Automatic Segmentation and 3D Reconstruction of the Inner Ear Vasculature from Histology Slides

Final Report

Aseem Jain

Table of Contents

Clinical Motivation

Researchers have often relied on analysis of histopathology slides of the inner ear from deceased humans/mammals to understand the impact that different structures have on the various pathologies in otolaryngology. Understanding the pathophysiology of the inner ear remains crucial to guiding treatment strategies in otolaryngology. For instance, the stria vascularis, a structure in the inner ear analyzed on histology slides, has been thought to be critical to understanding sensorineural hearing loss in mammals.[1] Additionally, through studying histology slides of patients with inner ear malformation, researchers were able to create a better model for cochlear implants in this population.[2]

One area of interest within the realm of inner ear histopathology is the study of vascular in the inner ear diseases. Currently, vascular disorders of the inner ear remain poorly understood. Some researchers have hypothesized that abnormal vasculature may play a role in pathologies such as vestibular neuritis.[3] These researchers note that being able to analyze the vasculature with 3D models of histopathology could prove to be valuable as it would enable them to better understand the positional relationship between vasculature and adjacent structures.

However, despite the importance of analyzing histology slides, most researchers still rely on manual methods for segmentation of structures within the inner ear. Manual segmentation techniques can also lead to poor interrater reliability.[4] Furthermore, few researchers have the tools to be able to generate 3D models of structures within the inner ear. While automatic techniques to extract relevant features from whole slide histology images (WSI) is rapidly expanding, little work has been done to automate segmentation histology slides of the inner ear and construction 3D models from these segmentations.

Prior work

Prior groups have tackled the issue of segmenting and analyzing WSI using neural networks. Segmenting WSI poses many challenges including 1) large image files often on the order of 1-2 GB per image 2) stain variability between slices 3) extracting clinically relevant features from large images. Khened et al proposed a generalized framework for addressing these issues; the researchers used an ensemble of neural networks that were based upon a U-net with various backbones to improve the overall accuracy of the model.[4] This group primarily used a Jaccard Index, measure of overlap between generated labels and ground truth, to quantify their accuracy and achieved a score of .67-78 for certain datasets. Guo et al focused on creating quick, yet still accurate, deep learning methods to analyze breast images. They used a patch-based method to extract cancerous regions and then evaluated these regions at different resolutions to gather course and fine features.[5] Isensee et. al, in 2021, demonstrated that their implementation of U-net, nnUnet, which automatically optimizes hyperparameters of traditional U-net creates a superior framework for image segmentation.[7] While originally designed for 3D CT scans, yhey

have since updated the model proposed in their original paper to facilitate easier segmentation of 2D and 3D structures.

Previous groups have generated 3D reconstructions of histology slides for a variety of applications. Büki et. al manually aligned segmented histology slides together to generate a 3D model.[3] The researchers anecdotally described this process as incredibly time consuming. Other researchers have used tools such as *SimpleElastix* built in to perform alignment and registration of histology images to generate 3D models of bladder cancers.[6]

Goals and Significance

Broad Goals

The key deliverable at the end of this project is to create software that can

1) Use deep learning methods to segment histopathology slides of the ear
2) Use these segmentations to create a 3D reconstruction of the ear

Significance

Completion of this project will enable researchers to analyze WSI of the inner ear rapidly and accurately compared to existing manual techniques. Additionally, by adding the ability to view these segmentations in 3D, researchers will have new insights into the role that structures such as the vasculature can play in various pathologies. Creation of this software could also facilitate the ability to study other structures within inner ear WSI. The proposed project would be first example of reconstructing vasculature from inner ear histology slides using deep learning. Overall, the proposed project has the potential to dramatically improve the way researchers analyze WSI within the inner ear.

General Experimental Setup

Temporal bone slices of 4 macaque monkey ears have already been sectioned for this project. Each ear contains 110 slides spaced 10 micrometers apart. The ears themselves are assumed to have no pathologies associated with them. These slides were converted into digital images using an Olympus microscope setup in the Lauer lab at JHMI.

To facilitate quicker creation of the software and the associated deep learning framework, access to a virtual GPU has already been provided. The configuration of the GPU is a GeForce RTX 3090 with 32 Gb of RAM. Software for this project was primarily written in Python and Matlab using various dependencies. Additionally, 3D Slicer software was used for 3D image visualization and analysis.

Technical Approach

Broadly, the technical approach can be divided into two phases. 1) Creation of the deep learning framework for semantic segmentation of inner ear vasculature. 2) Alignment of segmentations

and generation of 3D model. An overview of the pipeline created to accomplish project goals is shown below.

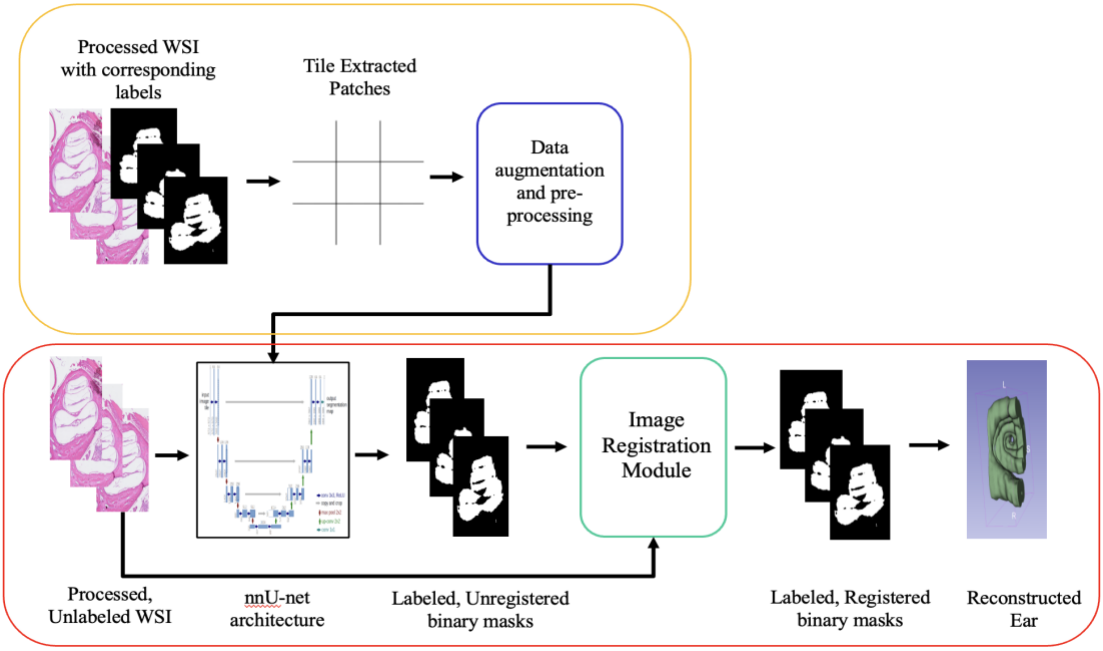


Figure 1. Overview of proposed framework; (top, orange) represents the training pipeline, (bottom, red) represents the inference and reconstruction pipeline.

Deep Learning Pipeline

As seen in Figure 2, two structures were specifically labeled for training: the vasculature in the mid-modiolus of the inner ear and the scala media, scala vestibuli, and scala tympani (these were combined and referred to as scalas). Segmenting the scalas was critical as it gave spatial context to the vasculature and could be a useful structure to study in the future. Each slide was ~10000-pixel x 10000-pixel. Of the 110 slides in the database, 30 were randomly sampled for labeling.

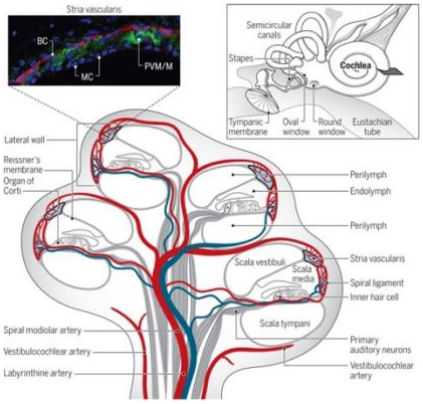


Figure 2. Graphical representation of inner ear[3]

The training process for both structures was done within a Roboflow© environment. For the scalas, each slide and corresponding label was first compressed to 512x512 pixel image and then augmented by rotation from -15 to 15 degrees as well as magnified from 0-20% percent. Overall, 30 scalas were represented in 267 images. Data augmentation remains a valuable tool to increase deep learning generalizability when working with relatively small datasets. Instead of compressing the original image by ~20x, training for the vascular segmentation model involved tiling a ~4x compressed slide (2048x2048 pixel) into 512x512 patches before data augmentation. The vasculature in the inner ear is a relatively small structure

and is often lost if represented in an overly compressed image. Overall, 1312 512x512 patches were used for vascular segmentation model.

The nnUnet architecture has already been described in detail in other works.[7] nnUnet has additional pre-processing steps including normalizing the intensity range of the image. The nnUnet was trained on 5 folds of data with a training, validation, testing split of 60-20-20. Each fold was trained for a maximum of 20-30 epochs. After training, the optimal weights of nnUnet based on the 5 folds were chosen and used for inference.

Image Registration and 3D Reconstruction

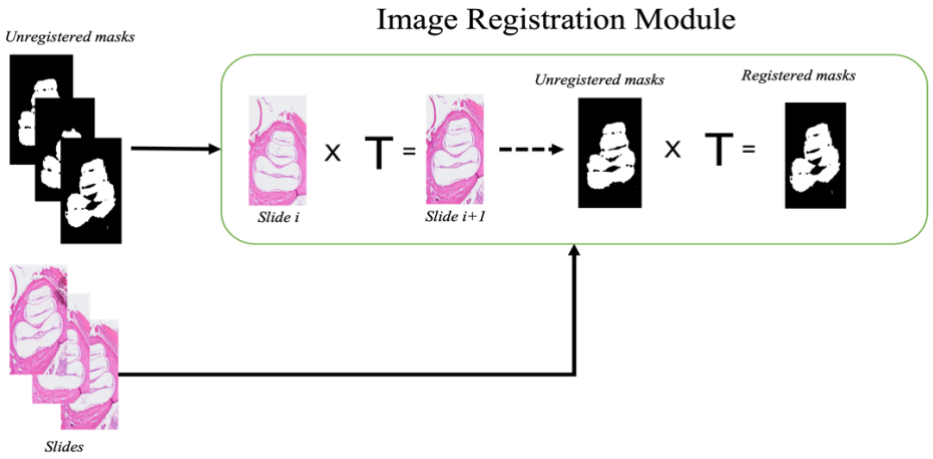


Figure 3. Representation of image registration module. Transform T is computed by aligning one WSI to the subsequent WSI; T is then applied to masks to align them.

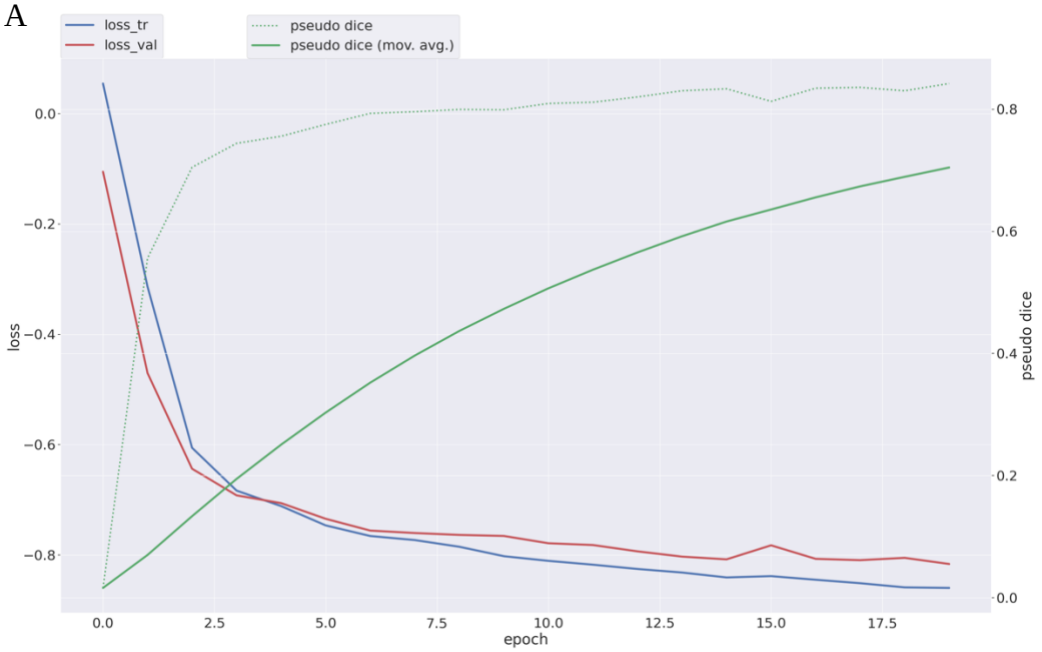
Once the deep learning pipeline produced binary masks, the image registration module aligned these masks. This done in Python using the *SimpleElastix* toolbox. The overall process of this module is outline in Figure 3. Broadly, a compressed (512 x 512px) WSI and its subsequent WSI were aligned by finding a transform T. This transform T was then applied to both the vascular mask as well as the scalas mask to align them to the previous slide. The transform T is represented by two separate processes. First, an affine transform T_1 was computed between the two images (I_1 and I_2) to generate an initial aligned image I_{1a}. Then, a rigid transform T_2 is computed between I_{1a} and I_2. The final transform T can be represented as $T_1 \times T_2$. The method for calculating each type of transform is a traditional iterative closest point algorithm implemented by *SimpleElastix.*

Once the aligned images were produced, they were uploaded in 3D Slicer to produce a volume rendering. This was done through the SlicerMorph extension which enables users to upload a series of images as a volume stack.

Results

Training Model Results

A



B

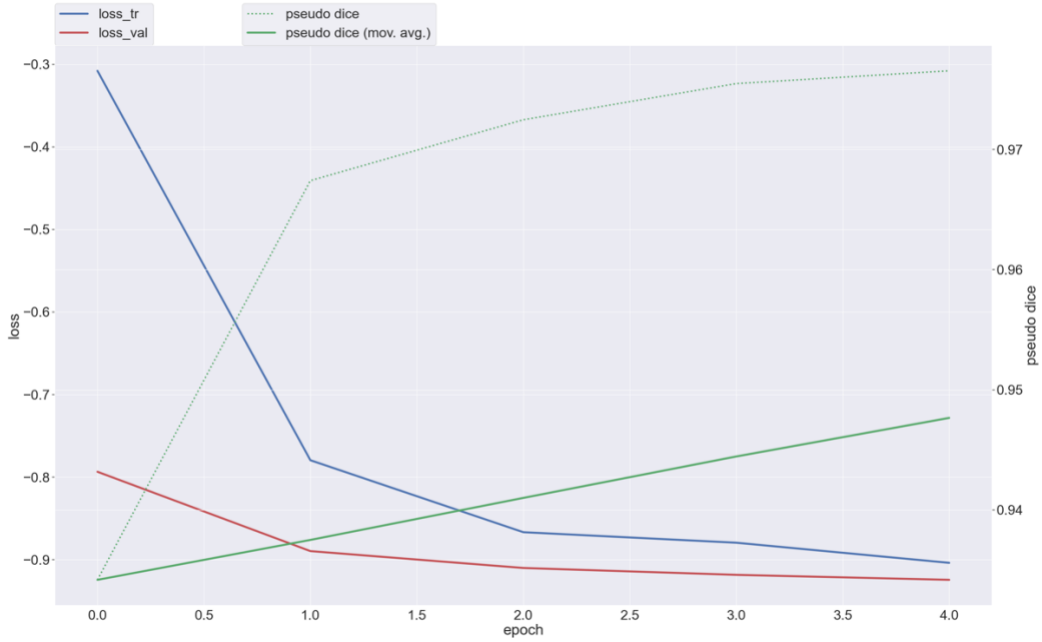


Figure 4. A) Training results for vasculature segmentation with nnUnet; B) Training results for scala segmentation model with nnUnet.

The results of the plots above are summarized in the table below. DICE and IoU were the two primary metrics for evaluating segmentation accuracy. In general, both are a measure of the amount of overlap between the ground truth and the labels generated. A DICE score or IoU score

closer to 1 means a greater degree of overlap and therefore more accurate model. The segmentation network for the scalas outperformed the vasculature one; however, both models achived DICE coefficients greater than 0.7 suggesting an overall accurate segmentation.

	Scalas Segmentation	Vasculature Segmentation
DICE	0.975	0.746
IoU	0.944	0.6636

Table 1. DICE and IoU scores for respective segmentation networks

A sample result is of the labels overlayed with the original WSI is shown below.

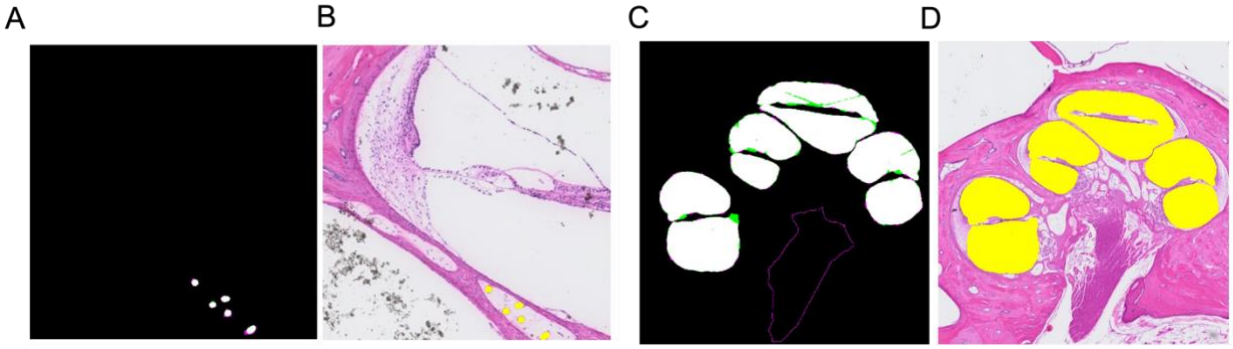


Figure 5. (A, C) Sample overlap between ground truth(GT) and label for vasculature and scala segmentation (purple = GT, green = label) respectively. (B,D) Overlay of label generated on WSI of vasculature and scala

An initial 3D reconstruction of the vasculature and the scalas outlined in 3D Slicer is demonstrated below.

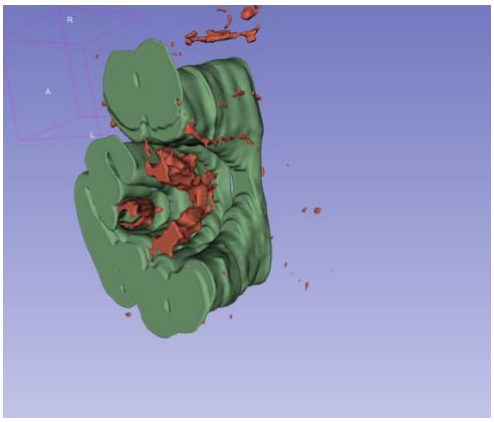


Figure 6. 3D Reconstruction of inner; green represents scalas, red represents vasculature

Discussion

Overall, this project demonstrates that we can segment and reconstruct the vasculature from histology in 3D. The deep learning aspect of this project performed well; meeting the acceptance criteria of achieving DICE scores greater than 0.7. The 3D reconstruction aspect of this project, while functional, requires further processing to generate smoother images.

This pipeline can easily be extended to analyze other layers as well as other structures within the inner ear. The clinical implications of visualizing and studying data in this manner are numerous. Being able to analyze structures in 3D that were previously processed in 2D give researchers a sense of how various pathologies globally affect regions with the inner ear. Researchers have hypothesized the there is a spatial distribution to vascular pathologies affecting the ear. As such, having 3D information about the vasculature could provide more insights into understanding hearing. Further, having a pipeline that can segment structures automatically and quickly facilitates larger scale studies that can determine how various pathologies impact structures within the inner ear at scale.

There are some key limitations to this proposal. The dataset used was primarily generated from one ear of a monkey. As such, the model trained may suffer from lack of generalizability. Further, some slides in this dataset were discarded due to poor quality or significant artifact from the sectioning process. Additionally, training nnUnet is computationally very expensive and requires certain hardware features such as multiple GPUs. That said, inference once a model is trained remain a relatively computationally inexpensive process.

Milestones and Deliverables

	Activities	**Deliverable**	**Status**
Minimum	Scan and Label Appropriate structure from 1 monkey ear	40 labeled WSI; 81 scanned slides total	Complete
	Implement U-Net with proposed algorithm	Functional software that segments vasculature from WSI	Complete
	Validation of U-Net on test data	Internal validation report that analyzes non-labeled slides	Complete
Expected	Align segmented images	Software that can align WSI slides; validate against manual alignment	Complete
	Reconstruction in 3D	3D mesh file that can be assessed by neurotologist	Complete
Maximum	Manuscript writing	Submitted manuscript	End of May
	Analysis of human temporal bone	3D mesh file that can be assessed by neurotologist. Report that includes clinically relevant features such as volume, curvature, etc.	TBD

Management Summary

Past Semester Evaluation

The timeline of this project changed throughout the course of the semester due to some technical limitations. Initial digitization of the WSI of the monkey ear was done incorrectly. Additionally, the time required for labeling datasets of WSI was grossly underrepresented. Both issues set the original timeline of the project significantly back; however, I was able to accomplish the expected goal of having a 3D mesh file that can be analyzed by an otolaryngologist. While more work is required to refine this process, the work done this semester demonstrates initial proof of concept.

Future work

There are 5 key deliverables I would like to continue to work on over the course of the next few months.

1. Write and publish a manuscript by the end of May on current pipeline.
2. Plan on expanding pipeline to include other structures such as cranial nerve VII and VIII and stria vascularis (initial results already collected)
3. Expand metrics used to evaluate both segmentation and subsequent registration.
4. Build model into an easy-to-use GUI that facilitates other researchers to study histology slides
5. Analyze 4 complete macaque ears as well as some human ears.

Dependencies

Of the 5 goals outlined, the first four can be accomplished virtually and as such have no physical dependencies. The team already has access to a virtual GPU for future analysis. Goal 5 may require use of the microscope in the Lauer Lab; however, other medical students may be able to digitize slides for the team.

What was learned

Working with WSI imaging is challenging for numerous reasons:

1. Unlike other modalities like CT which work well for 3D reconstruction, histology slides can have a lot of variation between adjacent slices. As such, aligning and segmenting them was difficult.

2. WSI are large and segmenting vasculature proved to be an incredibly time-consuming task despite limiting the scope to just a portion of the inner ear.

Overall, this project has demonstrated how deep learning and image registration can transform a tedious process into an efficient one that can facilitate improved analysis.

Technical Appendix

https://github.com/2014ajain/Vasculature_Seg_Recon.git

Reference

1. Yu W, Zong S, Du P, Zhou P, Li H, Wang E, Xiao H. Role of the Stria Vascularis in the Pathogenesis of Sensorineural Hearing Loss: A Narrative Review. Front Neurosci. 2021 Nov 19;15:774585. doi: 10.3389/fnins.2021.774585. PMID: 34867173; PMCID: PMC8640081.
2. Monsanto RDC, Sennaroglu L, Uchiyama M, Sancak IG, Paparella MM, Cureoglu S. Histopathology of Inner Ear Malformations: Potential Pitfalls for Cochlear Implantation. Otol Neurotol. 2019 Sep;40(8):e839-e846. doi: 10.1097/MAO.0000000000002356. PMID: 31361687; PMCID: PMC7377297.
3. Büki B, Mair A, Pogson JM, Andresen NS, Ward BK. Three-Dimensional High-Resolution Temporal Bone Histopathology Identifies Areas of Vascular Vulnerability in the Inner Ear. Audiol Neurootol. 2022;27(3):249-259. doi: 10.1159/000521397. Epub 2021 Dec 29. PMID: 34965531; PMCID: PMC9133178.
4. Khened, M., Kori, A., Rajkumar, H. *et al.* A generalized deep learning framework for whole-slide image segmentation and analysis. *Sci Rep* 11, 11579 (2021). https://doi.org/10.1038/s41598-021-90444-8
5. Guo, Z., Liu, H., Ni, H. *et al.* A Fast and Refined Cancer Regions Segmentation Framework in Whole-slide Breast Pathological Images. *Sci Rep* 9, 882 (2019). https://doi.org/10.1038/s41598-018-37492-9
6. Jansen, I., Lucas, M., Savci-Heijink, C.D. *et al.* Three-dimensional histopathological reconstruction of bladder tumours. *Diagn Pathol* 14, 25 (2019). https://doi.org/10.1186/s13000-019-0803-7
7. Isensee, F., Jaeger, P.F., Kohl, S.A.A. et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18, 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z