



OPEN Deep learning for 3D vascular segmentation in hierarchical phase contrast tomography: a case study on kidney

Ekin Yagis^{1✉}, Shahab Aslani^{1,4,10}, Yashvardhan Jain^{3,10}, Yang Zhou¹, Shahrokh Rahmani^{1,9}, Joseph Brunet^{1,2}, Alexandre Bellier⁵, Christopher Werlein⁶, Maximilian Ackermann⁷, Danny Jonigk⁸, Paul Tafforeau^{2,10}, Peter D. Lee^{1,10} & Claire L. Walsh^{1,10}

Automated blood vessel segmentation is critical for biomedical image analysis, as vessel morphology changes are associated with numerous pathologies. Still, precise segmentation is difficult due to the complexity of vascular structures, anatomical variations across patients, the scarcity of annotated public datasets, and the quality of images. Our goal is to provide a foundation on the topic and identify a robust baseline model for application to vascular segmentation using a new imaging modality, Hierarchical Phase-Contrast Tomography (HiP-CT). We begin with an extensive review of current machine-learning approaches for vascular segmentation across various organs. Our work introduces a meticulously curated training dataset, verified by double annotators, consisting of vascular data from three kidneys imaged using HiP-CT as part of the Human Organ Atlas Project. HiP-CT pioneered at the European Synchrotron Radiation Facility in 2020, revolutionizes 3D organ imaging by offering a resolution of around 20 $\mu\text{m}/\text{voxel}$ and enabling highly detailed localised zooms up to 1–2 $\mu\text{m}/\text{voxel}$ without physical sectioning. We leverage the nnU-Net framework to evaluate model performance on this high-resolution dataset, using both known and novel samples, and implementing metrics tailored for vascular structures. Our comprehensive review and empirical analysis on HiP-CT data sets a new standard for evaluating machine learning models in high-resolution organ imaging. Our three experiments yielded Dice similarity coefficient (DSC) scores of 0.9523, 0.9410, and 0.8585, respectively. Nevertheless, DSC primarily assesses voxel-to-voxel concordance, overlooking several crucial characteristics of the vessels and should not be the sole metric for deciding the performance of vascular segmentation. Our results show that while segmentations yielded reasonably high scores—such as centerline DSC ranging from 0.82 to 0.88, certain errors persisted. Specifically, large vessels that collapsed due to the lack of hydrostatic pressure (HiP-CT is an ex vivo technique) were segmented poorly. Moreover, decreased connectivity in finer vessels and higher segmentation errors at vessel boundaries were observed. Such errors, particularly in significant vessels, obstruct the understanding of the structures by interrupting vascular tree connectivity. Our study establishes the benchmark across various evaluation metrics, for vascular segmentation of HiP-CT imaging data, an imaging technology that has the potential to substantively shift our understanding of human vascular networks.

Keywords Deep learning, X-ray tomography, Semantic segmentation, 3D vascular segmentation

¹Department of Mechanical Engineering, University College London, London, UK. ²European Synchrotron Radiation Facility, Grenoble, France. ³Department of Intelligent Systems Engineering, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, USA. ⁴Centre for Medical Image Computing, University College London, London, UK. ⁵Laboratoire d'Anatomie Des Alpes Françaises, Grenoble, France. ⁶Institute of Pathology, Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany. ⁷Institute of Functional and Clinical Anatomy, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany. ⁸Institute of Pathology, RWTH Aachen University, Pauwelsstrasse 30, 52074 Aachen, Germany. ⁹National Heart and Lung Institute, Faculty of Medicine, Imperial College London, London, UK. ¹⁰These authors contributed equally: Shahab Aslani, Yashvardhan Jain, Paul Tafforeau, Peter D. Lee and Claire L. Walsh. ✉email: e.yagis@ucl.ac.uk

The vascular system consists of the network of blood vessels responsible for circulating blood, delivering oxygen, and facilitating the removal of waste from tissues and organs¹. Abnormalities in the structure or function of vascular networks can lead to, or be indicative of, various pathologies ranging from tumor growth and metastasis to strokes and cardiovascular disorders. Hence, accurate segmentation of vasculature and quantitative evaluation of its morphology are widely used to better understand these pathophysiological processes^{2–5}.

Annotation of images by experts has been regarded as the segmentation gold standard, but it is time-consuming and requires specialised knowledge⁶. The success of convolutional neural networks (CNN) in classification leads researchers to employ deep learning technologies for image segmentation^{7,8}. Consequently, a growing number of automated and semi-automatic vessel segmentation methods have been developed over the last years and applied in the diagnosis of diseases associated with the vascular system^{4,6,9}.

However, the complexity of vascular imaging data and the anatomical variability among subjects make vessel segmentation a challenging task. Despite the clinical need, a fully automated segmentation method for 3D vascular segmentation and subsequent feature extraction has not yet been developed due to several specific challenges¹⁰. The main hurdles are the variations across scales, differences in individual anatomies, complexities of slender structures, a limited volume ratio (i.e. vessel/background), as well as a scarcity of publicly available labeled data. In terms of individual anatomical variations, there are significant disparities in vessel length, diameter, and tortuosity, complicating the tasks of vascular tracking and segmentation^{10–13}. Additionally, narrower vessels with diameters at the border of the imaging resolution are particularly difficult to capture^{10,14}.

For segmentation algorithms to have application in biomedical imaging, they must handle a broad spectrum of anatomical and sample variations. They should also be versatile enough to operate across various imaging modalities that cover different length scales and be applicable to both in vivo and ex vivo imaging. For in vivo vascular imaging, magnetic resonance angiography (MRA) is one of the most common choices due to its noninvasive nature, the absence of ionizing radiation exposure, its capability for non-contrast examination, and its ability to provide volumetric representations that highlight vascular disease¹⁵.

Hierarchical Phase-Contrast Tomography (HiP-CT)¹⁶ is an ex vivo imaging technology developed in 2020 on the BM05 beamline at the European Synchrotron Radiation Facility (ESRF)'s Extremely Brilliant Source (EBS). The technique is a propagation-based phase-contrast tomography technique that utilises the increased brilliance and resulting high spatial coherence of the European Synchrotron Radiation Facility's upgrade to a 4th Generation X-ray source (The Extremely Brilliant Source (EBS)). With HiP-CT, it is possible to resolve very small refractive index difference (density difference) in soft tissue structures, even in very large samples such as whole adult human organs. The propagation through free space of the X-rays, after interaction with the sample allow the refractive index shifts to be converted to intensity variation through single-distance phase retrieval at reconstruction¹⁷. HiP-CT has been used to image ex vivo whole human organs, down to resolution of some single-cell types in regions of interest (e.g. Purkinje cells in cerebellar cortex)¹⁶.

A specific novelty of HiP-CT imaging as it relates to the vasculature, lies in its hierarchical nature. Whole human organs can be imaged with overview resolution of $\sim 20 \mu\text{m}/\text{voxel}$, then followed by high resolution (down to $\sim 1\text{--}2 \mu\text{m}/\text{voxel}$) regions of interest anywhere within the sample without sectioning (Fig. 1). This unique feature allows researchers to image the hierarchical structure of the vascular tree from the largest vessels down to near the capillary bed in intact human organs. This novel capability has already allowed unique insights into vascular abnormalities in COVID-19 lung lobes to be identified^{18–20}, as well as vascular quantification across the whole organ²¹. Despite its advantages and suitability for vessel imaging, HiP-CT also poses many challenges for CNN-based methods of vascular segmentation:

- Large data volume (routinely $\geq 1\text{Tb}$) require cutting-edge hardware
- Multi-scale vessels to segment
- Lack of available ground truth training data due to the novelty of the method
- Phase contrast generates fringes in intensity at vessel boundaries
- Large vessels can collapse
- No targeted vascular staining to enhance contrast between vessel lumen and surrounding tissue

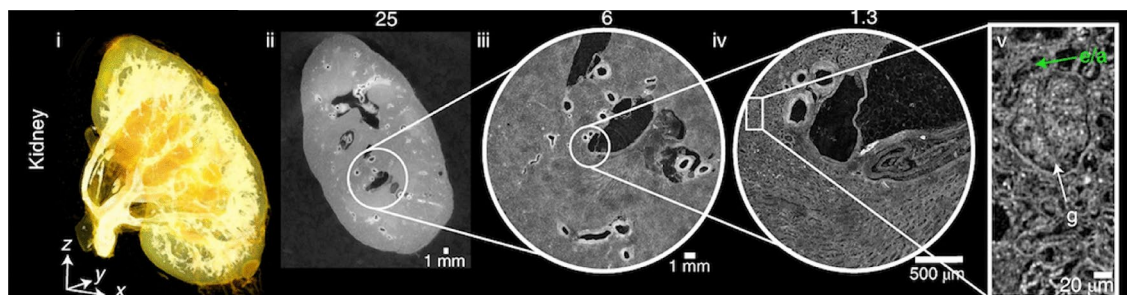


Fig. 1. Reprint from Walsh et al.¹⁶, HiP-CT of kidney; a 3D rendering (i) of the whole organ is shown using scans at $25 \mu\text{m}$ per voxel. Subsequent 2D slices (ii–iv) show positions of the higher-resolution VOI relative to the previous scan. (v), Digital magnification of the highest-resolution image with annotations depicting characteristic structural features in the kidney (e/a, efferent or afferent arteriole; g, glomerulus) and in the spleen (rp, red pulp; wp, white pulp; a, arteriole; ss, splenic sinus). All images are shown using $2\times$ binning.

- Continually developing technique with ongoing changes in image quality (e.g. improvement to resolution, SNR and CNR) result in heterogeneous datasets. In light of the opportunity and challenges that HiP-CT poses for vascular research, we review the literature on blood vessel segmentation with a particular focus on identifying current models that are most suited to the segmentation of blood vessels from HiP-CT imaging. We review segmentation on an organ-by-organ basis as the vascular and parenchymal tissue structures surrounding the vasculature differ vastly between different tissues. Our ultimate aim is to identify and then develop a model and training strategy which will provide a robust and generalisable method that can be easily adapted to segment multi-scale vascular structures across all organs imaged by HiP-CT.

The main contributions of the article are as follows:

- A comprehensive review of deep learning methods for vascular segmentation
 - Creation of a Hierarchical Phase-Contrast Tomography (HiP-CT) labeled vascular dataset in the human kidney
 - Implementation of the nnU-Net model, tailored for kidney vessel segmentation, using our novel Hierarchical Phase-Contrast Tomography (HiP-CT) dataset
 - In-depth discussion on the challenges of blood vessel segmentation and the significance of performance metrics
 - Establishment of a benchmark for future model evaluations in the context of phase contrast tomography imaging data
- This paper is arranged as follows: In “[Vessel segmentation](#)” Section, deep learning models used for vessel segmentation, the datasets used in their training, and the evaluation metrics in the literature are briefly explained. In “[nnU-Net for kidney vessel segmentation](#)” Section, the most prominent method in the literature is tested on HiP-CT kidney vascular data to establish a baseline of performance and highlight specific challenges in the adaptation of these models to HiP-CT data. The findings from our experiments are presented in “[Results](#)” Section, followed by a discussion of these results in “[Discussion](#)” Section. Finally, We conclude the paper in “[Conclusion and future work](#)” Section with a summary of our conclusions and suggestions for future research directions.

Vessel segmentation

Modalities for 3D vascular imaging

Whilst the ideal vessel segmentation algorithm or model is robust across a range of imaging modalities, in reality, the imaging modality will impose contrast and resolution limits that must be considered in segmentation approaches as well as imaging-specific artefacts (e.g. anisotropic voxels). Here, we review some of the most common imaging modalities for 3D vascular networks spanning both in vivo whole human organ imaging as well as ex vivo human tissue imaging; we contextualise the development of HiP-CT within this field.

The most widely adopted procedure for visualising vessels in human organs in vivo is angiography, i.e. where a contrast agent is injected into the vessels. The major angiographic modalities utilised in clinical practice are computed tomography angiography (CTA), magnetic resonance angiography (MRA), and digital subtraction angiography (DSA) also known as a conventional angiogram.

MRA emerged as the predominant modality in vessel segmentation studies we reviewed, as evidenced in Tables 1, 2, 3 and 4. Of the 31 deep learning-based vessel segmentation studies we reviewed, 9 employed MRA, which was primarily used for cerebral vascular evaluations due to its superior tissue contrast compared to CTA. Contrast agents used in MR and CT imaging do not cross the blood-brain barrier, making MRA the preferred choice in these cases. Commonly available datasets for these studies, such as PEGASUS²², MIDAS²³, SCAPIS²⁴, and 1000PLUS²⁵, also utilise MRA. Additionally, 9 studies utilised CTA, which was applied to different regions including the brain, kidney, coronary, and pulmonary vessels.

MRA can be divided into three categories depending on the technique behind the image acquisition/generation: TOF (time-of-flight), PC (phase contrast), and CE (contrast-enhanced). Among those, TOF-MRA is the most frequently used non-contrast bright-blood technique for imaging the human vascular system^{26,27}. As suggested by its name, TOF MRA relies on a principle known as flow-related enhancement, which happens when completely magnetised blood flows into a slab of magnetically saturated tissue whose signal has been muted by repeated RF-pulses²⁸. However, the long acquisition time and high operational costs make it difficult to employ for cerebral artery visualisation^{29,30}. Furthermore, MRA has been shown to overestimate stenosis compared to other modalities³¹.

CTA, on the other hand, is faster, takes a few minutes to complete, and is more accurate than MRA. However, unlike MRA, all CTAs need the administration of an IV contrast agent and involve radiation exposure, making it less safe.

DSA has been utilised as the gold standard for imaging vessels; nonetheless, this invasive and labor-intensive method is rather costly and comes with discomforts and possible dangers^{32,33}. For the in vivo methods described above the image resolution is on the order of 0.5–2 mm^{27,32}, with the state-of-the-art TOF techniques improving the resolution to at most 0.3–0.5 mm. Capturing the microvasculature (1–100 μm) is only possible in vivo at very shallow depths using e.g. contrasted ultrasound³⁴. Otherwise, microvascular imaging of human tissue is restricted to ex vivo techniques including optical, MRI, X-ray, or ultrasound modalities. These techniques are generally applied to smaller tissue sections^{35,36}, with some notable exceptions^{16,37,38}. Ex vivo techniques, capable of imaging a whole human organ down to 100 μm or less include only three examples to the authors’ knowledge: Light-sheet³⁷, 7T MRI³⁸ and HiP-CT¹⁶. Lightsheet imaging, where a fluorescence excitation beam is focused to a sheet which passes through the sample with the detector of the fluorescence emission arranged orthogonal to the illumination, relies on tissue clearing and staining to render the whole organ optically transparent and to target specific structure. High field MRI can provide higher spatial resolution but not below 100 μm . Of the three

References	Modality	Data source	No. of subj.	ML model	Input	DSC	Add'l perf. metrics
Livne et al. ⁴⁸	TOF MRA	Private (PEGASUS)	66	U-Net	2D patches	0.88	95HD = ~ 47 voxels an AVD = ~ 0.4 voxels
Phellan et al. ⁴⁹	TOF MRA	Private	5	2D CNN	2D patches	Between 0.764 0.786	–
Hilbert et al. ⁵⁰	TOF MRA	Private (PEGASUS+ 7UP+ 1000Plus)	264	3D CNN (BRAVE-NET)	3D patches	0.931	95HD = 29.153, AVD = 0.165
Patel et al. ⁵¹	DSA	Private	100	3D CNN (DeepMedic) 3D U-Net	3D patches	0.94 $\pm$ 0.02 0.92 $\pm$ 0.02 respectively	CAL = 0.84 $\pm$ 0.07 0.79 $\pm$ 0.06 respectively
Tetteh et al. ⁵²	TOF MRA μ CTA	Publicly available	Synthetic data	NN with 2D orthogonal cross-hair filters (DeepVesselNet)	3D volumes	0.79	Prec = 0.77 Recall = 0.82
Garcia et al. ⁵³	3DRA	Private	5	3DUNet-based architectures	3D patches	0.80 $\pm$ 0.06	Prec = 0.75 $\pm$ 0.10 and Recall = 0.90 $\pm$ 0.07
Vos et al. ⁵⁴	TOF-MRA	Private	69	2D and 3D U-Net	2D and 3D patches	0.74 $\pm$ 0.17 0.72 $\pm$ 0.15 respectively	MHD = 47.6 $\pm$ 40.4 5 81.3 $\pm$ 57.0 respectively
Chatterjee et al. ⁵⁵	TOF MRI	Not explicitly stated	11	U-Net MSS	3D patches	0.79 $\pm$ 0.091	IOU = 65.89 $\pm$ 1.25
Zhang and Chen ⁵⁶	TOF-MRA	Not explicitly stated	42	DD-Net	3D patches	0.67	Sensitivity = 67.86 IOU = 33.66
B.Zhang et al. ⁵⁷	TOF-MRA	Publicly available (MIDAS)	109	DD-CNN	3D patches	0.93	PPV = 96.4730, Sensitivity = 90.1443, Acc = 99.9463
Lee et al. ⁵⁸	MRA	Private (SNUBH)	26	2D U-Net with LSTM (Spider U-Net)	2D images	0.793	IOU = 74.3
Quintana et al. ⁵⁹	TOF-MRA	Private (UNAM)	4	Dual U-Net-Based cGAN	2D images	0.872	Prec = 0.895

Table 1. Review of the papers that applied deep learning for brain vessel segmentation. CNN convolutional neural network, GAN generative adversarial network, cGAN conditional GAN, DD-Net dense-dilated neural network, LSTM long short-term memory, U-Net MSS multi-scale supervised U-Net, TOF MRA time of flight magnetic resonance angiography, DSA digital subtraction angiography, μ CTA micro-computed tomography angiography, 3DRA 3D rotational angiographies, DSC Dice similarity coefficient, HD Hausdorff distance, AVD average distance, Prec Precision, MHD Mahalanobis distance, IOU intersection over union, PPV positive predictive value, CAL connectivity-area-length.

References	Modality	Data source	No. of subj.	ML model	Input	DSC	Add'l perf. metrics
Karpinski et al. ⁶⁰	WSI	Publicly available	35	2D UNet with Resnet34 backbone	2D images	–	Acc: 0.893
He et al. ⁶¹	CT	Private	170	DPA-DenseBiasNet	3D volumes	0.861	MCD:1.976
Taha et al. ⁶²	CT	Private	99	Kid-Net (a 3D CNN)	3D patches	–	F1 score: 0.72 (artery); 0.67 (vein)
He et al. ⁶³	CTA	Private	122	EnMcGAN	3D patches	0.89 $\pm$ 0.6 (artery); 0.77 $\pm$ 0.12 (vein)	–
Zhang et al. ⁶⁴	CT	Publicly available	392	DPA-DenseBiasNet	2D images	0.884	–
Xu et al. ⁶⁵	Micro-CT	Private	8	CycleGAN	3D patches	0.768 $\pm$ 0.3	Acc: 0.992
Li et al. ⁶⁶	CT	Publicly available	35	DUP-Net	3D patches	0.883	Precision:0.911; Recall:0.858

Table 2. Review of the papers that applied deep learning for kidney vessel segmentation. CNN convolutional neural network, GAN generative adversarial network, WSI whole slide imaging, DPA deep priori anatomy, MCD mean centerline distance, EnMcGAN ensemble multi-condition GAN, DUP-Net double UPoolFormer networks, DSC Dice similarity coefficient.

techniques, HiP-CT can perform multi-scale imaging far more simply than light sheet imaging as HiP-CT does not rely on staining of the vasculature, or clearing of the sample shortening sample preparation^{21,37}. In addition the imaging itself of a whole organ at 20 μ m/voxel can be performed in as little as 1 hr compared to the MRI acquisitions 100 μ m which took hundreds of hours^{21,38}.

Before the arrival of complex machine learning methods, vessel segmentation relied heavily on traditional techniques such as kernel-based methods, tracking methods, mathematical morphology-based methods,

References	Modality	Data source	No. of subj.	ML model	Input	DSC	Add'l perf. metrics
Dong et al. ⁶⁷	CCTA	Private	338	Di-Vnet	3D patches	0.902	Prec = 0.921, Recall = 0.97
Gao et al. ⁶⁸	XCA	Private	130	GBDT	2D images	–	F1 = 0.874, Sen = 0.902, Spec = 0.992
Wolterink et al. ⁶⁹	CCTA	Publicly available	18	GCN	2D images	0.74	MSD = 0.25 mm
Li et al. ⁷⁰	CCTA	Private	243	2D U-Net with 3DNet	2D images	0.771 ± 0.021	AUC = 0.737
Song et al. ⁷¹	CCTA	Private	68	3D FFR U-Net	3D patches	0.816	Prec = 0.77, Recall = 0.87
Zeng et al. ⁷²	CCTA	Publicly available	1000	3D-U-Net	3D volumes	0.82	–

Table 3. Review of the papers that applied deep learning for coronary vessel segmentation. *DSC* Dice similarity coefficient, *Prec* Precision, *Sen* sensitivity, *Spec* specificity, *GBDT* gradient boosting decision tree, *CCTA* coronary computed tomographic angiography, *XCA* X-ray coronary angiography, *GCN* graph convolutional networks, *MSD* mean surface distance, *AUC* area under the receiver operating characteristic curve, *3D FFR U-Net* 3D feature fusion and rectification U-Net.

References	Modality	Data source	No. of subj.	ML model	Input	DSC	Add'l perf. metrics
Tan et al. ⁷³	CT and CTA	Publicly available	16	2D-3D U-Net nnU-net	2D images 3D volume	0.786, 0.797, 0.77 (on CT)	OR = 0.281, 0.285, 0.304
Nam et al. ⁷⁴	CTA	Private	104	3D U-Net	3D patches	0.915 ± 0.31	AUC = 0.995
Guo et al. ⁷⁵	CT	Private	50	3D CNN	3D patches	0.943	–
Xu et al. ⁷⁶	CT	Private		2D CNN		–	–
Nardelli et al. ⁷⁷	CT	Publicly available	55	3D CNN	2D and 3D patches	–	Sen = 0.93 Prec = 0.83
Wu et al. ⁷⁸	CT	Private	143	MSI-U-Net	3D volume	0.7168	Sen = 0.7234, Prec = 0.7893

Table 4. Review of the papers that applied deep learning for pulmonary vessel segmentation. *CNN* convolutional neural network, *CT* computed tomography, *CTA* computed tomography angiography, *MSI-U-Net* multi-scale interactive U-Net, *OR* over segmentation rate, *AUC* area under the receiver operating characteristic curve, *Sen* sensitivity, *Spec* specificity, *Prec* precision.

and model-based methods³⁹. In kernel-based methods, edge detection techniques, like the Canny or Sobel operators, identify vessel boundaries using gradient information⁴⁰. More sophisticated approaches, such as Hessian-based methods can be categorised under model-based methods⁴¹. These methods utilise second-order intensity derivatives to capture vessel-like structures, and level-set methods evolve contours within the image domain to detect vessels. Active contours, or snakes, are tracking methods dynamically adjusted to capture vessel boundaries, and mathematical tools like B-splines offer precise modeling of vessel structures^{42,43}. These traditional methods provided a foundation that paved the way for machine learning in the realm of medical image analysis.

With the rapid advancement of computational capabilities, a paradigm shift has taken place in the realm of image processing and machine learning: the rise of deep neural networks (NNs). These NNs, employing multi-layered architectures, extracted patterns within data through multiple levels of abstraction, and transformed the way images are understood and processed.

Among various NN architectures, CNNs have stood out as highly effective for image analysis tasks. They possess the capability to learn image features hierarchically, progressing from simple to complex, thereby eliminating the need for manual feature extraction. This inherent feature has made CNNs especially popular in the field of medical image analysis.

Recently, the landscape of computer vision has seen another transformative shift with the emergence of Transformers⁴⁴. Initially successful in natural language tasks, Transformers have begun to find application in various computer vision challenges, prompting researchers to reevaluate the dominant role of CNNs. These advancements in computer vision have also generated significant interest in the medical imaging domain⁴⁵. Transformers, with their ability to capture a more extensive global context compared to CNNs have now become a subject of considerable attention in the medical imaging community^{46,47}.

Deep learning models for vessel segmentation

Whilst the acquisition of these large vascular network datasets is becoming increasingly fast, automated segmentation has also been a rapidly developing area of research⁷⁹. Before the introduction of machine learning

models, traditional vessel segmentation methods such as the Frangi⁸⁰ and Sato⁸¹ filters were widely employed for their ability to enhance tubular structures in medical images. These filters have played a crucial role in extracting vascular structures, particularly in medical imaging contexts like angiography, and continue to be referenced as foundational techniques in vessel segmentation. These techniques fundamentally rely on the gradient of intensity along the long axis of the vessel being smaller than the intensity gradient orthogonal to the vessel where the vessel wall is crossed. Due to the label-free nature of HiP-CT and to the effects of partial blood infilling, large vessel collapse, and intensity gradients that can occur in tomographic reconstruction such as ring artefact removal⁸², such gradient-based approaches were unsuccessful when applied in an early investigation into HiP-CT vessel segmentation.

In this section, we delve into the latest deep-learning (DL) strategies used for blood vessel segmentation. We've grouped the DL-based techniques for segmenting vessels under three primary network structures: Convolutional Neural Networks, Generative Adversarial Networks, and Vision Transformers.

Convolutional neural network based models (CNNs)

Traditional CNNs have been foundational in vessel segmentation before the widespread adoption of architectures like U-Net. These neural networks, based on fundamental concepts of learning features and building hierarchical representations, leverage the presence of spatial structures and patterns inherent in image data. Unlike the Fully Convolutional Networks (FCN) that perform dense predictions, traditional CNNs often operate on patches or regions, aiming at classifying central pixels or aggregates.

Yao et al. used a 2D CNN architecture to extract blood vessels from fundus images⁸³. The output of their CNN architecture is the confidence level of each pixel being blood vessels. In 2018, Tetteh et al. developed an architecture called DeepVesselNet. Their design leverages 2-D orthogonal cross-hair filters using 3-D context information while minimizing computational demands. Additionally, they introduce a class balancing cross-entropy loss function with false positive rate correction specifically tailored to address the prevalent issues of significant class imbalances and high false positive rates observed with traditional loss functions⁸².

Cervantes-Sanchez et al. trained a multilayer perceptron with X-ray Coronary Angiography (XCA) images enhanced by using Gaussian filters in the spatial domain and Gabor filters in the frequency domain for segmentation of coronary arteries in X-ray angiograms⁸⁴. Nasr-Esfahani et al. presented a multi-stage model where a patch around each pixel is fed into a trained CNN to determine whether the pixel is of vessel or background regions⁸⁵. Samuel and Veeramalai proposed a two-stage vessel extraction framework to learn well-defined vessel features from global features (learned by the pre-trained VGG-16 base network) using the Vessel Specific Convolutional (VSC) blocks, Skip chain Convolutional (SC) layers, and feature map summations⁸⁶. In 2021, Iyer et al. developed a new CNN for angiographic segmentation: AngioNet, which combines an Angiographic Processing Network (APN) with Deeplabv3+ because of its ability to approximate more complex functions. The APN was tailored to tackle several challenges specific to angiographic segmentation, including low-contrast images and overlapping bony structures⁸⁷.

Fully convolutional neural network (FCN) based models: U-Net and its variants

Fully Convolutional Networks (FCNs) are a subtype of CNNs specialised in semantic segmentation tasks. Deep learning techniques have revolutionised medical image segmentation, achieving precise pixel-level classification with the help of FCN⁸⁹. FCNs are designed to produce an output of the same size as the input (i.e., a pixel-wise map). They replace the dense layers with convolutional layers. Expanding upon FCN's foundation, many researchers have pioneered advanced 2D fully convolutional neural networks, including U-Net, SegNet, DeepLab, and PSPNet^{88,90–92}. Khan et al. presented a fully convolutional network called RC-Net for retinal image segmentation. The network itself is relatively small and the number of filters per layer is optimised to reduce feature overlapping and complexity as compared to alternatives elsewhere in the literature. In the model, they kept pooling operations to a minimum and integrated skip connections into the network to preserve spatial information⁹³.

U-Net, distinguished by its unique U-shaped architecture, has become particularly popular for medical image segmentation tasks, especially when the dataset is limited⁸⁸. It leverages skip-layer connections combined with encoder-decoder multi-scale features, enhancing segmentation precision. An example of a U-Net architecture is illustrated in Fig. 2.

Given U-Net's success in medical image segmentation, Livne et al. streamlined it by halving the channels in each layer, coining it "half U-Net" for brain vessel segmentation⁴⁸. Even though there has been extensive exploration of 2D methodologies in the literature (see Tables 1, 2, 3 and 4), especially in coronary vessel segmentation (see Table 3), it is observed that relying solely on 2D segmentation networks may overlook inter-slice relationships, leading to subpar segmentation⁷⁸. Additionally, given the nonplanar nature of blood vessels, attempting to segment them without adequate volumetric information, such as on a single slice or in 2.5D scenarios, poses challenges. Although direct comparison of different studies is hampered due to the use of different datasets, the results from the studies detailed in see Tables 1, 2, 3 and 4 suggest that 3D models outperform 2D models. Specifically, 3D models achieve an average Dice score coefficient (DSC) of 0.83, compared to the 0.79 average for 2D models, underscoring the significance of interslice relationships.

Consequently, researchers have proposed 3D segmentation networks like 3D U-Net, V-Net, and VoxResNet^{94–96}. Huang et al. utilised 3D U-Net with data augmentation for liver vessel segmentation and introduced a weighted DSC loss function to address the voxel imbalance between vessels and other tissues⁹⁷.

To enhance the segmentation abilities of the U-Net network, several modifications have been proposed. One notable advancement is the attention U-Net, introduced by Oktay et al., which is designed to suppress irrelevant regions in an input image while highlighting salient features useful for a specific task⁹⁸. Bahdanau et al. originally conceived the attention mechanism to tackle the challenges arising from using a fixed-length encoding vector, which restricted the decoder's access to input information⁹⁹. This mechanism mimics human attention,

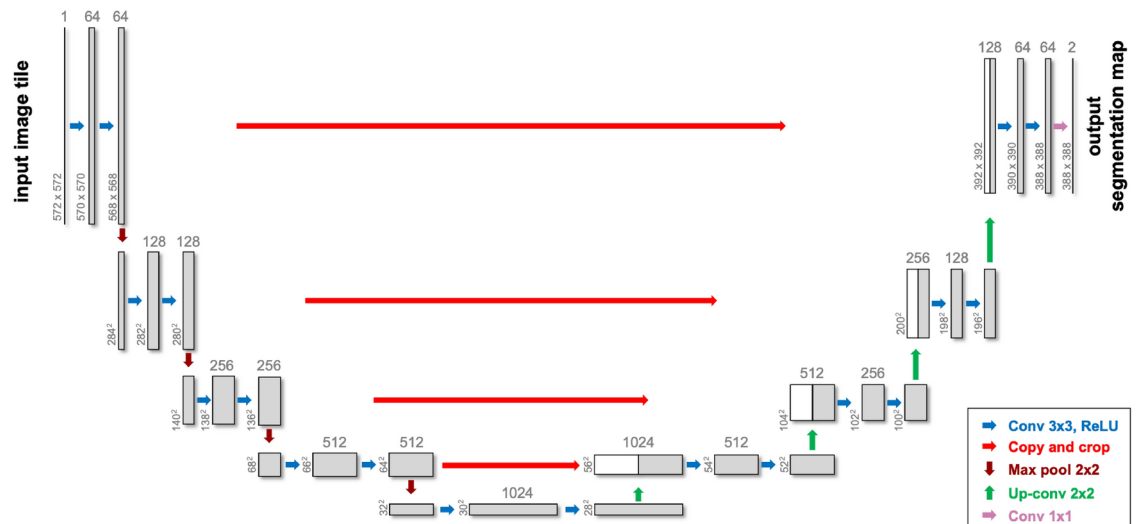


Fig. 2. A sample U-Net architecture used in medical image segmentation tasks-modified after⁸⁸.

enabling the model to concentrate on specific input areas while generating an output. The core idea behind this mechanism is to assign different weights to different parts of the input data, indicating how much “attention” each part should receive relative to others when producing a specific output.

Building upon the U-Net framework, Zhou and colleagues implemented U-Net ++ with the premise that the network would face a more straightforward learning challenge when the feature maps from both the decoder and encoder networks have semantic similarities¹⁰⁰. With this goal in mind, they reconfigured the skip pathways to diminish the semantic disparity between the encoder’s and decoder’s feature maps.

Sanches et al. melded 3D U-Net and Inception, dubbing it “Uception” for brain vessel segmentation¹⁰¹. Dong and colleagues introduced a cascaded residual attention U-Net, termed CRAUNet, for a layered analysis of retinal vessel segmentation. The architecture capitalizes on the advantages of U-Net, coupled with cascaded atrous convolutions and residual blocks that are further enhanced by squeeze-and-excitation features¹⁰². Drawing inspiration from U-Net and DropBlock, Guo et al. introduced the Structured Dropout U-net (SD-Unet) for coronary vessel segmentation. This design amalgamates the U-Net and DropBlock frameworks to omit specific semantic details, thereby preventing the network from overfitting¹⁰³. In 2021, Pan et al. adopted a 3D Dense-U-Net model and replaced the standard DSC loss function with the focal loss function to tackle the issue of class imbalance to achieve fully automated segmentation of the coronary artery¹⁰⁴. In 2022, Wu and colleagues introduced the Multi-Scale Interactive U-Net (MSI-U-Net), an enhancement of the 3D U-Net that enhances the precision of segmenting smaller vessels⁷⁸. They offered a strategy for interacting with information at multiple scales, where features are transferred among vessels of varying sizes through shared convolution kernel parameters, strengthening the relationship between small, medium, and large vessels in lung CT scans.

Moreover, to address the challenge of extensive annotation requirements in vessel segmentation, Koziński et al.¹⁰⁵ proposed a method that utilizes 2D tracing to reduce the annotation effort for 3D deep delineation of linear structures. Similarly, Dang et al.¹⁰⁶ introduced Vessel-CAPTCHA, an efficient learning framework designed to streamline vessel annotation and segmentation tasks.

nnU-Net¹⁰⁷ is one of the most widely utilised frameworks for vessel segmentation tasks, consistently delivering impressive results. It is considered the state-of-the-art for a wide variety of medical image segmentation tasks, with the original or variant of nnU-Net winning international biomedical challenges across 11 tasks^{107,108}. It has also recently been proposed as the network for the standardised evaluation of continual segmentation¹⁰⁸. Rather than introducing a novel model architecture, nnU-Net is designed to autonomously adapt and configure the entire segmentation framework, encompassing preprocessing, network architecture, training, and post-processing, to suit any new task. This deep learning framework is known for its flexibility, scalability, and ease of use in tackling various medical imaging segmentation problems. To this end, it is frequently used to benchmark new model approaches, or combined with other models as in ensembles to improve performance.

It should be noted that as Moccia et al.⁴ and Garcia et al.⁵³ recently mentioned, the success of the segmentation approaches is highly influenced not only by the algorithm but also by factors such as imaging modalities the presence/absence of noise or artifacts, and the anatomical region of interest. This makes direct comparisons among the studies in the literature challenging. Although different modalities, such as CTA and MRA, were used, the majority of the reviewed papers (71%) relied on private data, making it difficult to replicate or compare their findings.

The data presented in see Tables 1, 2, 3 and 4 reveals a predominant use of U-Net or variants of CNNs in the majority of studies, accounting for 80% of the cases.

Generative adversarial networks (GANs)

Goodfellow et al. introduced the Generative Adversarial Network (GAN) for synthesizing images from random noise¹⁰⁹. A sample GAN architecture can be seen from Fig. 3. The generator crafts artificial images to deceive the discriminator, while the discriminator endeavours to distinguish genuine images from fabricated ones, a process termed “adversarial training”. This methodology can be adapted to train segmentation networks, where a generator is tasked with generating segmented images, and the discriminator differentiates between the predicted segmentation maps and the authentic ones. This modification prompts the segmentation network to yield more anatomically accurate segmentation maps^{110,111}.

Son et al. introduced a technique employing generative adversarial training to create retinal vessel maps in the context of vascular segmentation. Their method enhanced the segmentation efficacy by employing binary cross-entropy loss during the training of the generator¹¹². In 2020, K.B. Park et al. introduced a novel architecture called M-GAN, which aimed to enhance the accuracy and precision of retinal blood vessel segmentation by combining the conditional GAN with deep residual blocks¹¹³. The M-generator incorporated two deep FCNs interconnected through short-term skip and long-term residual connections, supplemented by a multi-kernel pooling block. This setup ensured the scale-invariance of vessel features across the dual-stacked FCNs. They integrated a set of redesigned loss functions to optimise performance, encompassing BCE, LS, and FN losses. Recently, Amran et al. introduced an adversarial DL-based model for automatic cerebrovascular vessel segmentation¹¹⁴. Their BV-GAN model utilised attention techniques, allowing the generator to focus on voxels more likely to contain vessels. This is achieved by leveraging latent space features derived from a prior vessel segmentation map, effectively addressing the issue of imbalance settings. In related research, Subramaniam et al. introduced a 3D GAN-based approach for cerebrovascular segmentation¹¹⁵. While their method used GANs for dataset augmentation (generating an extensive set of examples with self-supervision), Amran et al. aimed to enhance the U-NET-based segmentation directly through the GAN technique. Finally, earlier in 2023, Xie et al. introduced the MLP-GAN for brain vessel segmentation¹¹⁶. This method divided a 3D brain vessel image into three separate 2D views (sagittal, coronal, and axial) and processed each through distinct 2D conditional GANs. Every 2D generator incorporated a modified skip connection pattern integrated with the MLP-Mixer block. This design enhances the ability to grasp global details.

Vision transformers

Transformers are primarily designed for sequence-to-sequence tasks but have shown significant promise in various domains, including computer vision. Vision Transformers (ViTs) are a type of neural network architectures that use transformer mechanisms to process images. Following their success in the realm of computer vision, researchers began to explore the potential of vision transformers for medical image segmentation tasks¹¹⁸. A sample vision transformer framework for vasculature segmentation is illustrated in Fig. 4. The input is a 3D medical image with dimensions H (height), W (width), D (depth), and C representing the number of channels (e.g., different modalities or scans). The image is divided into 3D patches, which are flattened and linearly projected into a lower-dimensional space before being passed into a Transformer encoder. The asterisk represents the extra [class] token, a learnable embedding added at the start of the patch sequence. This token interacts with all the patch embeddings within the transformer layers and eventually contributes to the final segmentation output through the decoder.

In 2021, Chen et al. came up with a model called TransU-Net, where a hybrid CNN-Transformer architecture is created to leverage both detailed high-resolution spatial information from CNN features and the global context encoded by Transformers by treating the image features as sequences¹¹⁹.

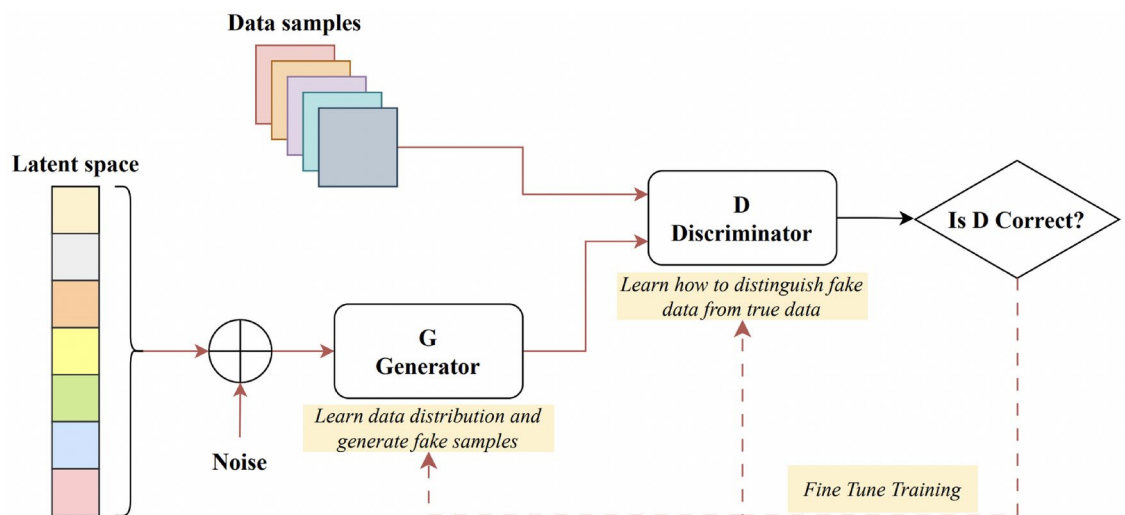


Fig. 3. A sample generative adversarial network architecture—modified after¹¹⁷.

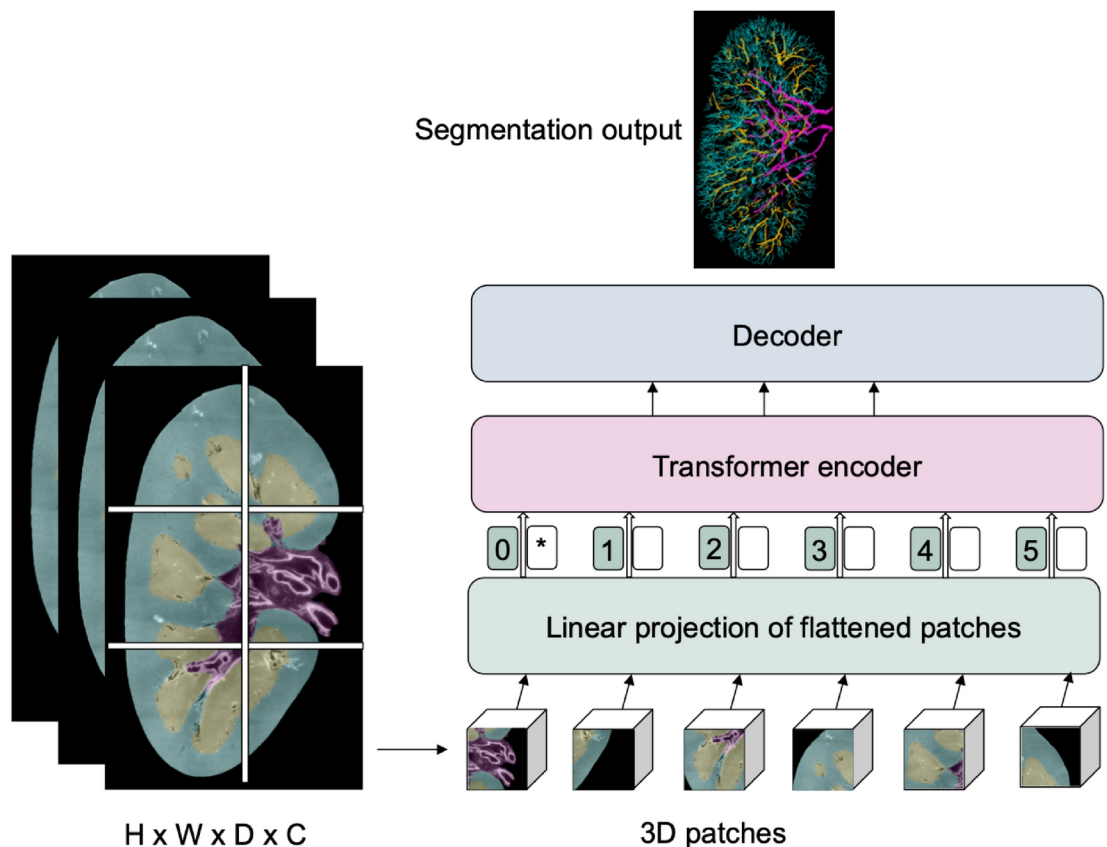


Fig. 4. A sample vision transformer architecture for renal vasculature segmentation.

Pan et al. introduced the Cross Transformer Network (CTN), a new approach designed to understand 3D vessel features while considering their overall structure. CTN accomplishes this by combining the U-Net and transformer modules, allowing the U-Net to be more globally aware and better handle issues like disconnected or imprecise segmentation¹²⁰.

Yu et al. introduced two new deep learning modules, called CAViT (Channel Attention Vision Transformer) and DAGC (Deep Adaptive Gamma Correction), to solve the problem of retinal vessel segmentation. CAViT combines two components: efficient channel attention (ECA) and the vision transformer (ViT). The ECA module examines how different parts of the image relate to each other, while the ViT identifies important edges and structures in the entire image. On the other hand, the DAGC module figures out the best gamma correction value for each input image. It does this by training a CNN model together with the segmentation network, ensuring that all retinal images have the same brightness and contrast settings¹²¹.

Zhang and his team, on the other hand, introduced a model named TiM-Net for efficient retinal vessel segmentation¹²². To capitalize on multiscale data, TiM-Net takes in multiscale images after maximum pooling as its inputs. Following this, they incorporated a dual-attention mechanism after the encoder to minimize the effects of noisy features. Concurrently, they utilised the MSA mechanism from the Transformer module for feature re-coding to grasp the extensive relationships within the fundus images. Finally, a weighted SideOut layer was created to complete the final segmentation.

The analysis of Tables 1, 2, 3 and 4 indicates a less frequent application of GANs and transformers (around 20%), suggesting that while U-Net and CNNs are the preferred methods in the field, there is still room for exploration and potential growth in the utilisation of GANs and transformers for vessel segmentation.

Evaluation metrics

In order to evaluate any automated vessel segmentation model or algorithm, a quantitative assessment of a segmentation task can be done through various performance metrics. Depending on the theory behind them, these metrics can be divided into overlap-based, volume-based, pair-counting-based, information-theoretic-based, probabilistic-based, and spatial distance-based measures³⁹. Three commonly used metrics and their calculations are illustrated in Fig. 5. Note that the volumetric and surface DSC are identical, but applied to volume and surface area, respectively.

When assessing the performance of segmentation algorithms in comparison to ground truth, a contingency table is often employed, featuring True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP) values. In this context, positive and negative denote pixels that, according to the ground truth segmentation, are attributed to vessels and background respectively.

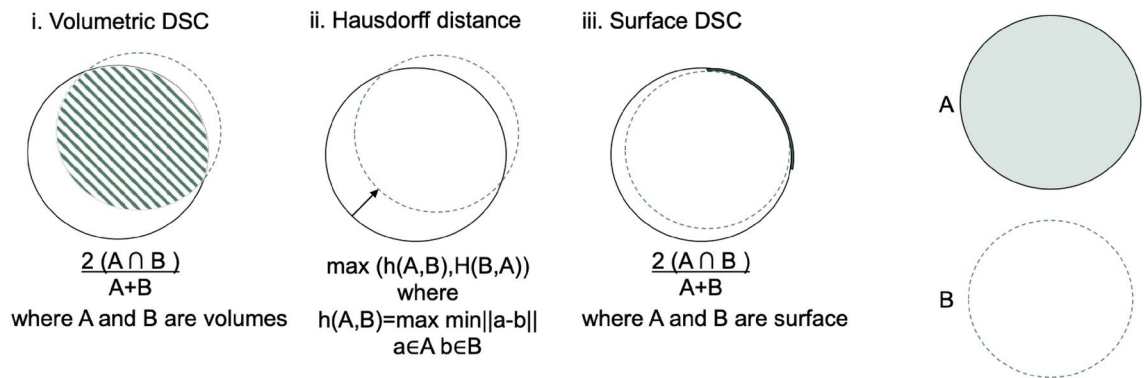


Fig. 5. Illustration for calculation of volumetric Dice score coefficient (DSC), Hausdorff distance, and surface DSC. The solid line represents the ground truth contour, whereas the dashed line is the prediction. i. A Volumetric DSC, defined as the union of two volumes (green volume region) normalised by the mean of the two volumes. ii. Hausdorff distance, defined as the maximum nearest neighbor Euclidean distance (the arrow). iii. Surface DSC, is defined as the union of two contours (yellow contour region) normalised by the mean surface of the two contours.—modified after¹²⁹.

A review of the literature reveals varied metric reporting across studies (as evidenced in Tables 1, 2, 3 and 4). This section delves into the most frequently used performance metrics, elucidating their underlying theory, strengths, and weaknesses.

Overlap-Based Measures Overlap-based measures are used to evaluate the similarity between the segmented result and the ground truth based on their overlapping regions. Such measures are the most widely utilised metrics with all studies we reviewed reporting at least one overlap based metric. However they have some limitations when applied to vascular segmentation which will be discussed further. The DSC is one of the most common overlap-based metrics that calculates the spatial overlap of two segmentations, with 26 of the 31 studies reviewed utilised DSC (see Tables 1, 2, 3 and 4). Given its properties, DSC is particularly used in vascular segmentation where the vasculature accounts for a small fraction (e.g. around 3% in the brain) of the organ volume, leading to unbalanced data challenges¹³⁰.

Sensitivity, also known as TP Rate or Recall, is another overlap-based metric. It measures how well a model identifies TPs (e.g., correctly segmented vessels). It provides insight into how much of the ground truth overlaps with the predicted positive segment. Specificity, on the other hand, measures the proportion of TNs (e.g., correctly identified non-vessel areas) that are correctly detected. It reflects on how much of the ground truth's negative area overlaps with the predicted negative segment. The calculations for these two measures, along with other metrics, are provided in Table 5.

Another related measure is precision, often referred to as the positive predictive value (PPV). While not frequently utilised in medical image validation, it plays a role in determining the F-Measure. The calculation for precision is presented in Table 5, while the F-Measure, also known as the F1 score, is the harmonic mean of precision and recall.

In 2020, Shit et al. developed centerline DSC (*clDice*), a unique topology-preserving loss function for tubular structure segmentation, as another DSC modification¹²⁵. Their work highlighted that DSC metric does not necessarily assess the connectedness of a segmentation because DSC does not equally weight tubular structures with large, medium, and small radii. As connectivity is a crucial biological feature of vascular networks, bespoke metrics should be used to evaluate vascular segmentations. The authors demonstrated that training on a globally averaged loss causes a considerable bias towards the volumetric segmentation of big arteries in real vascular datasets compared to arterioles and capillaries (radius ranges e.g., 30 μm for arterioles and 5 μm for capillaries). To enable topology preservation, centerline DSC (*clDice*) is formulated using topological precision (T_{prec}) and topological sensitivity (T_{sens}). The formula for *clDice* is provided in Table 5, while T_{prec} and T_{sens} are calculated as follows:

$$T_{\text{prec}} = \frac{|S_P \cap V_L|}{|S_P|}, \quad (1)$$

$$T_{\text{sens}} = \frac{|S_L \cap V_P|}{|S_L|}, \quad (2)$$

here V_L is the ground truth mask, V_P is the predicted segmentation mask, and S_P (Predicted Skeleton) and S_L (Labeled Skeleton) represent the skeletons extracted from the predicted and ground truth segmentations, respectively.

Volume-based measures These metrics focus on the volume or size of the segmented structures. As the name implies, Relative Volume Difference (RVD) is a typical volume-based metric that is based on the absolute difference in volumes of the segmented structure and the ground truth, usually normalised by the ground

Metric category	Metric	Formula	Strengths	Limitations
Overlap-based	Dice Similarity Coefficient (DSC)	$DSC = \frac{2TP}{FP+FN+2TP}$ ¹²³	Easy to implement - widely used (26/31 studies). Good for overlap assessment.	Sensitive to class imbalance and lacks connectivity assessment.
	Sensitivity (Recall or TPR)	$Sensitivity = \frac{TP}{TP+FN}$ ¹²⁴	Measures true positive rates.	Ignores structural/topological features.
	Specificity (TNR)	$Specificity = \frac{TN}{FP+TN}$ ¹²⁴	Measures true negative rates.	Ignores structural/topological features.
	Precision	$Precision = \frac{TP}{TP+FP}$ ¹²⁴	Accounts for false positives/negatives.	Rarely used in medical image segmentation.
	F1-Score	$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$ ¹²⁴	Balances precision and recall.	Rarely used in medical image segmentation.
Topology-preserving	Centerline Dice (<i>clDice</i>)	$clDice = \frac{2T_{prec}T_{sens}}{T_{prec}+T_{sens}}$ ¹²⁵	Preserves connectivity of thin, tubular structures like vessels.	Computationally expensive - less commonly used (2/31 studies).
Volume-based	Relative Volume Difference (RVD)	$RVD = \frac{ V_p - V_g }{V_g}$ ¹²³	Quantifies the disparity in volume.	Sensitive to small structures
Spatial distance-based	95th Percentile Hausdorff Distance (95HD)	$percentile(h(A, B), 95)$ ¹²⁶	Captures boundary accuracy.	Sensitive to outliers or small discrepancies.
	Average Symmetric Surface Distance (ASSD)	$ASSD = \frac{1}{2} \left(\frac{1}{ A } \sum_{a \in A} \min_{b \in B} d(a, b) + \frac{1}{ B } \sum_{b \in B} \min_{a \in A} d(b, a) \right)$ ¹²⁷	Provides average boundary accuracy.	Less sensitive to topology, focuses on boundary alignment.
Probabilistic-based	Area Under the Curve (AUC)	$AUC = \int_0^1 TPR(FPR) d(FPR)$ ¹²⁸	Common in binary classification models, reflects overall performance.	Rarely used (3/31 studies).
Pair-counting-based	Adjusted Rand Index (ARI)	$ARI = \frac{RI - \text{Expected } RI}{\text{Max } RI - \text{Expected } RI}$ ¹²⁸	Adjusts for chance in clustering assessments.	Not widely used in vascular segmentation.
Information-theoretic	Mutual Information (MI)	$MI(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right)$ ¹²⁸	Captures shared info between segmented and true data.	Complex to compute and rarely used in vascular segmentation.

Table 5. Summary of Evaluation Metrics for Vascular Segmentation: TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative, V_p = Predicted Volume, V_g = Ground Truth Volume, A, B = Sets of Points for Spatial Distance, T_{prec} = Topological Precision, T_{sens} = Topological Sensitivity, RI = Rand Index, x = Predicted Segmentation Value, y = Ground Truth Value, $p(x, y)$ = Joint Probability Distribution of x and y , $p(x)$ = Marginal Probability Distribution of x , $p(y)$ = Marginal Probability Distribution of y , TPR = True Positive Rate, FPR = False Positive Rate, $d(FPR)$ = Differential of the False Positive Rate.

truth volume. It should be noted that RVD does not account for the spatial agreement between segmentations, as metrics like Dice would. For example, two objects of the same volume may receive a perfect score even if their shapes differ significantly, such as a square and a tree. Therefore, RVD fails to capture spatial correctness and should be complemented with metrics that assess spatial overlap and shape similarity, such as the Dice coefficient, to provide a more comprehensive evaluation of segmentation performance.

Spatial distance-based measures

Spatial distance-based measures focus on the spatial discrepancies between the segmented structures and the ground truth. For instance, the 95th Percentile Hausdorff Distance (95HD) assesses the greatest distance between a point in the true segmentation and the nearest point in the segmented result (Table 5). More formally, if we define $h(A, B)$ as the set of all distances from points in set A to their nearest points in set B , then the 95HD is calculated as the 95th percentile of this set of distances. This measures the largest error in spatial alignment between the two sets, but by focusing on the 95th percentile, it avoids sensitivity to extreme outliers.

The Average Hausdorff Distance (AHD) is another commonly used metric in image segmentation and shape analysis to measure the similarity between two surfaces, typically the predicted segmentation and the ground truth (Table 5). It is derived from the Hausdorff Distance (HD), which captures the maximum distance between

any point on one surface to the closest point on the other surface. The AHD refines this concept by averaging the distances between all points on the surfaces, making it a more robust measure than the standard Hausdorff Distance, which only considers the single greatest distance. Smaller AHD values indicate that the surfaces are more similar, with lower distances between corresponding points, while larger values reflect greater deviations in the segmentation's boundaries.

Boundary-based techniques are another important subcategory of distance-based measures¹³¹. Normalised surface DSC, mean average surface distance, and average symmetric surface distance (ASSD) are three key metrics, especially in the field of vessel segmentation. Surface DSC is a measurement to evaluate the similarity between the segmented surface and a ground truth surface. Normalised surface DSC takes this a step further by normalising the measurement and making it less sensitive to size differences between the predicted segmentation and the ground truth. This is particularly useful for segmenting smaller structures, or size variability between subjects.

Surface distance measures the shortest distance of each point on the surface of the segmented structure to the surface of the ground truth. The ASSD is the average of these distances, calculated symmetrically between the two surfaces (Table 5). Specifically, the distance from each point on surface *A* to the nearest point on surface *B* is averaged, and the same is done for points on *B* with respect to *A*, ensuring a balanced comparison between the surfaces.

The use of the appropriate assessment metric is crucial, as each measurement has specific biases dependent on the characteristics of the segmented structures. Whilst our review of literature shows that DCS or other generalised overlap measures are still the most frequently used metrics, more specialised vascular specific metrics are available and are becoming more widely implemented. The evaluation metric can strongly influence the choice of the final optimal model and hence should be chosen in accordance with the segmentation task's nature and taking into consideration what downstream analyses are to be conducted on the segmentations, or the biological implications of a particular metric. For vascular structures, quantitative analyses of network topology for flow simulations are common end goals for segmentation^{3,21}. In such cases meshing or skeletonising are common steps after segmentation. Meshing or skeletonisation algorithms are often highly sensitive to holes or cavities in the segmentations, or to breaks in connectivity in vascular components¹³². Thus, evaluation methods highly sensitive for connectivity breaks (e.g. cDice) should be implemented in these cases.

Probabilistic-Based Measures

Probabilistic-based measures evaluate the performance of segmentation and classification algorithms based on predicted probabilities rather than strict binary or discrete decisions. In binary classification, for example, a probabilistic model may assign a probability indicating the likelihood that an item belongs to the positive class rather than simply categorizing it as positive or negative.

The ROC curve (Receiver Operating Characteristic) stands as a prime example of a probabilistic-based measure. It graphs the sensitivity against (1-Specificity = 1-TNR = FPR) over a spectrum of decision thresholds. This curve offers a holistic perspective on a model's capability to distinguish between classes.

The Area Under the Curve (AUC) encapsulates the model's overall discriminative ability between positive and negative classes. From a mathematical standpoint, the AUC represents the integral of the ROC curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR).

This reflects the model's ability to distinguish between classes at various threshold settings. AUC appears in 3 of the 31 models reviewed in Tables 1, 2, 3 and 4.

Details on the pair-counting-based and information-theoretic-based metrics can be found in the Supplementary Material, as these metrics are not commonly employed in vessel segmentation.

nnU-Net for kidney vessel segmentation

From the above review of the literature, we concluded that segmentation of blood vessels from HiP-CT data should be initially attempted using nnU-Net¹⁰⁷. As nnU-Net is one of the most widely utilised frameworks for vessel segmentation tasks and has consistently delivered impressive results, we felt that this would provide the initial baseline against which any future development of more HiP-CT-specific frameworks should be benchmarked. Therefore, we prepared a training dataset from three different human kidneys, imaged with HiP-CT and semi-manually segmented by expert annotators. The goal was to assess how much training data is needed to provide adequate segmentation for a single kidney but also to investigate if nnU-Net trained on a subset of kidney data would generalise to a new kidney dataset given the high inter-sample variability seen between human organs.

In this section, we will provide a description of the HiP-CT kidney datasets, including its acquisition protocol and the segmentation process ("Hierarchical phase-contrast tomography (HiP-CT) kidney dataset" Section). Following that, we will present detailed information about the employed nnU-Net framework configuration ("nnU-Net framework" Section), and finally, we will present the results for nnU-Net application to kidney vascular segmentation of HiP-CT data "Evaluation" Section.

Hierarchical phase-contrast tomography (HiP-CT) kidney dataset

Three human kidneys were used to create the training dataset - termed kidney 1, kidney 2, and kidney 3. Kidney 1 and kidney 3, were collected from donors who consented to body donation to the Laboratoire d'Anatomie des Alpes Françaises before death. Kidney 2 was obtained after a clinical autopsy at the Hannover Institute of Pathology at Medizinische Hochschule, Hannover (Ethics vote no. 9621 BO K 2021). The transport and imaging protocols were approved by the Health Research Authority and Integrated Research Application System (HRA and IRAS) (200429) and the French Health Ministry. Post-mortem study was conducted according to Quality Appraisal for Cadaveric Studies scale recommendations¹³³. Sample preparation, scanning and tomographic

Donor identifier	Experiment name	Dataset size (x, y, z) (pixels)	Gender	Age	Scanning voxel size (um)	Scan energy (keV)	Binned voxel size (um)
LADAF-2021-17	Kidney 1	1303, 912, 2279	M	63	25.0	81	50.0
S20-28	Kidney 2	1041, 1511, 2217	M	84	25.0	88	50.0
LADAF-2020-27	Kidney 3	1706, 1510, 501	F	94	25.08	93	50.16

Table 6. The size of the kidney volumes from the dataset used in the experiments together with their gender and resolution information.

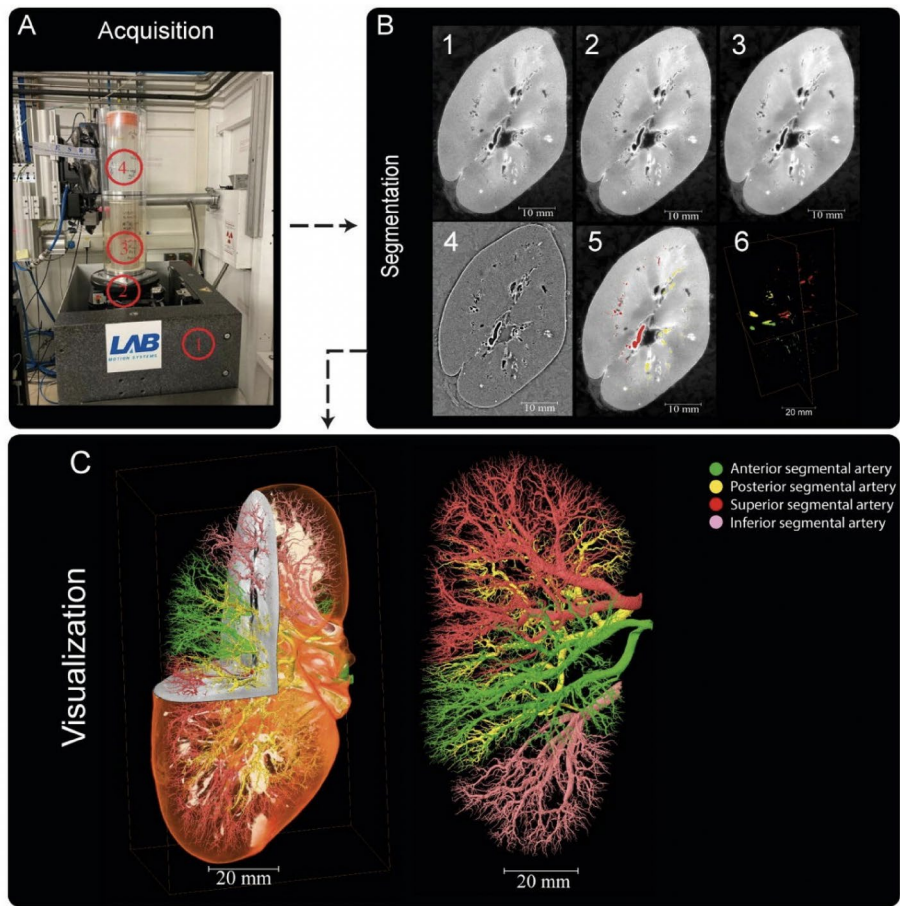


Fig. 6. Processing pipeline for HiP-CT imaging and segmentation of human kidney vasculature. (A) Setup for imaging acquisition using HiP-CT at BM05; (1) tomographic stage, (2) platform, (3) sample, (4) reference sample (B) Image processing pipeline; (1) a 2D reconstructed image at 25 $\mu\text{m}^3/\text{voxel}$ resolution; (2) binning the image by (2, 3) applying 3D median filter to increase signal-to-noise ratio, (4) Image normalisation using background detection correction, (5) Segmentation and thresholding, (6) Labelling the four main arterial branches (C) 3D rendering of the segmented vascular network of a human kidney. Each of the main four branching of the renal artery entering the kidney are colour-coded, Figure after²¹.

reconstruction protocols are described in the references^{16,21,82,134}. Basic scan parameters and demographic information are provided in Table 6

The annotation process of the three kidneys was carried out in Amira Version 2021.1. The reconstructed raw image data Fig. 6 B1 first underwent average binning x 2 Fig. 6 B2 from the acquired resolution (ca. 25 μm) to ca. 50 μm , 3D median filtering was applied in Amira-Avizo v2021.1 (3 iterations 26 voxel neighbourhood) Fig. 6 B3 and filtering to visually enhance vessels appearance performed using background detection correction (Amira v2021.1; default parameter settings) Fig. 6B4. The annotation was conducted in a semi-manual fashion using the Amira v2021.1 magic wand tool. This is an interactive 3D region growing tool. Using this tool, annotators select a seed voxel within a vessel in a slice (slices can be in any one of three orthogonal directions); in addition, the annotator selects and refines a combination of intensity threshold, contrast threshold and hard-drawn limits, which are used to specify the stopping criteria of the 3D region growing. In some cases, it is necessary to manually draw using a paintbrush tool in slice-by-slice locations where vessels are infilled with blood or have largely collapsed. To provide assurance for labeling quality, an expert annotation validation process was implemented.

3D annotations were performed by a first expert annotator, after they were complete a second independent annotator conducted a 3D proofreading of these binary labels adding additional missing vessel labels again using three orthogonal planes for segmentation. After this second annotation round, five randomly selected 2D circular regions of the image were selected, and a third annotator marked all the labeled regions (one mark per region, which are the cross sections of the vessels). The third annotator then counted the True Positive (correctly segmented vessels), and False Negative (missed vessels), (note false positives are rare and easily removed due to the connected nature of the arterial tree). This provides a metric for what proportion of vessels have been traced or missed, and by sampling random areas in the dataset, directs the first two annotators to regions that may be poorly segmented. This triple validation process was iterated to achieve final annotations which can be accessed via¹³⁵. It should be noted that vessels with diameters as small as 1-2 pixels could be segmented with the binning, median filter, and manual approach. Given the resolution of the images and their location within the generation vascular tree, these vessels represent interlobular arteries within the renal vascular tree and have been estimated from higher resolution HiP-CT of the same kidney, to be two branching generations from the kidney glomeruli (where the capillary bed is found). It is also of note that these samples come from a diverse age grouping 63-94yr old and have two males and one female represented. Finally, it is also worthy of note that in the case of kidney 2 (S20-28), the donor's cause of death was COVID-19 which can cause microthrombi in the renal vasculature, and is seen as hyperintensities, particularly around the edge of the kidney context, this can be seen in Supplementary movies (1-3) which show slice by slice views of each kidney.

The number of slices and size of each slice segmentation were as followed (Z, X, Y): kidney 1: (2279, 1303, 912); kidney 2: (2217, 1041, 1511); and kidney 3: (501, 1706, 1510). The network topologies automatically generated for 3D full resolution configuration for experiment 1 were a patch size (Z, X, Y) of [192, 112, 112], a batch size of 2 and the number of pool per axis (Z, X, Y) were [5, 4, 4].

The imaging datasets used in this study are publicly available through the Human Organ Atlas data portal, where they can be accessed and visualised at <https://human-organ-atlas.esrf.eu>. Additionally, a Kidney Vessel Segmentation Challenge has been organised to facilitate the development of new deep learning methods for 3D blood vessel segmentation using HiP-CT data, with details available on Kaggle at <https://www.kaggle.com/competitions/blood-vessel-segmentation/>.

nnU-Net framework

Preprocessing. Image preprocessing is a necessary step in machine learning, especially for image segmentation tasks. Notably, this step is crucial for enhancing model performance, generalization, and robustness when handling real-world data. nnU-Net incorporates a comprehensive set of automated preprocessing steps within its data preparation pipeline. These steps include:

- *Data Augmentation:* Rotations, scaling, Gaussian noise, Gaussian blur, brightness and contrast adjustments, simulation of low resolution, gamma correction, and mirroring.
- *Intensity Normalization:* Global dataset percentile clipping, z-score normalization with a global foreground mean, and z-score normalization on a per-image basis.
- *Resampling Strategies:* In-plane resampling using third-order spline interpolation and out-of-plane resampling with either nearest neighbor or third-order spline interpolation.
- *Image Target Spacing:* Determined based on the lowest resolution axis tenth percentile, axes median, and median spacing for each axis. During preprocessing, nnU-Net automatically extracts a dataset fingerprint, which encompasses dataset-specific properties such as image sizes, voxel spacings, and intensity information. It also computes relevant dataset statistics and configures the training parameters based on both the dataset characteristics and the available computational resources. As a result, parameters like batch size and other configuration settings are automatically determined by nnU-Net and used directly in the experiments. For a detailed overview of the automated method configuration, refer to Supplementary Fig. 1¹⁰⁷.

Architecture. It is important to mention that the nnU-Net framework does not have a new deep learning architecture. However, it covers the U-Net family, which has encoder-decoder architectures, including 2D U-Net⁸⁸, 3D U-Net⁹⁵, and Cascaded 3D U-Net¹³⁶. It also provides an ensemble option, which explores 2D U-Net, 3D U-Net, or 3D cascade results and chooses the best model (or combination of two) according to cross-validation performance.

Postprocessing. The nnU-Net framework also provides optional configuration for postprocessing on the full set of training data and annotations, including treating all foreground classes as one individual class (depending on the largest component suppression increases in cross-validation performance)

Configuration and Training Process. We trained the nnU-Net framework using the 3D U-Net version with 3-fold cross-validation implemented on the training sets. The evaluation results presented in the following section are the averages on testing sets of 3-fold cross-validation. We used the nnU-Net default auto-generated hyper-parameters for training, which include the learning rate⁹¹ initialised as 0.01 with a polynomial decay policy of $(1 - \text{epoch}/\text{epoch}_{\text{max}})^{0.9}$, the loss function as the sum of cross-entropy and DSC loss¹³⁷, an optimizer based on stochastic gradient descent (SGD) with Nesterov momentum value of 0.99 and epoch number of 1000 with mini-batches of 250. Notably, the default nnU-Net typically requires > 30 million parameters for the training. The proposed model was implemented in Python language¹ using Pytorch¹³⁸. All experiments were on an NVIDIA TITAN RTX 24GB GPU. The training processes for Experiments 1, 2, and 3 are shown in Supplementary Figs. 2, 3 and 4, respectively.

¹<https://www.python.org>.

Evaluation

We chose to perform four experiments. In three experiments, we used two of the kidneys as training data and one as test data. These aimed to investigate how well the nnU-Net approach would generalise to a new dataset where the donor and thus anatomy and sample prep may differ slightly; it also had the virtue of testing how different-sized training datasets impacted nnU-Net training output. The overview of the experiments can be seen in Fig. 7.

The fourth experiment aimed to see how well nnU-Net could segment the remainder of a dataset given a limited amount of labeled slices from the same dataset. Due to the size of HiP-CT datasets, such a model still has utility as segmenting vascular networks for entire organs is highly time-consuming²¹. In this section, we selected kidney 1 for our experimental evaluation. Half of the kidney has been utilised as a training set (1139 slices), and the remaining half as a testing set (1140 slices). The kidney is divided into two symmetrical halves horizontally (when the kidney figures shown in the manuscript are considered), ensuring that both sides contain a similar distribution of large and small vessels. Please note that the purpose of this experiment is to determine whether the annotation process can be accelerated by annotating only half of the kidney. As a result, it inherently differs from the other experiments and does involve some degree of data leakage.

The importance of evaluation metrics in automatic image processing with machine learning cannot be emphasised enough, as they serve as a basis for determining the choice or practical applicability of a method. However, much of the research has been on creating novel image processing algorithms leaving the critical issue of reliably and objectively evaluating the performance of these algorithms largely unexplored¹³¹. Moreover, some of the commonly used evaluation metrics do not always correlate with clinical applicability^{129,139}, and the specific features of a biomedical problem may make certain metrics unsuitable, such as when the DSC is utilised to evaluate extremely small structures¹⁴⁰. As a result, carefully choosing the right evaluation metric for a given problem becomes important for validating and comparing the performance of image processing methods.

In vessel segmentation, class imbalance presents a significant challenge where the pixel count for the foreground class (vasculature) is notably lesser than that of the background class (non-vasculature).

Class imbalance can hinder effective network training, as most data points are typically negative samples (or backgrounds) that often do not offer valuable learning insights. Moreover, these negative samples can dominate the positive ones (such as arteries) during the training process as loss values are predominantly generated from the negative samples.

A combination of overlap-based and boundary-based metrics is selected to evaluate different properties of the model predictions¹³¹. For overlap-based metrics, DSC is computed and is complemented by the Centerline DSC (cDice)¹²⁵ due to the tubular nature of blood vessels and the importance of connectivity for vascular structures. Overlap-based metrics have certain limitations such as shape unawareness and inaccurate assessment when dealing with small structures hence, boundary-based metrics, specifically Normalised Surface DSC (NSD) and Average Symmetric Surface Distance (ASSD), are also computed using MONAI v1.2.0¹⁴¹. The results of the four experiments are provided in Table 7 and representative images for each kidney shown in Fig. 8.

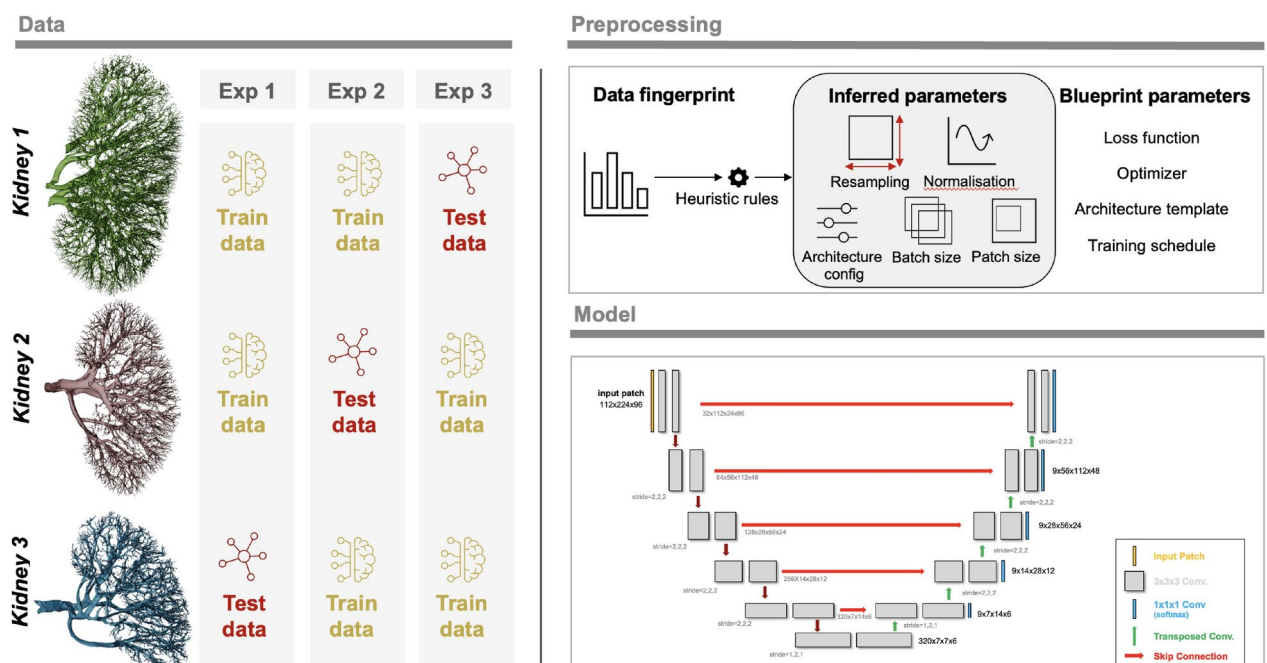


Fig. 7. Overview of the experiments together with a sample 3D nnU-Net architecture.

Expt.	Train data	Test data	DSC	CLD	NSD ($t = 1$)	NSD ($t = 0$)	ASSD
1	Kidney 1,2	Kidney 3	0.9410	0.8886	0.9651	0.7631	4.300
2	Kidney 1,3	Kidney 2	0.9523	0.8533	0.9518	0.7120	0.8639
3	Kidney 2,3	Kidney 1	0.8585	0.8228	0.8968	0.7132	2.9270
4	Half of kidney 1	Other half of kidney 1	0.9513	0.8631	0.9404	0.8549	2.1561

Table 7. Results of all four experiments done using nnU-Net. *SD* surface distance, *NSD* normalised surface DSC, *CLD* centerline DSC, *t* tolerance (in voxels), *ASSD* average symmetric surface distance

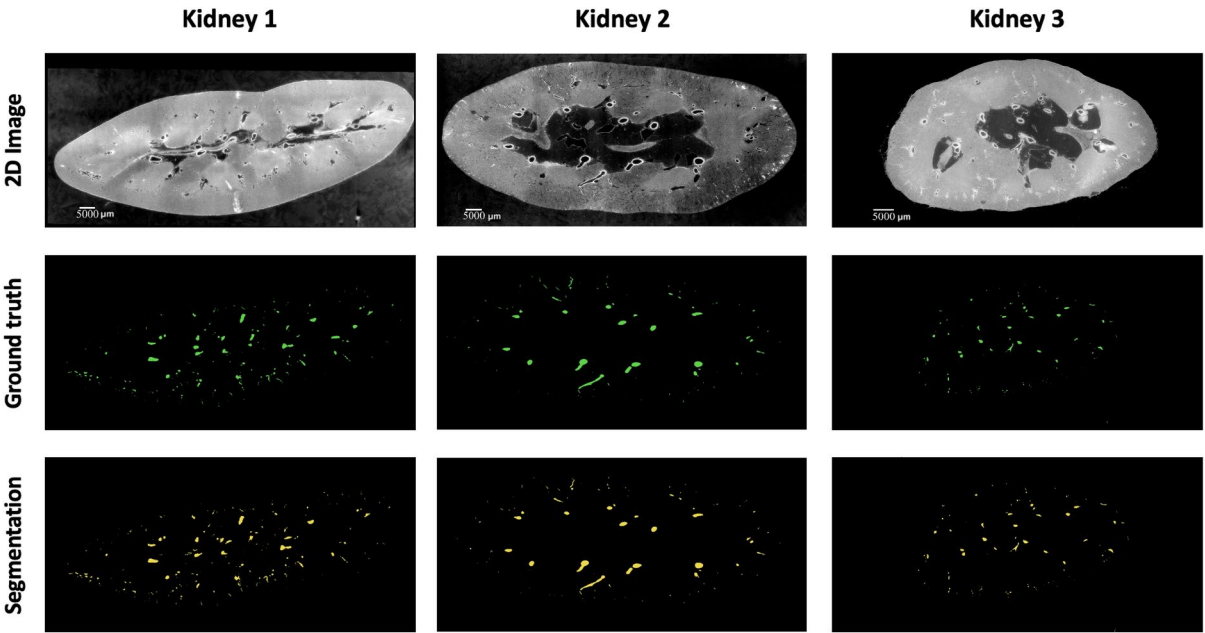


Fig. 8. A sample 2D slice from each kidney together with their corresponding ground truth labels and the output of the model segmentations (for kidney 1 the segmentation is provided by Expt. 3).

Results

In this study, we evaluated model performance for the segmentation of the arterial network from HiP-CT images of whole human kidney. Five metrics: DSC, cDice, NSD ($t=1$), NSD ($t=0$), and ASSD, with results presented in Table 7 were used to evaluate the outputs. The range of metrics we employed in our experiments provides a thorough overview of model performance and allows the state-of-the-art presented to be rigorously benchmarked against future research. For each experiment, one unseen kidney is selected as test data for evaluation of the model's generalisation capability. The 5 metrics are calculated on test data after the model training is finished. Beyond quantitative evaluations, we provided visual insights through 3D representations of the ground truth, model predictions, and false negatives.

Based on our literature review (see Tables 1, 2, 3 and 4), it is clear that the DSC, also known as Dice, is the predominant metric for evaluating vascular segmentation model performance. Therefore, employing DSC for result evaluation is essential to align with existing literature, such as in⁵⁰. This metric quantifies the overlap between segmentation predictions and ground truth. In our three experiments, the first and second experiments yielded superior segmentation results on the unseen kidney, with DSC of 0.9523 and 0.9410, respectively, whereas the third experiment attained only 0.8585. 3D examination of kidney 1, as shown in Fig. 9; revealed more collapsed vessels compared to kidney 2 and 3, potentially explaining the lower DSC when the first kidney was the test subject. Nevertheless, DSC primarily assesses voxel-to-voxel concordance, overlooking several crucial characteristics of the vessels. Hence, it should not be the sole metric for deciding the performance of vascular segmentation.

In “Evaluation” Section, we describe an additional experiment (experiment four) to evaluate the nnU-Net's performance with a limited dataset. In this case we used just kidney 1. We trained the model on half the data and tested it on the remaining half. This experiment's training-test division was designed to test the model's capabilities, to extrapolate from a small subset of images within a single organ to the entire dataset. This reduces the challenge of the segmentation task but also reduces the utility of the final model, however given the size of HiP-CT datasets, such an approach is likely to still have utility in particularly challenging or anatomically unique datasets. The performance metrics for the test set were DSC = 0.9513, cDice = 0.8631, NSD ($t=1$) = 0.9404, NSD ($t=0$) = 0.8549, and ASSD = 2.1561. These results indicate that nnU-Net is effective at segmenting 'unseen'

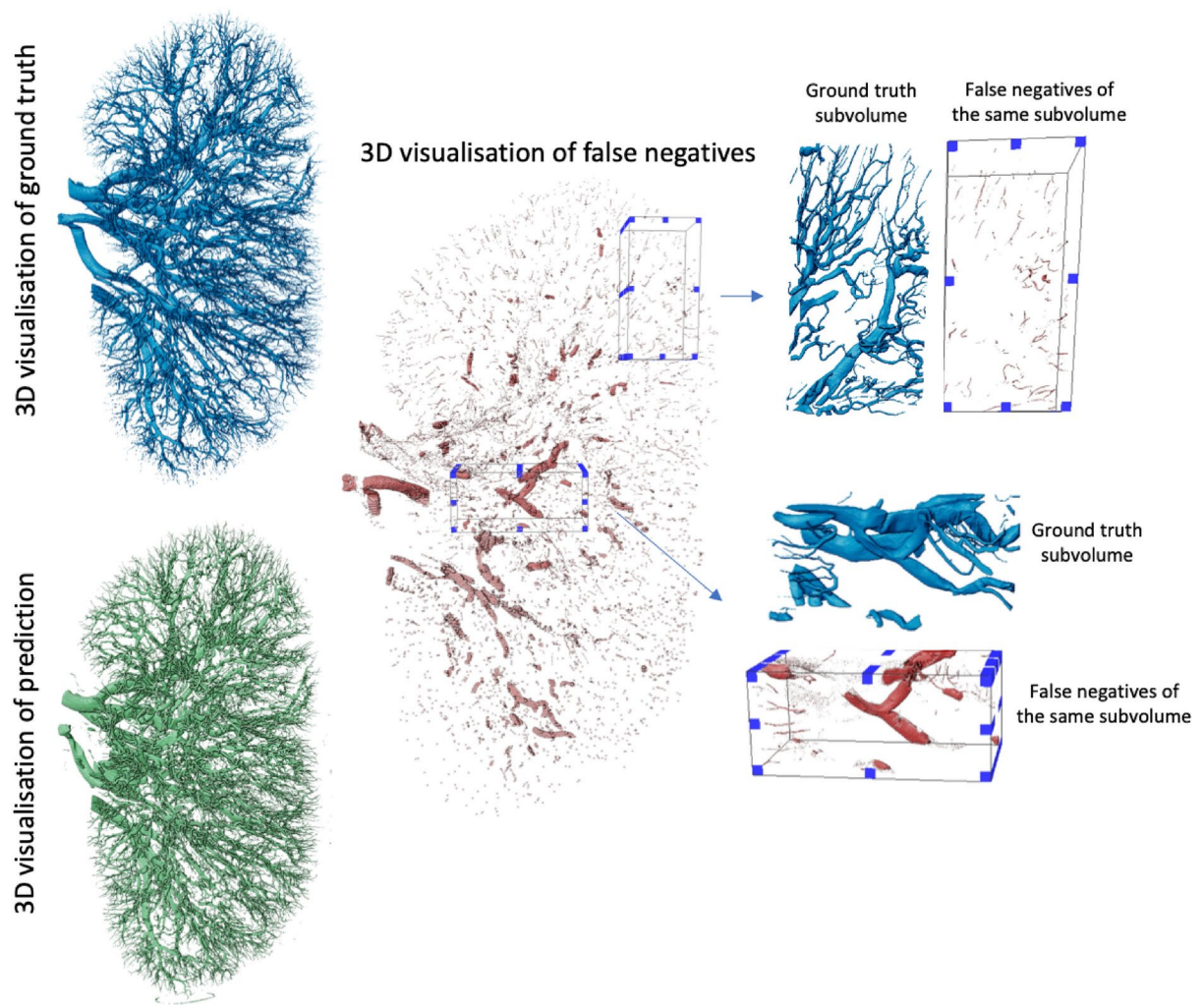


Fig. 9. The visualisation of kidney 1 predictions from Experiment 3 showing the model's false negative prediction together with subvolumes taken from it indicating the regions where model did perform worst.

kidney tissues from the same organ, reflecting its capability to learn kidney anatomy and thereby enhancing the efficiency of the labeling process. It also shows the expected improvement by comparison to the other three experiments, indicating the challenges with accounting for anatomical variation across human samples, or variations in the data acquisition and data reconstruction parameters see^{16,134,142} for details on protocol variations.

Discussion

The nnU-Net framework demonstrated significant success in accurately segmenting kidney vessels. The results provide key insights into how nnU-Net can address the challenges posed by hierarchical, connected vascular networks, making it highly relevant for applications in high-resolution imaging, such as HiP-CT. Given the nature of vasculature networks - hierarchical connected networks of thin and elongated tubular structures, ensuring sensitivity to small as well as large vessels, mitigating boundary ambiguities and ensuring connectivity are all important factors. Therefore, we reported a number of different metrics, and provided qualitative outputs to understand what features each metric best represents. Distance metrics applied included the ASSD (see equation 15) and the NSD to evaluate the similarity of the segmented regions' surfaces (or boundaries) rather than the volumetric overlap. It is particularly useful in scenarios where small boundary deviations can be critical, such as in segmenting thin structures or lesions. Tolerance (t) is an important hyper-parameter when it comes to NSD, which relates to the acceptable amount of deviation in the segmentation boundary in voxels. If a point on the boundary of the prediction is within this tolerance distance from a point on the boundary of the ground truth, it is considered a TP. Therefore, an appropriate NSD tolerance should usually be selected based on inter-annotator variability or some other heuristic. In our experiment, we initially set tolerance to a strict 0 voxels as for the smallest vessels the diameter can be as small as two voxels. We also calculated the value with 1 voxel tolerance, with the latter achieving around 25% increase in the NSD score. The choice of $t=0$ voxels for our primary experiment was driven by the small diameter of the kidney vessels, ensuring that even the smallest vessels were accurately assessed. For larger vessels, an increased tolerance would likely be acceptable without drastically

impacting the performance assessment, as the proportional effect of the tolerance would be smaller relative to the vessel size. However, for the smaller vessels, any increase in tolerance could lead to an overestimation of the segmentation accuracy. This is why, in cases where vessels of various sizes are involved, using a metric with a variable tolerance based on the ground truth vessel radius could provide a more nuanced evaluation. Such a method would allow larger vessels to be evaluated with a higher tolerance and smaller vessels with a stricter one, potentially offering a more balanced assessment across varying vessel sizes.

Capturing vessel continuity is vital for vascular network analysis. To this end, we applied cDice (see equation 7), a metric specifically developed to be sensitive to leaks in the continuity of vessel centerline irrespective of vessel radius. This metric reduces bias toward algorithms that predict larger structures such as large arteries but miss microvasculature structures. The cDice scores across all experiments are lower than the DSC or NSD ($t = 1$) reflecting its lack of bias towards the larger vessels and sensitivity to discontinuity in small vessels as seen in Fig. 9. In addition, the ranking of outputs does not follow the same pattern when the cDice rather than the DSC, NSD or ASSD scores are considered, with Expt. 1 having the highest score, and Expt. 3 the lowest. It should be noted that since cDice metric incorporates a morphological thinning approach to skeletonize i.e. reduce the vessels to their centerlines, small holes or highly multi-scale vasculature such as that imaged by HiP-CT can result in highly irregular skeletons e.g. ball-like or star-like structures which impact the ability of cDice to provide a comparative metric (See Supplementary Fig. 5) data.

The introduction of the HiP-CT technique has brought both opportunities and challenges to vascular segmentation. While its high resolution enables detailed visualisation of intricate structures, this comes at the cost of significantly increased manual annotation time, which is both time-consuming and subjective. Also, while HiP-CT offers significant advantages over techniques like MRA and CTA, its application is currently restricted to research scenarios and cannot be used in clinical settings where in-vivo imaging is required. At this stage, it is also important to acknowledge that some of the challenges in detecting the smallest vessels may come from annotation. Human annotators were responsible for the labeling process of this dataset utilising a semi-manual, interactive 3D region growing tool. As with all data created by human annotators there is the likelihood that some of smallest vessels in these datasets were missed by the independent annotators. This highlights why methods which treat human-annotated data as a gold standard rather than a ground truth, can be so powerful in e.g. active learning frameworks¹⁴³. Deep learning offers a valuable solution by automating the segmentation process, reducing annotation efforts, and improving consistency across datasets. By leveraging advanced algorithms, the need for manual intervention can be minimized, enhancing scalability and accuracy in segmenting complex vascular networks. Our study shows that nnU-Net successfully captures some of the small vessels missed by manual segmentation (see Supplementary Fig. 6).

It should be noted that the evaluation in this study was performed specifically on kidney vessels, which may limit the generalisability of the results to other types of vessels. Different anatomical structures, vessel sizes, and complexities in other organs could require adjustments in model or further validation to ensure applicability beyond kidney vessel segmentation. Also, another key limitation is that only the 3D nnU-Net framework was evaluated, which limits the ability to assess the performance of other segmentation models that might handle small, intricate structures differently. Additionally, inter-rater reliability assessment of the vessel segmentation was not performed, due to the time-consuming nature of manual segmentation at this high level of resolution and our approach for triple annotator verification and our approach for triple annotator verification.

From the discussion on metrics above, the benchmarking methodology of nnU-Net underperforms in several scenarios when applied to HiP-CT data on vascular segmentation. With whole organ HiP-CT we are able to manually segment and create training data down to the level of the interlobular arteries or approx. two generations from the capillary bed. These arteries have a minimum radius of ca. $40 \mu\text{m}^{21}$, (1–2 pixels in diameter). However, the experiment scores reveal that the failed prediction on such small vessels leads to discontinuity of the vascular network. The work in¹²⁵ has shown an improved continuity performance on vasculature segmentation with cDice as part of the loss function when training the network, though this comes at a high computational cost and has challenges already highlighted with the skeletonisation of multi-scale structures. Integration of rich hierarchical representations of thin and long structures by use of spatial attention and channels attention modules could also potentially improve the outcomes¹⁴⁴. Another challenge that may hinder vasculature continuity is edge identification and segmentation. Considering the imbalance of edge and non-edge voxels, an edge-reinforced neural network (ER-Net)¹⁴⁵ combining a reverse edge attention module⁵⁷ and an edge-enforced optimisation loss to discover spatial information of the edge structures could potentially increase the vasculature continuity on segmentation results. In both the small vessel and edge enhancement cases it is important to remember that HiP-CT is a propagation-based phase contrast technique and thus has intrinsically higher contrast for small thin structures and edges between structures. Tomographic reconstruction is a part of the process of HiP-CT data volume data production and can have a large impact on the ease of segmentation. Tuning the tomographic reconstruction pipeline to enhance specific features, whilst challenging could be a powerful approach to segmenting specific smaller structures.

Relative to the ground truth, the benchmark model appears to overlook certain vasculature, particularly struggling to identify certain large vessels as shown in Fig. 9. Upon closer examination of the pertinent sub-volumes, it becomes evident that these misdetected large vessels correspond to collapsed vessels within the organs. As shown in Fig. 9 of the prediction on kidney 1, collapsed vessels have a flat appearance, morphologically different from the noncollapsed vessels seen elsewhere. This suggests that the model's performance falters notably in the presence of disrupted structures, and these could be the potential cause for the observed lower scores when kidney 1 is used as the test set. In future approaches, addressing the challenge of collapsed vessels could involve employing a data augmentation method. By applying transformations that accurately simulate the morphological changes observed in collapsed vessels, the algorithm's ability to accurately segment collapsed vessels can be enhanced. Key transformations could include selectively compressing or flattening portions of

the vessels in certain images. Additionally, adjusting the vessels' shape while ensuring these modifications are medically realistic would be essential.

HiP-CT scans in high-resolution present challenges to deep learning methods applied to different tasks such as segmentation. These scans highlight more details that must be captured by the models. Using a larger patch size for training can ease the problem. However, there is a trade-off between the GPU memory and the size of the receptive field. As a result, integrating regional features and their global dependencies remains a research direction.

Conclusion and future work

The vascular system is vital for nurturing and supporting every organ in the body, and its abnormalities can be indicative of various diseases. Developing accurate and automated segmentation of vasculature, that can be applied to nascent imaging technologies such as HiP-CT, could pave the way for scalable vascular segmentation across large-scale datasets. Scalability in turn enables quantification across a large demographic population, therefore assisting the construction of a data-driven Vascular Common Coordinate Framework^{146,147} for human atlasing projects such as the Human BioMolecular Atlas Program¹⁴⁸, the Human Organ Atlas Project¹⁴⁹ and the Cellular Senescence Network Program¹⁵⁰.

This paper offers a comprehensive review of recent literature focusing on deep-learning techniques for blood vessel segmentation. We primarily delve into the recent deep-learning methodologies, emphasizing the challenges associated with vasculature segmentation. Our aim with this review is to lay a solid foundation for researchers, building robust models for vessel segmentation, especially using phase contrast imaging.

For these models to be widely accepted in the medical field, we need to address certain gaps in the current research. From our literature survey, we deduce the following:

- (i) Publicly accessible datasets for vascular segmentation are limited. Even within private datasets, the size available for training is often limited.
- (ii) Labeling the vascular region remains a formidable challenge and requires extensive effort.
- (iii) Most studies employ private data and provide internal validation. This makes direct comparisons between studies challenging and prevents the assessment of algorithm efficacy.
- (iv) Every phase of the imaging process; from sample preparation, imaging, modality specific artefact correction, pre-processing and algorithm optimisation to post-processing, holds immense importance. The post-processing stage, especially, has a great impact on the final segmentation^{151,152}. Many studies focus on mapping out complex blood vessel networks, mostly concentrating on larger vessels yet the significance of smaller vessels, which are frequently overlooked, cannot be understated from a (patho-)physiological point of view^{19,153}. Moreover, the choice of evaluation metrics and loss functions in vessel segmentation is often unclear, calling for clearer guidelines. Through our study, we aim to assist researchers in identifying the most suitable metrics for their analyses. One significant challenge with vessel segmentation models is their limited generalisability, as most studies rely on private datasets. To enhance adaptability, models should be cross-validated against diverse data. Providing both more public datasets and pre-trained model weights could benefit future research. By making our data publicly accessible, through a Kaggle competition (<https://www.kaggle.com/competitions/blood-vessel-segmentation/data>) and following its conclusion, through the human organ atlas portal (www.human-organ-atlas.esrf.eu), we invite researchers to evaluate their algorithms on HiP-CT data, fostering advancements in the field.

Data availability

The datasets analysed in the present study are publicly accessible and available online: <https://human-organ-atlas.esrf.eu/>.

Received: 20 June 2024; Accepted: 23 October 2024

Published online: 08 November 2024

References

- Pugsley, M. & Tabrizchi, R. The vascular system: An overview of structure and function. *J. Pharmacol. Toxicol. Methods* **44**, 333–340 (2000).
- Potente, M. & Carmeliet, P. The link between angiogenesis and endothelial metabolism. *Annu. Rev. Physiol.* **79**, 43–66 (2017).
- Sweeney, M. D., Kisler, K., Montagne, A., Toga, A. W. & Zlokovic, B. V. The role of brain vasculature in neurodegenerative disorders. *Nat. Neurosci.* **21**, 1318–1331 (2018).
- Moccia, S., De Momi, E., El Hadji, S. & Mattos, L. S. Blood vessel segmentation algorithms-review of methods, datasets and evaluation metrics. *Comput. Methods Programs Biomed.* **158**, 71–91 (2018).
- Rust, R. et al. A practical guide to the automated analysis of vascular growth, maturation and injury in the brain. *Front. Neurosci.* **14**, 244 (2020).
- Zhao, F., Chen, Y., Hou, Y. & He, X. Segmentation of blood vessels using rule-based and machine-learning-based methods: A review. *Multimed. Syst.* **25**, 109–118 (2019).
- Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P. & Johansen, H. D. Doubleu-net: A deep convolutional neural network for medical image segmentation. in *2020 IEEE 33rd International symposium on computer-based medical systems (CBMS)*, 558–564 (IEEE, 2020).
- Mortazi, A. & Bagci, U. Automatically designing CNN architectures for medical image segmentation. in *Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9*, 98–106 (Springer, 2018).
- Fraz, M. M. et al. Blood vessel segmentation methodologies in retinal images-a survey. *Comput. Methods Programs Biomed.* **108**, 407–433 (2012).

10. Deshpande, A. et al. Automatic segmentation, feature extraction and comparison of healthy and stroke cerebral vasculature. *NeuroImage: Clin.* **30**, 102573 (2021).
11. Luo, S. & Zhong, Y. Extraction of brain vessels from magnetic resonance angiographic images: Concise literature review, challenges, and proposals. In *IEEE Engineering in Medicine and Biology 27th Annual Conference*, 1422–1425 (IEEE, 2006).
12. Lesage, D., Angelini, E. D., Bloch, I. & Funka-Lea, G. A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Med. Image Anal.* **13**, 819–845 (2009).
13. Zhao, Y. et al. Automatic 2-D/3-D vessel enhancement in multiple modality images using a weighted symmetry filter. *IEEE Trans. Med. Imaging* **37**, 438–450 (2017).
14. Ajam, A. et al. A review on segmentation and modeling of cerebral vasculature for surgical planning. *IEEE Access* **5**, 15222–15240 (2017).
15. Kuo, A. H., Nagpal, P., Ghoshhajra, B. B. & Hedgire, S. S. Vascular magnetic resonance angiography techniques. *Cardiovasc. Diagn. Ther.* **9**, S28 (2019).
16. Walsh, C. et al. Imaging intact human organs with local resolution of cellular structures using hierarchical phase-contrast tomography. *Nat. Methods* **18**, 1532–1541 (2021).
17. Paganin, D., Mayo, S. C., Gureyev, T. E., Miller, P. R. & Wilkins, S. W. Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *J. Microsc.* **206**, 33–40 (2002).
18. Ackermann, M. et al. Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in Covid-19. *N. Engl. J. Med.* **383**, 120–128 (2020).
19. Ackermann, M. et al. The fatal trajectory of pulmonary Covid-19 is driven by lobular ischemia and fibrotic remodelling. *EBioMedicine* **85**, 104296 (2022).
20. Mentzer, S. J., Ackermann, M. & Jonigk, D. Endothelialitis, microischemia, and intussusceptive angiogenesis in Covid-19. *Cold Spring Harbor Perspect. Med.* **12**(10), a041157 (2022).
21. Rahmani, S. et al. Micro to macro scale analysis of the intact human renal arterial tree with Synchrotron Tomography. *BioRxiv* 2023–03 (2023).
22. Mutke, M. A. et al. Clinical evaluation of an arterial-spin-labeling product sequence in steno-occlusive disease of the brain. *PLoS ONE* **9**, e87143 (2014).
23. Bullitt, E. et al. Vessel tortuosity and brain tumor malignancy: A blinded study. *Acad. Radiol.* **12**, 1232–1240 (2005).
24. Bergström, G. et al. The Swedish CARDIOpulmonary BioImage study: Objectives and design. *J. Intern. Med.* **278**, 645–659 (2015).
25. Hotter, B. et al. Prospective study on the mismatch concept in acute stroke patients within the first 24 h after symptom onset-1000plus study. *BMC Neurol.* **9**, 1–8 (2009).
26. Forkert, N. D. et al. Automatic brain segmentation in time-of-flight MRA images. *Methods Inf. Med.* **48**, 399–407 (2009).
27. Bollmann, S. et al. Imaging of the pial arterial vasculature of the human brain in vivo using high-resolution 7t time-of-flight angiography. *Elife* **11**, e71186 (2022).
28. Baghaie, A. et al. Curvelet transform-based volume fusion for correcting signal loss artifacts in time-of-flight magnetic resonance angiography data. *Comput. Biol. Med.* **99**, 142–153 (2018).
29. Klimont, M. et al. Deep learning for cerebral angiography segmentation from non-contrast computed tomography. *PLoS ONE* **15**, e0237092 (2020).
30. Katz, D. A., Marks, M. P., Napel, S. A., Bracci, P. M. & Roberts, S. L. Circle of Willis: Evaluation with spiral CT angiography, MR angiography, and conventional angiography. *Radiology* **195**, 445–449 (1995).
31. Buerke, B., Wittkamp, G., Seifarth, H., Heindel, W. & Kloska, S. P. Dual-energy CTA with bone removal for transcranial arteries: Intraindividual comparison with standard CTA without bone removal and TOF-MRA. *Acad. Radiol.* **16**, 1348–1355 (2009).
32. Willinsky, R. A. et al. Neurologic complications of cerebral angiography: Prospective analysis of 2899 procedures and review of the literature. *Radiology* **227**, 522–528 (2003).
33. Kaufmann, T. J. et al. Complications of diagnostic cerebral angiography: Evaluation of 19,826 consecutive patients. *Radiology* **243**, 812–819 (2007).
34. Kierski, T. M. & Dayton, P. A. Perspectives on high resolution microvascular imaging with contrast ultrasound. *Appl. Phys. Lett.* **116**, 210501 (2020).
35. Jafree, D. J. et al. Spatiotemporal dynamics and heterogeneity of renal lymphatics in mammalian development and cystic kidney disease. *Elife* **8**, e48183 (2019).
36. Reichardt, M. et al. 3D virtual histopathology of cardiac tissue from Covid-19 patients based on phase-contrast x-ray tomography. *Elife* **10**, e71359 (2021).
37. Zhao, S. et al. Cellular and molecular probing of intact human organs. *Cell* **180**, 796–812 (2020).
38. Edlow, B. L. et al. 7 tesla MRI of the ex vivo human brain at 100 micron resolution. *Sci. Data* **6**, 244 (2019).
39. Taha, A. A. & Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **15**, 1–28 (2015).
40. Chatterjee, S. et al. Retinal blood vessel segmentation using edge detection method. In *Journal of Physics: Conference Series*, vol. 1717, 012008 (IOP Publishing, 2021).
41. Foruzan, A. H., Zoroofi, R. A., Sato, Y. & Hori, M. A Hessian-based filter for vascular segmentation of noisy hepatic CT scans. *Int. J. Comput. Assist. Radiol. Surg.* **7**, 199–205 (2012).
42. Zhao, Y., Rada, L., Chen, K., Harding, S. P. & Zheng, Y. Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images. *IEEE Trans. Med. Imaging* **34**, 1797–1807 (2015).
43. Shang, Y. et al. Vascular active contour for vessel tree segmentation. *IEEE Trans. Biomed. Eng.* **58**, 1023–1032 (2010).
44. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30** (2017).
45. Shamshad, F. et al. Transformers in medical imaging: A survey. *Med. Image Anal.* **88**, 102802 (2023).
46. Al-Hammuri, K., Gebali, F., Kanan, A. & Chelvan, I. T. Vision transformer architecture and applications in digital health: A tutorial and survey. *Vis. Comput. Ind. Biomed. Art* **6**, 1–28 (2023).
47. He, K. et al. Transformers in medical image analysis. *Intell. Med.* **3**, 59–78 (2023).
48. Livne, M. et al. A U-Net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease. *Front. Neurosci.* **13**, 97 (2019).
49. Phellan, R., Peixinho, A., Falcão, A. & Forkert, N. D. Vascular segmentation in TOF MRA images of the brain using a deep convolutional neural network. In *Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, 39–46 (Springer, 2017).
50. Hilbert, A. et al. BRAVE-NET: Fully automated arterial brain vessel segmentation in patients with cerebrovascular disease. *Front. Artif. Intell.* **3**, 552258 (2020).
51. Patel, T. R. et al. Multi-resolution CNN for brain vessel segmentation from cerebrovascular images of intracranial aneurysm: A comparison of U-Net and DeepMedic. in *Medical Imaging*, vol. 11314, 677–685 (SPIE, 2020).
52. Tetteh, G. et al. Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-D angiographic volumes. *Front. Neurosci.* **14**, 592352 (2020).
53. García, C. et al. A deep learning model for brain vessel segmentation in 3DRA with arteriovenous malformations. *arXiv preprint arXiv:2210.02416* (2022).
54. de Vos, V. et al. Automatic cerebral vessel extraction in TOF-MRA using deep learning. *arXiv preprint arXiv:2101.09253* (2021).

55. Chatterjee, S. et al. DS6: Deformation-aware learning for small vessel segmentation with small, imperfectly labeled dataset. arXiv preprint [arXiv:2006.10802](https://arxiv.org/abs/2006.10802) (2020).
56. Zhang, Y. & Chen, L. DDNet: a novel network for cerebral artery segmentation from MRA images. In *12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 1–5 (IEEE, 2019).
57. Zhang, B. et al. Cerebrovascular segmentation from TOF-MRA using model-and data-driven method via sparse labels. *Neurocomputing* **380**, 162–179 (2020).
58. Lee, K., Sunwoo, L., Kim, T. & Lee, K. J. Spider U-Net: Incorporating inter-slice connectivity using LSTM for 3D blood vessel segmentation. *Appl. Sci.* **11**, 2014 (2021).
59. Quintana-Quintana, O. J., De León-Cuevas, A., González-Gutiérrez, A., Gorrostieta-Hurtado, E. & Tovar-Arriaga, S. Dual U-Net-based conditional generative adversarial network for blood vessel segmentation with reduced cerebral MR training volumes. *Micromachines* **13**, 823 (2022).
60. Salvi, M. et al. Karpinski score under digital investigation: A fully automated segmentation algorithm to identify vascular and stromal injury of donors' kidneys. *Electronics* **9**, 1644 (2020).
61. He, Y. et al. Dpa-densebiasnet: Semi-supervised 3d fine renal artery segmentation with dense biased network and deep priori anatomy. in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI* **22**, 139–147 (Springer, 2019).
62. Taha, A., Lo, P., Li, J. & Zhao, T. Kid-net: Convolution networks for kidney vessels segmentation from ct-volumes. in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part IV* **11**, 463–471 (Springer, 2018).
63. He, Y. et al. EnMcGAN: adversarial ensemble learning for 3D complete renal structures segmentation. in *Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings* **27**, 465–477 (Springer, 2021).
64. Zhang, S. et al. A novel 3D deep learning model to automatically demonstrate renal artery segmentation and its validation in nephron-sparing surgery. *Front. Oncol.* **12**, 997911 (2022).
65. Xu, P., Sosnovtseva, O., Sørensen, C. M., Erleben, K. & Darkner, S. Extremely weakly-supervised blood vessel segmentation with physiologically based synthesis and domain adaptation. arXiv preprint [arXiv:2305.17054](https://arxiv.org/abs/2305.17054) (2023).
66. Li, B. et al. DUP-Net: Double U-PoolFormer networks for renal artery segmentation in CT urography. in *2023 International Joint Conference on Neural Networks (IJCNN)*, 1–8 (IEEE, 2023).
67. Dong, C., Xu, S. & Li, Z. A novel end-to-end deep learning solution for coronary artery segmentation from CCTA. *Med. Phys.* **49**, 6945–6959 (2022).
68. Gao, Z. et al. Vessel segmentation for x-ray coronary angiography using ensemble methods with deep learning and filter-based features. *BMC Med. Imaging* **22**, 10 (2022).
69. Wolterink, J. M., Leiner, T. & Išgum, I. Graph convolutional networks for coronary artery segmentation in cardiac CT angiography. in *Graph Learning in Medical Imaging. GLMI 2019. Lecture Notes in Computer Science*, 62–69 (Springer, 2019).
70. Li, Y. et al. Automatic coronary artery segmentation and diagnosis of stenosis by deep learning based on computed tomographic coronary angiography. *Eur. Radiol.* **32**, 6037–6045 (2022).
71. Song, A. et al. Automatic coronary artery segmentation of CCTA images with an efficient feature-fusion-and-rectification 3D-U-Net. *IEEE J. Biomed. Health Inform.* **26**, 4044–4055 (2022).
72. Zeng, A. et al. ImageCAS: A large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images. *Comput. Med. Imaging Graph.* **109**, 102287 (2023).
73. Tan, W. et al. Automated vessel segmentation in lung CT and CTA images via deep neural networks. *J. X-Ray Sci. Technol.* **29**(6), 1123–1137 (2021).
74. Nam, J. G. et al. Automatic pulmonary vessel segmentation on noncontrast chest CT: Deep learning algorithm developed using spatiotemporally matched virtual noncontrast images and low-keV contrast-enhanced vessel maps. *Eur. Radiol.* **31**, 9012–9021 (2021).
75. Guo, W., Gu, X., Fang, Q. & Li, Q. Comparison of performances of conventional and deep learning-based methods in segmentation of lung vessels and registration of chest radiographs. *Radiol. Phys. Technol.* **14**, 6–15 (2021).
76. Xu, Y., Mao, Z., Liu, C. & Wang, B. Pulmonary vessel segmentation via stage-wise convolutional networks with orientation-based region growing optimization. *IEEE Access* **6**, 71296–71305 (2018).
77. Nardelli, P. et al. Pulmonary artery-vein classification in CT images using deep learning. *IEEE Trans. Med. Imaging* **37**, 2428–2440 (2018).
78. Wu, R., Xin, Y., Qian, J. & Dong, Y. A multi-scale interactive U-Net for pulmonary vessel segmentation method based on transfer learning. *Biomed. Sign. Process. Control* **80**, 104407 (2023).
79. Wagner, W. L. et al. 3d analysis of microvasculature in murine liver fibrosis models using synchrotron radiation-based microtomography. *Angiogenesis* **24**, 57–65 (2021).
80. Sato, Y. et al. Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. *Med. Image Anal.* **2**, 143–168 (1998).
81. Frangi, A. F., Niessen, W. J., Vincken, K. L. & Viergever, M. A. Multiscale vessel enhancement filtering. in *Medical Image Computing and Computer-Assisted Intervention–MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings* **1**, 130–137 (Springer, 1998).
82. Mirone, A., Brun, E., Gouillart, E., Tafforeau, P. & Kieffer, J. The pyhst2 hybrid distributed code for high speed tomographic reconstruction with iterative reconstruction and a priori knowledge capabilities. *Nucl. Instrum. Methods Phys. Res. Sect. B: Beam Interact. Mater. Atoms* **324**, 41–48. <https://doi.org/10.1016/j.nimb.2013.09.030> (2014).
83. Yao, Z., Zhang, Z. & Xu, L.-Q. Convolutional neural network for retinal blood vessel segmentation. in *IEEE 9th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 1, 406–409 (IEEE, 2016).
84. Cervantes-Sanchez, F., Cruz-Aceves, I., Hernandez-Aguirre, A., Hernandez-Gonzalez, M. A. & Solorio-Meza, S. E. Automatic segmentation of coronary arteries in x-ray angiograms using multiscale analysis and artificial neural networks. *Appl. Sci.* **9**, 5507 (2019).
85. Nasr-Esfahani, E. et al. Segmentation of vessels in angiograms using convolutional neural networks. *Biomed. Sign. Process. Control* **40**, 240–251 (2018).
86. Samuel, P. M. & Veeramalai, T. VSSC Net: Vessel specific skip chain convolutional network for blood vessel segmentation. *Comput. Methods Programs Biomed.* **198**, 105769 (2021).
87. Iyer, K. et al. Angionet: A convolutional neural network for vessel segmentation in x-ray angiography. *Sci. Rep.* **11**, 18066 (2021).
88. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* **18**, 234–241 (Springer, 2015).
89. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440 (2015).
90. Badrinarayanan, V., Kendall, A. & Cipolla, R. Segnet: A deep convolutional encoder–decoder architecture for image segmentation. *IEEE Trans. Patt. Anal. Mach. Intell.* **39**, 2481–2495 (2017).
91. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Patt. Anal. Mach. Intell.* **40**, 834–848 (2017).

92. Zhao, H., Shi, J., Qi, X., Wang, X. & Jia, J. Pyramid scene parsing network. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2881–2890 (2017).
93. Khan, T. M., Robles-Kelly, A. & Naqvi, S. S. RC-Net: A convolutional neural network for retinal vessel segmentation. in *Digital Image Computing: Techniques and Applications (DICTA)*, 01–07 (IEEE, 2021).
94. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II* 19, 424–432 (Springer, 2016).
95. Milletari, F., Navab, N. & Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. in *4th International Conference on 3D vision (3DV)*, 565–571 (IEEE, 2016).
96. Chen, H., Dou, Q., Yu, L., Qin, J. & Heng, P.-A. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *Neuroimage* **170**, 446–455 (2018).
97. Huang, Q., Sun, J., Ding, H., Wang, X. & Wang, G. Robust liver vessel extraction using 3D U-Net with variant dice loss function. *Comput. Biol. Med.* **101**, 153–162 (2018).
98. Oktay, O. *et al.* Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* (2018).
99. Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
100. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. & Liang, J. Unet++: A nested u-net architecture for medical image segmentation. in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings* 4, 3–11 (Springer, 2018).
101. Sanches, P., Meyer, C., Vigon, V. & Naegel, B. Cerebrovascular network segmentation of MRA images with deep learning. in *IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, 768–771 (IEEE, 2019).
102. Dong, F. *et al.* CRAUNet: A cascaded residual attention U-Net for retinal vessel segmentation. *Comput. Biol. Med.* **147**, 105651 (2022).
103. Guo, C., Szemenyei, M., Pei, Y., Yi, Y. & Zhou, W. SD-UNet: A structured dropout U-Net for retinal vessel segmentation. in *IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, 439–444 (IEEE, 2019).
104. Pan, L.-S. *et al.* Coronary artery segmentation under class imbalance using a U-Net based architecture on computed tomography angiography images. *Sci. Rep.* **11**, 14493 (2021).
105. Kozłowski, M., Mosinska, A., Salzmänn, M. & Fua, P. Tracing in 2d to reduce the annotation effort for 3d deep delineation of linear structures. *Med. Image Anal.* **60**, 101590 (2020).
106. Dang, V. N. *et al.* Vessel-captcha: An efficient learning framework for vessel annotation and segmentation. *Med. Image Anal.* **75**, 102263 (2022).
107. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
108. González, C., Ranem, A., Pinto dos Santos, D., Othman, A. & Mukhopadhyay, A. Lifelong nnu-net: A framework for standardized medical continual learning. *Sci. Rep.* **13**, 9381 (2023).
109. Goodfellow, I. *et al.* Generative adversarial nets. *Adv. Neural. Inf. Process. Syst.* **27**, 2672–2680 (2014).
110. Luc, P., Couprie, C., Chintala, S. & Verbeek, J. Semantic segmentation using adversarial networks. *arXiv preprint arXiv:1611.08408* (2016).
111. Chen, C. *et al.* Deep learning for cardiac image segmentation: A review. *Front. Cardiovasc. Med.* **7**, 25 (2020).
112. Son, J., Park, S. J. & Jung, K.-H. Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks. *J. Digit. Imaging* **32**, 499–512 (2019).
113. Park, K.-B., Choi, S. H. & Lee, J. Y. M-GAN: Retinal blood vessel segmentation by balancing losses through stacked deep fully convolutional networks. *IEEE Access* **8**, 146308–146322 (2020).
114. Amran, D., Artzi, M., Aizenstein, O., Ben Bashat, D. & Bermanno, A. H. BV-GAN: 3D time-of-flight magnetic resonance angiography cerebrovascular vessel segmentation using adversarial CNNs. *J. Med. Imaging* **9**, 044503–044503 (2022).
115. Subramaniam, P. *et al.* Generating 3D TOF-MRA volumes and segmentation labels using generative adversarial networks. *Med. Image Anal.* **78**, 102396 (2022).
116. Xie, B., Tang, H., Duan, B., Cai, D. & Yan, Y. MLP-GAN for Brain Vessel Image Segmentation. in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5 (IEEE, 2023).
117. Kate, P., Ravi, V. & Gangwar, A. Fingan: Generative adversarial network for analytical customer relationship management in banking and insurance. *arXiv preprint arXiv:2201.11486* (2022).
118. Jain, Y. *et al.* Segmenting functional tissue units across human organs using community-driven development of generalizable machine learning algorithms. *bioRxiv* 2023–01 (2023).
119. Chen, J. *et al.* Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306* (2021).
120. Pan, C. *et al.* Deep 3D vessel segmentation based on cross transformer network. in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1115–1120 (IEEE, 2022).
121. Yu, H., Shim, J.-h., Kwak, J., Song, J. W. & Kang, S.-J. Vision transformer-based retina vessel segmentation with deep adaptive gamma correction. in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1456–1460 (IEEE, 2022).
122. Zhang, H. *et al.* T²IM-Net: Transformer in M-Net for retinal vessel segmentation. *J. Healthc. Eng.* **2022**, 9016401 (2022).
123. Nai, Y.-H. *et al.* Comparison of metrics for the evaluation of medical segmentations using prostate MRI dataset. *Comput. Biol. Med.* **134**, 104497 (2021).
124. Müller, D., Soto-Rey, I. & Kramer, F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC. Res. Notes* **15**, 210 (2022).
125. Shit, S. *et al.* cldice-a novel topology-preserving loss function for tubular structure segmentation. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16560–16569 (IEEE Computer Society, Los Alamitos, CA, USA, 2021).
126. Fedorov, A. *et al.* Evaluation of brain MRI alignment with the robust hausdorff distance measures. in *International Symposium on Visual Computing*, 594–603 (Springer, 2008).
127. Yeghiazaryan, V. & Voiculescu, I. Family of boundary overlap metrics for the evaluation of medical image segmentation. *J. Med. Imaging* **5**, 015006–015006 (2018).
128. Bertels, J. *et al.* Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice. in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II* 22, 92–100 (Springer, 2019).
129. Vaassen, F. *et al.* Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy. *Phys. Imaging Radiat. Oncol.* **13**, 1–6 (2020).
130. Nicholson, C. Diffusion and related transport mechanisms in brain tissue. *Rep. Prog. Phys.* **64**, 815 (2001).
131. Maier-Hein, L., Menze, B. *et al.* Metrics reloaded: Pitfalls and recommendations for image analysis validation. *arXiv.org* (2022).
132. Zhang, J., Wu, F., Chang, W. & Kong, D. Techniques and algorithms for hepatic vessel skeletonization in medical images: A survey. *Entropy* **24**, 465 (2022).

133. Wilke, J. et al. Appraising the methodological quality of cadaveric studies: Validation of the QUACS scale. *J. Anat.* **226**, 440–446 (2015).
134. Brunet, J. et al. Preparation of large biological samples for high-resolution, hierarchical, synchrotron phase-contrast tomography with multimodal imaging compatibility. *Nat. Protoc.* **18**, 1441–1461 (2023).
135. Jain, Y. et al. Sennet + hoa - hacking the human vasculature in 3d (2023).
136. Roth, H. R. et al. An application of cascaded 3D fully convolutional networks for medical image segmentation. *Comput. Med. Imaging Graph.* **66**, 90–99 (2018).
137. Drozdal, M., Vorontsov, E., Chartrand, G., Kadoury, S. & Pal, C. The importance of skip connections in biomedical image segmentation. in *International Workshop on Deep Learning in Medical Image Analysis, International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, 179–187 (Springer, 2016).
138. Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. in *Advances in Neural Information Processing Systems*, vol. 32, 8024–8035 (Curran Associates, Inc., 2019).
139. Alidoost, M. et al. Model utility of a deep learning-based segmentation is not Dice coefficient dependent: A case study in volumetric brain blood vessel segmentation. *Intell.-Based Med.* **7**, 100092 (2023).
140. Antonelli, M. et al. The medical segmentation decathlon. *Nat. Commun.* **13**, 4128 (2022).
141. Cardoso, M. J. et al. Monai: An open-source framework for deep learning in healthcare. arXiv preprint [arXiv:2211.02701](https://arxiv.org/abs/2211.02701) (2022).
142. Xian, R. P. et al. A multiscale x-ray phase-contrast tomography dataset of a whole human left lung. *Sci. Data* **9**, 264 (2022).
143. Yang, L., Zhang, Y., Chen, J., Zhang, S. & Chen, D. Z. Suggestive annotation: A deep active learning framework for biomedical image segmentation. in *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11–13, 2017, Proceedings, Part III* 20, 399–407 (Springer, 2017).
144. Mou, L. et al. CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging. *Med. Image Anal.* **67**, 101874 (2021).
145. Xia, L. et al. 3D vessel-like structure segmentation in medical images by an edge-reinforced network. *Med. Image Anal.* **82**, 102581 (2022).
146. Boppana, A. et al. Anatomical structures, cell types, and biomarkers of the healthy human blood vasculature. *Sci. Data* **10**, 452 (2023).
147. Börner, K. et al. Anatomical structures, cell types and biomarkers of the human reference atlas. *Nat. Cell Biol.* **23**, 1117–1128 (2021).
148. the nih human biomolecular atlas program. lcrai@caltech.edu 21 b Shendure Jay 9 Trapnell Cole 9 Lin Shin shinlin@uw.edu 2 e Jackson Dana 9, C.-U. T. C. L. et al. The human body at cellular resolution. *Nature* **574**, 187–192 (2019).
149. Walsh, C. et al. Multiscale three-dimensional imaging of intact human organs down to the cellular scale using hierarchical phase-contrast tomography. *Biorxiv* (2021).
150. 34, B. U. T. W. S. S. et al. Nih sennet consortium to map senescent cells throughout the human lifespan to understand physiological health. *Nature aging* **2**, 1090–1100 (2022).
151. Pain, C. D., Egan, G. F. & Chen, Z. Deep learning-based image reconstruction and post-processing methods in positron emission tomography for low-dose imaging and resolution enhancement. *Eur. J. Nucl. Med. Mol. Imaging* **49**, 3098–3118 (2022).
152. Hesamian, M. H., Jia, W., He, X. & Kennedy, P. Deep learning techniques for medical image segmentation: Achievements and challenges. *J. Digit. Imaging* **32**, 582–596 (2019).
153. Ackermann, M. et al. The bronchial circulation in Covid-19 pneumonia. *Am. J. Respir. Crit. Care Med.* **205**, 121–125 (2022).

Acknowledgements

The authors would like to express their gratitude to P. Masson (LADAF) for dissections of body donors, the Hannover Biobank for provision of donor organs, Emer O’Leary for segmentation input, E. Boller, K. Dollman, C. Berruyer and C. Jarnias for their help in setup, developments, improvements and imaging. The authors would like to express gratitude for the financial support provided by the Chan Zuckerberg Initiative DAF (2020-225394), an advised fund of SVCF, the MRC (MR/R025673/1), and ESRF beamtimes (md1252 & md1290). P.D.L. is supported by a Royal Academy of Engineering Chair in Emerging Technologies (CiET1819/10). This research was funded in part by the Wellcome Trust [209553/Z/17/Z]. The grant of the European Research Council (ERC); European Consolidator Grant, XHale to Danny Jonigk (ref. no.771883). This work was supported by the German Registry of COVID-19 Autopsies (DeRegCOVID, www.DeRegCOVID.ukaachen.de); supported by the Federal Ministry of Health-ZMV11-2520COR201), and the Federal Ministry of Education and Research as part of the Network of University Medicine (DEFEAT PANDEMIcs, 01KX2021). This research has been funded by the NIH Common Fund through the Office of Strategic Coordination/Office of the NIH Director under awards OT2OD033756 and OT2OD026671, by the Cellular Senescence Network (SenNet) Consortium through the Consortium Organization and Data Coordinating Center (CODCC) under award number U24CA268108, by the Kidney Precision Medicine Project grant U2CDK114886, by the NIDDK under awards U24DK135157 and U01DK133090 and by The Multiscale Human CIFAR project. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

E.Y. led this research project. E.Y., S.A., Y.J., and Y.Z. performed experiments. S.R., E.Y., and Y.Z. annotated data. E.Y., S.A., Y.J., and Y.Z. wrote the paper. J.B., A.B., C.W., M.A., D.J., and P.T. generated and organised the data. P.L., C.W., and P.T. managed and coordinated the planning and execution of research activities. All authors read and revise the final manuscript.

Declaration

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-77582-5>.

Correspondence and requests for materials should be addressed to E.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024