

4 voxels in an empty $32 \times 32 \times 32$ grid in locations corresponding to the electrode sites in the preoperative CT. We also added small high intensity cubic blocks (with a dimensionality of $2 \times 2 \times 2$ voxels) between the electrodes to represent the wires that connect to the electrodes. This resulted in an ideal image with a synthetic electrode array (shown in Fig. 2) but lacking realistic reconstruction artifacts. To get a realistic metal artifact, we applied radon transformation on the high intensity blocks to project the data on the detector space. Then, with the resulting sinogram, we applied inverse radon transformation to backproject the sinogram into the world space. The backprojection process is done two times separately: first time with low frequency scaling (to achieve blurry metal edge) and second time with high frequency scaling (to achieve dominant metal artifact). A Hamming filter was used in the backprojection process. Finally, the preoperative CT and the images with back-projected synthetic EA with realistic metal artifact were merged together additively to generate a synthetic postoperative CT image. Wang et al. [23] also presented a method to generate synthetic CT images with metal artifact, however, the aim of the study was to remove metal artifact from post-implant CT images.
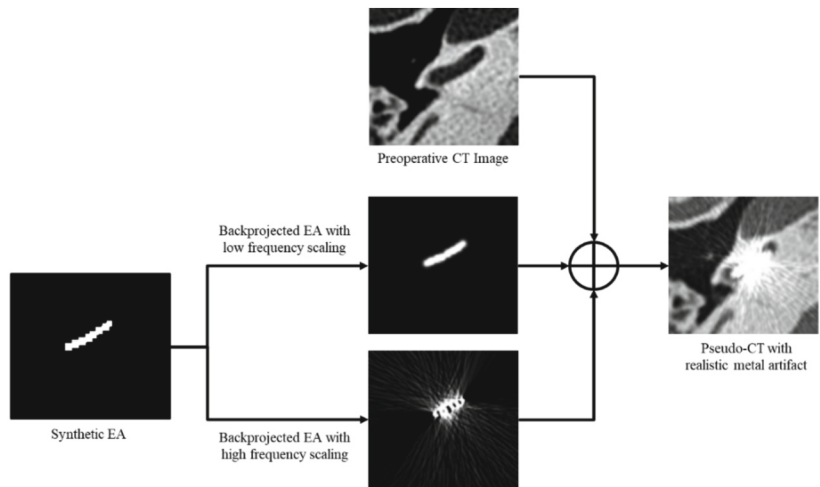

Fig. 2. Synthetic postoperative CT with realistic metal artifact generation process from preoperative CT image.

Table 1. Overall distribution of the dataset.

	Training		Validation		Testing	
	PCT	Real CT	PCT	Real CT	PCT	Real CT
Foldover	135	0	5	0	15	7
Normal	40	160	5	12	15	200
Subtotal	175	160	10	12	30	207
Total	335		22		237	

Using the described synthetic CT generation process, we produced 215 synthetic pseudo CTs (PCTs) with random electrode locations sampled from 100 preoperative real CT images with stratified sampling into the training, validation, and testing datasets. In these PCTs, 155 images have a synthetic EA with tip fold-over and the remaining 60 do not have any fold over. The combined dataset including 379 real CTs and 215 synthetic CTs was divided into training (335), validation (22) and testing (237) subsets. The overall distribution of the dataset is presented in Table 1. As the number of real fold-over cases is very low (about 1.85% in this study), we allotted all of them in the testing dataset.
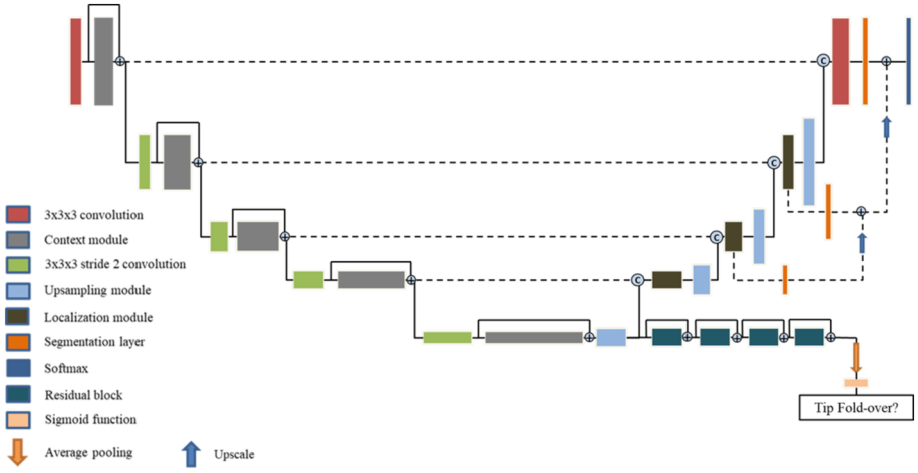

Fig. 3. Multitasking 3D U-Net architecture for EA segmentation and tip fold-over classification.

2.3 Multi-task Deep Learning Network

The neural network model proposed in this study for EA fold-over detection was inspired by the 3D U-Net architectures proposed by Isensee et al. and Ronneberger et al. [18, 24]. The model is a multitasking network where the outputs are the segmentation of the EA and fold-over classification. Our hypothesis is that this multi-task approach helps the network focus attention on the shape of the EA when learning to classify the CT, rather than overfitting to a spurious local minima driven by non-EA features in the training dataset. The architecture is comprised of a context pathway for encoding the increasingly abstract representation of the input as we advance deeper into the neural network. A localization pathway recombines the representation for localizing the EA [20]. In the model, the context modules compute the context pathways. Each of these modules is a pre-activation residual block [25] which has a dropout layer $(p_{dropout} = 0.2)$ in between two convolutional layers of $3 \times 3 \times 3$ dimensionality.

The localization pathways collect features at lower spatial resolution where the contextual information is encoded and transfer it to the higher resolution. This is done by using an upsampling step followed by a convolutional layer. The upsampled features are then concatenated with the corresponding context pathway level. Segmentation layers from different levels of the architecture are convolutional layers that are combined by elementwise summation to build the segmentation in a multi-scale fashion and obtain the final segmentation output. This approach was inspired by Kayalibay et al. [26].

A classification branch is added to the network at the point where the contextual information is encoded at the lowest resolution (shown in Fig. 3). The classification branch consists of 4 residual blocks [27] followed by an average pooling layer and a sigmoid function layer.

We used binary cross entropy (BCE) loss between the EA ground truth, created by manually selected thresholding of the CT image, and the predicted segmentation. For fold-over classification, we also used BCE loss between the ground truth and the predicted class. However, to place emphasis on the classification performance of the model, the classification loss was weighted 5 times the loss for the segmentation. The learning rate and the batch size were considered $3e-5$ and 20, respectively. While training the model, random horizontal flipping and $90°$ rotation were used as data augmentation techniques for generalization.

To evaluate the performance of the proposed model compared to some other neural network models, we implemented 3D versions of ResNet18 [15], Variational Autoencoder (VAE) [28] and Generative Adversarial Network (GAN) [29]. The overall performance comparison among these network architectures is presented in the result section. Similar to the multitasking 3D U-Net, proposed in this study, the VAE and the GAN architectures were designed with the same multi-task objective. In the VAE network, the information of the input image was encoded in 128 latent variables in the encoder section of the model. With these latent variables the decoder section reconstructs a 3D image with the EA segmentation. A classification branch with the same architectures as proposed above was added to the encoder section for fold-over detection. Similarly, the GAN also had a classification branch in the generator model of the architecture. The ResNet18 performs only classification, and thus represents classification performance achievable without multi-task training.

3   Results

The training and validation loss curves of multitasking 3D U-Net are presented in Fig. 4. The graph at the top presents the overall training and validation loss. A rapid drop is visible in the segmentation loss curves where the validation loss curve swiftly follows the training loss curve. On the other hand, the classification loss demonstrates a gradual drop.

Next, we compare the performance of different models for fold-over classification. In addition to the performance analysis of these networks for the whole testing data, we separately reported the performance analysis for the synthetic as well as the real portions of the testing data. The separate analysis provides insight about the applicability of the trained model for the real CT images.
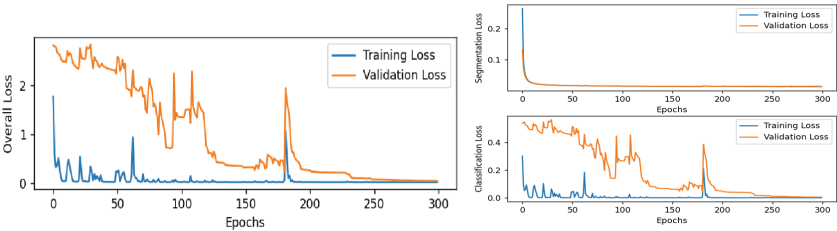
Fig. 4. Training and validation loss curves of 3D U-Net.

As reported in Table 2, the proposed multitask network has a classification accuracy of 98% for all the testing data (237 CT images). Among those 207 (7 tip fold-over and 200 non fold-over cases) are real CT images where the proposed model is 99% accurate regarding the classification by misclassifying 3 non fold-over test cases. In addition, the model misclassified 1 synthetic CT with folded over condition which degraded its accuracy to 97% for synthetic data. Although segmentation is not our primary goal, the model was able to consistently capture the location of the EA (example shown in Fig. 5). Inference time for our network was 0.60 ± 0.10 s.

Table 2. Tip fold-over detection result comparison among different networks.

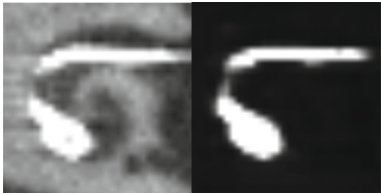| Network architectures		**3D U-Net**	ResNet18	VAE	GAN
	Overall	0.98	0.93	0.87	0.98
Accuracy	Synthetic	0.97	0.90	1.00	1.00
	Real	**0.99**	0.93	0.83	0.96
	Overall	0.95	0.54	0.88	0.88
Sensitivity	Synthetic	0.93	0.80	1.00	1.00
	Real	**1.00**	0.00	0.00	0.00
	Overall	0.99	0.97	0.87	0.99
Specificity	Synthetic	1.00	1.00	1.00	1.00
	Real	**0.99**	0.97	0.86	0.99

Fig. 5. Real CT with tip fold-over condition (left) and segmentation output of the multitasking 3D U-Net (right).

On the other hand, ResNet18 demonstrated lower classification accuracy (around 87%) while misclassifying several synthetic CTs and the real CTs with tip folded over.

In our study, the 3D versions of VAE and the GAN networks rendered promising segmentation results comparable to those of the 3D U-Net. However, in case of tip fold-over detection both the networks classified all the fold-over cases as non fold-overs. This suboptimal performance made our implemented VAE and GAN impractical for intra and/or postoperative fold-over detection.

Table 3 presents results of a hyperparameter settings evaluation study. As the optimizer, we considered the Adam and Stochastic Gradient Descent (SGD). The learning rate was varied between $1e-3$ to $1e-5$; however, the best classification output was obtained using a learning rate of $3e-4$. As stated in the methodology, we assigned a higher weight with the classification loss for emphasizing on the classification branch during the training process. Better classification accuracy was obtained for real CTs when the weight was either 2 or 5. From the classification accuracy, sensitivity, and specificity analysis in Table 3 it is evident that the Adam optimizer with a learning rate of $3e-5$ outperforms the other hyperparameter combinations for real CT fold-over detection. The classification loss weight and the batch size were considered 5 and 20, respectively. Similar hyperparameter analysis was done to select parameters for the other networks evaluated in this study, but not included here for the sake of brevity.

Using Adam optimizer with learning rate $3e-4$, batch size of 20, and classification loss weight of 5, we repeated the training process 8 times to investigate training stability. In one out of eight cases, the resulting model could again correctly classify all the real CTs (7 with folded EA and 13 without) in the testing dataset. For the remaining 7 of 8 models, the network could correctly classify 19 out of 20 real CTs (misclassifying one fold-over case), which results in a classification accuracy of 95% for real postoperative CT images.

Table 3. Tip fold-over detection result comparison among different hyperparameter settings.

Hyperparameters							
	Optimizer	Adam	SGD	SGD	Adam	Adam	Adam
	Learning rate	3E-04	3E-04	5E-05	3E-04	3E-04	3E-05
	Classification loss weight	2	2	2	2	5	5
	Batch size	30	30	30	30	20	20
Accuracy	Overall	0.97	0.98	0.99	0.95	0.97	0.98
	Synthetic	0.97	1.00	1.00	1.00	1.00	0.97
	Real	0.96	0.98	0.99	0.95	0.97	0.99
Sensitivity	Overall	0.77	0.82	0.91	0.95	0.95	0.95
	Synthetic	0.93	1.00	1.00	1.00	1.00	0.93
	Real	0.43	0.43	0.71	0.86	0.86	1.00
Specificity	Overall	1.00	0.99	1.00	0.95	0.97	0.99
	Synthetic	1.00	1.00	1.00	1.00	1.00	1.00
	Real	1.00	1.00	1.00	0.95	0.98	0.99

4   Discussion and Conclusion

In a CI surgical procedure, the relative positioning of the EA influences the overall outcome of the surgery [30, 31]. A tip fold-over case results in poor positioning of the apical electrodes and, hence, can lead to severe consequences including trauma, damage to the residual hearing region and poor hearing restoration. Upon intraoperative detection of such a case, the surgeon can extract and reimplant the array to avoid any folding. The conventional detection processes require experience yet are prone to failure. In addition, due to the incidences of fold-over cases being low, training a model to detect these cases with real data is difficult. Therefore, in this study, we generated a dataset of CT images with folded synthetic electrode arrays with realistic metal artifact. A multitask custom network was proposed and trained with the dataset for array fold detection. We tested the trained model on real post-implantation CTs (7 with folded arrays and 200 without). We were able to train a model that could correctly classify all the fold-over cases while misclassifying only 3 non fold-over cases. In future work, clinical deployment of the model will be investigated.

References

1. US Department of Health and Human Services, National Institute on Deafness and Other Communication Disorders, Cochlear implants, No. 11–4798 (2014)
2. Yukawa, K., et al.: Effects of insertion depth of cochlear implant electrodes upon speech perception. Audiol. Neurotol. **9**(3), 163–172 (2004)
3. Stakhovskaya, O., et al.: Frequency map for the human cochlear spiral ganglion: implications for cochlear implants. J. Assoc. Res. Otolaryngol. **8**(2), 220 (2007)
4. Zuniga, M.G., et al.: Tip fold-over in cochlear implantation: case series. Otol. Neurotol. Official Publ. Am. Otological Soc. Amer. Neurotol. Soc. Eur. Acad. Otol. Neurotol. **38**(2), 199 (2017)
5. Dhanasingh, A., Jolly, C.: Review on cochlear implant electrode array tip fold-over and scalar deviation. J. Otol. **14**(3), 94–100 (2019)
6. Ishiyama, A., Risi, F., Boyd, P.: Potential insertion complications with cochlear implant electrodes. Cochlear Implants Int. **21**(4), 206–219 (2020)
7. Dirr, F., et al.: Value of routine plain x-ray position checks after cochlear implantation. Otol. Neurotol. **34**(9), 1666–1669 (2013)
8. McJunkin, J.L., Durakovic, N., Herzog, J., Buchman, C.A.: Early outcomes with a slim, modiolar cochlear implant electrode array. Otol. Neurotol. **39**(1), e28–e33 (2018)
9. Garaycochea, O., Manrique-Huarte, R., Manrique, M.: Intra-operative radiological diagnosis of a tip roll-over electrode array displacement using fluoroscopy, when electrophysiological testing is normal: the importance of both techniques in cochlear implant surgery. Braz. J. Otorhinolaryngol. **86**, s38–s40 (2020)
10. Cohen, L.T., Saunders, E., Richardson, L.M.: Spatial spread of neural excitation: comparison of compound action potential and forward-masking data in cochlear implant recipients. Int. J. Audiol. **43**(6), 346–355 (2004)

11. Pile, J., Wanna, G.B., Simaan, N.: Robot-assisted perception augmentation for online detection of insertion failure during cochlear implant surgery. Robotica **35**(7), 1598–1615 (2017)
12. Gabrielpillai, J., Burck, I., Baumann, U., Stöver, T., Helbig, S.: Incidence for tip foldover during cochlear implantation. Otol. Neurotol. **39**(9), 1115–1121 (2018)
13. Ahn, B.B.: The compact 3D convolutional neural network for medical images. Standford University (2017)
14. Jin, T., Cui, H., Zeng, S., Wang, X.: Learning deep spatial lung features by 3d convolutional neural network for early cancer detection. In: 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE, Piscataway (2017)
15. Hara, K., Kataoka, H., Satoh, Y.: Learning spatio-temporal features with 3d residual networks for action recognition. In: Proceedings of the IEEE international Conference on Computer Vision workshops, pp. 3154–3160 (2017)
16. Zhang, D., Wang, J., Noble, J.H., Dawant, B.M.: HeadLocNet: deep convolutional neural networks for accurate classification and multi-landmark localization of head CTs. Med. Image Anal. **61**, 101659 (2020)
17. Jnawali, K., Arbabshirani, M.R., Rao, N., Patel, A.A.: Deep 3D convolution neural network for CT brain hemorrhage classification. In: Medical Imaging 2018: Computer-Aided Diagnosis, vol. 10575, pp. 307–313. SPIE (2018)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
19. Noble, J.H., Labadie, R.F., Majdani, O., Dawant, B.M.: Automatic segmentation of intra-cochlear anatomy in conventional CT. IEEE Trans. Biomed. Eng **58**(9), 2625–2632 (2011)
20. Noble, J.H., Dawant, B.M., Warren, F.M., Labadie, R.F.: Automatic identification and 3D rendering of temporal bone anatomy. Otol. Neurotol. **30**(4), 436–442 (2009)
21. Noble, J.H., Warren, F.M., Labadie, R.F., Dawant, B.M.: Automatic segmentation of the facial nerve and chorda tympani using image registration and statistical priors. In: Medical Imaging 2008: Image Processing, vol. 6914, p. 69140P. International Society for Optics and Photonics (2008)
22. Khan, M.M., Banalagay, R., Labadie, R.F., Noble, J.H.: Sensitivity of intra-cochlear anatomy segmentation methods to varying image acquisition parameters. In: Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 12034, pp. 111–116. SPIE (2022)
23. Wang, Z., et al: Deep learning based metal artifacts reduction in post-operative cochlear implant CT imaging. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 121–129. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_14
24. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 287–297. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_25
25. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 630–645. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_38
26. Kayalibay, B., Jensen, G., van der Smagt, P.: CNN-based segmentation of medical imaging data (2017). arXiv preprint arXiv:1701.03056
27. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

28. Tan, Q., Gao, L., Lai, Y.K., Xia, S.: Variational autoencoders for deforming 3d mesh models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5841–5850 (2018)
29. Cirillo, M.D., Abramian, D., Eklund, A.: Vox2Vox: 3D-GAN for brain tumour segmentation. In: Crimi, A., Bakas, S. (eds.) BrainLes 2020. LNCS, vol. 12658, pp. 274–284. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72084-1_25
30. Rubinstein, J.T.: How cochlear implants encode speech. Curr. Opin. Otolaryngol. Head Neck Surg. **12**(5), 444–448 (2004)
31. Wilson, B.S., Dorman, M.F.: Cochlear implants: current designs and future possibilities. J. Rehabil. Res. Dev. **45**(5), 695–730 (2008)