**Human Treelike Tubular Structure Segmentation: A Comprehensive Review and Future Perspectives**

Hao Li\textsuperscript{a,b,1}, Zeyu Tang\textsuperscript{b,1}, Yang Nan\textsuperscript{a,2} and Guang Yang\textsuperscript{a,c,\ast,2}

\textsuperscript{a}National Heart and Lung Institute, Faculty of Medicine, Imperial College London, London, United Kingdom
\textsuperscript{b}Department of Bioengineering, Faculty of Engineering, Imperial College London, London, United Kingdom
\textsuperscript{c}Royal Brompton Hospital, London, United Kingdom

**ARTICLE INFO**

**Keywords:**
Treelike tubular structure
Medical imaging
Segmentation
Review
Airways
Blood vessels

**ABSTRACT**

Various structures in human physiology follow a treelike morphology, which often expresses complexity at very fine scales. Examples of such structures are intrathoracic airways, retinal blood vessels, and hepatic blood vessels. Large collections of 2D and 3D images have been made available by medical imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT), optical coherence tomography (OCT) and ultrasound in which the spatial arrangement can be observed. Segmentation of these structures in medical imaging is of great importance since the analysis of the structure provides insights into disease diagnosis, treatment planning, and prognosis. Manually labelling extensive data by radiologists is often time-consuming and error-prone. As a result, automated or semi-automated computational models have become a popular research field of medical imaging in the past two decades, and many have been developed to date. In this survey, we aim to provide a comprehensive review of currently publicly available datasets, segmentation algorithms, and evaluation metrics. In addition, current challenges and future research directions are discussed.

**1. Introduction**

Anatomical structures with tree-shaped topologies are very commonly found in the human body. For example, in our transport system, blood vessels branch into networks to provide material exchange between tissues and cells while airway branches allow the exchange of gases in and out of the body. In our nervous system, tens of billions of neurons form hundreds of trillions of connections with one another (Fig 1). The analysis of such topology structure can reveal pathological conditions of the corresponding organs.

**Figure 1:** Examples of treelike structures in human body [1, 2, 3]

Human airways or respiratory tract start from the nasal cav-

\textsuperscript{\ast}This study was supported in part by the BHF (TG/18/5/34111, PG/16/78/32/402), the ERC IMI (101005122), the H2020 (952172), the MRC/MIC/PC/21013), the Royal Society (IEC/NSFC/211235), the SABER project supported by Boehringer Ingelheim Ltd, the Imperial College Undergraduate Research Opportunities Programme (UROP), the NVIDIA Academic Hardware Grant Program, and the UKRI Future Leaders Fellowship (MR/V023799/1).

\textsuperscript{\ast\ast}Corresponding author
g.yang@imperial.ac.uk (G. Yang)

ORCID(s): 0000-0003-2476-8664 (H. Li); 0000-0003-3789-2906 (Z. Tang); 0000-0002-654-3336 (Y. Nan); 0000-0001-7344-7733 (G. Yang)

\textsuperscript{1}H. Li and Z. Tang contributed equally to this review

\textsuperscript{2}Y. Nan and G. Yang are the co-last senior authors

Blood vessels are also tubular structures commonly found in our body anatomy, and are of great clinical importance as a diagnostic and prognostic factor. For example, the morphological shape of retinal blood vessels, which is imaged using fundus photography, can help diagnose hypertension, diabetes, and atherosclerosis [5, 6]. Many algorithms have been proposed for the segmentation of the blood vessels as well. We aim to provide a comprehensive review of the algorithm for human treelike structure segmentation, particularly for airways in this article. Pu et al. [7] has done a review on airway segmentation 10 years ago; Therefore, we will mainly focus on newly developed methods afterwards. In addition, we found the segmentation algorithms...
for human blood vessels are very similar and will include a few of them here for comparison and discussion. A detailed systematic review on blood vessel segmentation has already been done by [8, 9, 10, 11, 12] However, none of them focused on the big picture of summarising algorithms for segmenting all treelike structures in human anatomy. In this review, we would like to first briefly introduce different modalities used for imaging treelike structures. Then we move on to the segmentation, summarising representative conventional and ML/DL algorithms for delineating treelike anatomical regions in the past two decades. A list of open datasets and model implementations is compiled for future research work. Finally, we discuss the trend of development and current limitations.

1.1. Searching criteria
The literature search was conducted mainly by Z.T. and later reviewed by Y.N. and H.L. The inclusion of an article was based on agreement by at least two of the three authors. We used the following keywords for a systematic searching on Scopus: "airway" OR "vessel" AND "segmentation" AND "deep learning" OR "network" and 2603 results were found; We further limited the results to journal articles written in English that are already published and were left with 1190 papers for the title and abstract screening. 897 papers were excluded as they are either duplicated or have completely irrelevant topics with the keywords only being mentioned few times. During the writing of this review, a small fraction of papers that have relevant contents but were missed by our search on Scopus or excluded by an error during screening were added back, which in total gave us 72 of them for this study.

1. Imaging treelike structures in human
2.1. CT
X-rays used in medical imaging are usually called soft X-rays because of their low energy level. X-rays are generated inside the X-ray tube as shown in Fig. 3 which is a vacuum tube that utilizes high voltage to accelerate electrons released from a hot cathode. One type of X-ray, known as the characteristic X-ray, is generated when an electron in the metal atom is kicked out by the high-energy incoming electrons leaving a hole later filled by electrons from a higher energy state. The other type of X-ray, known as the Bremsstrahlung X-ray, is generated when the high-energy incoming electron passes through the atom and emits an X-ray when it loses kinetic energy. The generated X-ray is then directed to traverse the patients from various angles and is recorded by electronic detectors. Beer-Lambert law governs X-ray attenuation, which is influenced by a variety of factors such as beam energy, tissue type, sample thickness, and others. The higher the attenuation of the beam, the lighter the tissue appears on the reconstructed CT images; The lower the attenuation, the darker the tissue appears on the reconstructed CT images.

2.2. MRI
Instead of using ionizing radiation, strong external magnetic field is generated surrounding the patient lying in an MRI
Figure 4: Lung CT scans from EXACT’09 dataset [2]; from left to right is axial, coronal and sagittal plane respectively.

Figure 5: When strong external magnetic field is applied, the hydrogen nuclei precess along the direction of the field. Then a radio-frequency pulse with frequency equal to the Larmor frequency is applied perpendicular to the field. The RF pulse makes the magnetic moment of the hydrogen nuclei tilt. When the RF pulse stops, the nuclei realign themselves to the external magnetic field.

Figure 6: Time-of-flight (TOF) magnetic resonance angiography (MRA) of cerebral vessels [14], from left to right is the axial, coronal and sagittal plane respectively.

machine. The stronger the magnetic field, the faster the protons (nuclei of the hydrogen atom) precess. The precessing frequency lies in the range of radio frequency (RF) electromagnetic wave. The MRI machine sends RF pulses that have the same frequency of the processing and excite the proton to a higher energy state. When the RF pulse is switched off, the protons gradually relax themselves and release excess energy in the form of RF waves which can be detected. This signal is referred to as the free-induction decay (FID) response signal which is measured by a conductive field coil placed around the object being imaged. By evaluating the relaxation rate (longitudinal T1 and transverse T2) of protons in different tissues, we can obtain 3-D images of deep tissues with high resolution. A schematic diagram is shown below [13].

Figure 7: Light is projected through a set of filters and lenses, resulting in a focused doughnut-shaped form into the eye. Assuming the illumination and the object are aligned and focused, the image of the retinal exits the cornea through the central dark region of the doughnut.

Figure 8: Examples of fundus images from STARE dataset [15] captured by a TopCon TRV-50 fundus camera at 35 field of view. Each slide was digitized to produce a 605 x 700 pixel image, 24 bits per pixel (standard RGB).

2.3. Fundus Photography
Fundus retinal images allow non-invasive capture of the major anatomical structures of the retina, including the deep microvascular system. Depending on the instrument and the associated acquisition protocol, the main fundus imaging techniques are colour fundus photography (CFP) and scanning laser ophthalmoscope (SLO). The CFP utilizes a low-power microscope with a camera attachment to image the retina. The optical principle is similar to the indirect ophthalmoscope to provide a magnified view of the inner surface of the eye (see Figure 7). The SLO uses laser scanning to provide high-contrast images of blood vessels and uses wide-field instruments.

2.4. OCT
OCT creates cross-sectional images by performing a series of low time-coherence interferometry (LCI) depth-scans which measures absolute distances. The low time-coherence light source travels through the beam splitter, one being directed to the sample the other being directed to the reference mirror. OCT bathes the sample in electromagnetic waves and uses principles of interference to determine from what depth a photon was reflected inside a sample. Each layer in the sample has sharp change of refractive index which manifest themselves as intensity peaks in the interference pattern. The interference only occurs when the distances to the sample and reference are matched within coherence length of the light. A schematic diagram is shown below created with BioRender [16]. OCT is mainly used in ophthalmic and intravascular imaging. It can identify most retinal pathologies.
Human Treelike Tubular Structure Segmentation

3. Segmentation Algorithms

3.1. Conventional Segmentation

3.1.1. Basic image processing operations

For most biomedical images, different substances have different pixel intensity values (e.g., 0 for water and -1000 for air in CT scans). Therefore, the tubular structure can be roughly segmented by certain basic image processing operations, for instance, the intensity thresholding (to provide candidate regions) [19, 20], morphology operations (to remove noise or leakages) [21, 22, 23], filtering (to remove noise or enhance tubular structures) [24], or a combination of these operations [25, 26].

Adaptive thresholding (AT): Adaptive thresholding refers to methods that use a dynamic threshold strategy for threshold-based segmentation. For instance, Jiang et al. [19] proposed an adaptive local thresholding method based on multi-threshold probing. The airway prediction was obtained by combining the probed thresholding and non-maximum suppression post-processing. Pu et al. [20] relaxed the airway segmentation into the geometrical domain by analysing the principal curvature and lung anatomical structures. A progressively adaptive threshold-based modeling strategy was adopted to address the wide intensity range of airway walls. Though the dynamic threshold promotes robustness to some extent, it still faces the challenge of false-positive predictions, especially for those abnormal scans.

Mathematical morphology and cross-curvature evaluation (MM-CCE): Zana & Klein [21] segmented vessel-like patterns using a hybrid framework of morphological filtering and cross-curvature analysis. The vessels in the monochrome retinal image can be identified by mathematical morphology methods because they are natively straight, and connected, and their curvature changes gradually along the crest line. Cross-curvature analysis is applied to evaluate which features have linearly consistent curvature.

Knowledge-guided mathematical morphology (MM): Passat et al. [22] proposed a preliminary approach to enhance brain vessel segmentation in Magnetic Resonance Angiography (MRA) images by incorporating high-level anatomical knowledge into the segmentation process. The first stage integrates anatomical knowledge of vessels such as density, size, and orientation into a cerebral vascular atlas extracted by a skeletonization-based algorithm, utilizing a topology-preserving non-rigid registration method. This atlas then guides MM using adaptive sets of grey-level hit-or-miss operators in the second stage. The parameters of these morphology operators are adapted to the vascular structures using anatomical knowledge modelled by the cerebral vascular atlas. A similar strategy was employed in [23] to segment coronary arteries from 3D X-ray data sequences. However, it combines a region-growing algorithm in addition to the grey-level hit-or-miss transform.

Multi-scale line detection (MSLD): Nguyen et al. [24] suggested a multi-scale line detection (MSLD) based approach. Their approach enhances pixels representing the
Thresholding and mathematical morphology (T-MM): Aykac et al. [25] identified the candidate airways on 2D slices by intensity thresholding and morphology. After noise reduction, the grey reconstruction (via morphological closing operation) was conducted to extract airways by identifying the grey-scale minima. The intensity thresholding then removed most of the non-airway candidates to obtain the final airway structure.

Coarse-to-fine AT and MM (CTF AT-MM): Câmara Neto
et al. [26] presented an unsupervised statistical coarse-to-fine method for detecting blood vessels in fundus pictures. After background harmonization and noise reduction, an adaptive local thresholding approach is employed to coarsely estimate the vessel tree. The threshold is computed based on the spatial dependence between pixels using the cumulative distribution function and the normalized grey-level co-occurrence matrix to evaluate the regional influence, especially in the centre and border. The coarse segmentation is refined by a morphological reconstruction using a binary curvature map as a marker to eliminate pixel mislabelling. The suggested method effectively solved major image distortion by minimizing mislabelling of central vascular reflex areas and false-positive identification of abnormal patterns.

3.1.2. Template matching

A basic template matching method involves creating an image patch (template) that is suited to a certain aspect of the search image that we want to detect. Given the intensities of pixels in the search image $I_s$ and template image $I_t$, the template matching is normally defined by calculating the sum of absolute differences (SAD) between $I_s$ and $I_t$.

$$SAD(x, y) = \sum_{i=0}^{T_{rows}} \sum_{j=0}^{T_{cols}} |I_s(x+i, y+j) - I_t(i, j)|, \quad (1)$$

where $T_{rows}$ and $T_{cols}$ denote the rows and the columns of the template image, respectively. The segmentation performance of tubular structures heavily relies on the continual detection of identified candidates, while some pseudotracheal regions may occur in noisy cases, leading to leakage when dealing with small branches. To alleviate these issues, researchers designed novel filters or preset templates for image convolution to acquire the response image. Then the tubular structures can be located by analysing textual or shape features through template matching methods. Among different template matching approaches, one way is to use a preset circular mask to match the section of tubular structures [27, 28, 29], and another solution is to match a 3D model with multiple constraints [30, 31].

**Matched filtering (MF):** MF is a classic template matching method that employs predefined convolution kernels to convolute images to acquire intense responses or features of tubular structures. Based on the assumption that the cross-section of tubular structures can be regarded as a Gaussian function, one common solution is to match different Gaussian-shaped filters for segmentation. The binary segmentation can then be obtained by thresholding the response image. For instance, Hoover et al. [27] proposed an adaptive thresholding method by probing an area of the matched filter response image, decreasing the threshold iteratively. However, MF approaches will obtain intense responses to both tubular structures and edges, leading to false-positive predictions.

**Matched filtering with the first-order derivative of Gaussian (MF-FDOG):** To alleviate the redundant response of MF approaches, Zhang et al. [28] extended conventional MF approaches by integrating the first-order derivative of the Gaussian (FDOG). By checking the response intensity to the MF and FDOG, the non-vessel structures can be distinguished. This approach considerably reduced the false positive predictions of the original MF and discovered more small vessels.

**B-COSFIRE:** Based on the combination of shifted filter responses, Azzopardi et al. [29] suggested a non-linear filter named B-COSFIRE for detecting elongated patterns in images. The weighted geometric mean (of a group of different Gaussian filters) that provides orientation selectivity is calculated with the B-COSFIRE filter. Meanwhile, the rotation invariance is achieved by applying simple-shifting operations. However, these non-linear filters need to be manually adjusted to achieve optimal results on different datasets.

**The 3D cylindrical model matching (3D CMM):** Kris-sian et al. [30] presented an adaptive approach for detecting three-dimensional (3D) tubular structures based on the eigenvalues and eigenvectors of the Hessian matrix. The method begins with an analytical examination of those features across a variety of models, demonstrating that a response function based on both eigenvectors and gradient is the most robust. The method then employs a cylindrical vessel model with a circular Gaussian cross-section to guide the tubular structure detection on the feature map. Because the intensity profile of the cross-section affects size
estimation based on multi-scale analysis, a more accurate model with a bar-like convolved cross-section is proposed, along with an improved radius estimation algorithm based on this new model. Worz & Roehr [31] developed a new 3D cylindrical parametric intensity model for vessel segmentation which is directly fitted to image intensities using an incremental (segment-wise) Kalman filter. This mathematical model represents a cylindrical structure of a variable radius and directly depicts the image intensities of vessels and neighbouring tissues and a parameter that characterizes image blurring. Compared to previously presented Gaussian-shaped models, the new model more accurately describes a smoothed Gaussian cylinder, a more realistic representation that produces improved outcomes.

3.1.3. Region growing (RG)
The region growing or tracking methods segment the tubular structures by setting specific region growing rules [32, 33, 34, 35, 36, 37, 38] (e.g., most approaches predefine a 3D cylinder as a template to match the tubular structures), which can also be adaptive. Starting with an initial seed (given manually or automatically), the orientation and position of the tubular structure can be calculated by image-derived features such as intensities and gradients. However, due to the inhomogeneous intensity and noise in images, these approaches are prone to segmentation leakages and cannot segment small tubular structures effectively. Therefore, previous methods usually use region growing/tracking algorithms to obtain the main bronchi, then combine them with other image processing techniques, such as morphology-based methods, to segment the small trachea for further fine segmentation [39, 40, 41, 42].

![Region Growing](image)

**Figure 14:** Region growing.

**Adaptive RG:** To prevent airway leakages, Kiraly et al. [32] carried out a preset explosion parameter to detect leakages. Specifically, the threshold of the region growing algorithm is dynamically changed by assessing the current growth conditions (decreasing threshold while explosion occurs and vice versa). In addition, they also combined adaptive RG with morphology-based segmentation to accurately segment airway trees.

**Energy-constrained RG:** Fetita et al. [33] proposed an energy-constrained region-growing algorithm for 3D airway segmentation. The RG is controlled by the local energy minimization exploiting the bronchial structure in terms of intensity and 3D topology. To acquire more accurate predictions of small bronchus, they also developed a mathematical morphology operator based on selective marking and depth-constrained connection costs. The experimental results indicated that their approach was robust to a large spectrum of airway pathologies, including severe stenosis.

**Multi-seeded fuzzy connectivity (MSFC):** Tschirren et al. [34] employed a 3D cylindrical ROI (region of interest) with multi-seeded fuzzy connectivity (MSFC) to overcome image noise and gradients. The ROI for segmenting the airway adapts to its geometrical dimensions, orientation, size, and position. The MSFC grows two regions (foreground and background) simultaneously in a competitive manner, based on the similarity between input voxels and the seed voxel as fuzzy membership values. Although the prediction from MSFC is robust with fewer airway leakages, it has high computational costs, especially for large 3D volumes.

**Geometrical RG:** Different from the conventional region-growing approaches that perform growth through pixel/voxel intensity, the geometrical region growing utilizes geometrical features as a constraint for region growing. These geometrical features are usually used to approximate the width, size, and orientation of tubular structures. For instance, Martinez-Perez et al. [35] calculated two geometrical features (gradient magnitude and maximum principal curvature of the Hessian matrix) based on the first and second derivatives of the intensity value. The growth is then confined to low gradient magnitude regions in the first stage, along with spatial information about the 8-neighbouring pixels. Fabijańska [36] proposed a two-pass region growing algorithm to segment 3D airway structures, one (classic region growing) for rough airway extraction and another (based on morphological gradient) for refining segmentation. Besides, Cetin et al. [37] utilized a second-order tensor derived from directional intensity measurements for extracting coronaries from computed tomography angiography (CTA) volumes. The anisotropic second-order tensor fits inside a vessel drives the segmentation evolution starting from a seed point. As a result, the second-order tensor effectively copes with complex structures like bifurcations. Cetin & Unal [38] utilized a similar method with higher-order tensors.

**CTF RG:** Due to the limitation of region-growing approaches, researchers proposed a multi-stage strategy to segment peripheral branches. The semantic predictions obtained by region growing are usually regarded as rough segmentation, followed by refinement algorithms. For instance, Bartz et al. [39] combined 3D region growing with 2D wave propagation and template matching to segment small branches of airway tree structures. The 2D wave propagation is presented to address the partial volume and low-resolution effects, while the 2D template matching is used to segment the small lumen further. This approach is further investigated in [40]. Graham et al. [41] first applied region growing for conservative segmentation of trachea and main branches, then they scanned and located all the airway cross-sections by applying an elliptic decision function on three 2D planes (transverse, coronal and sagittal planes). After that, they
built a graph to connect all these airway segment candidates.

### 3.1.4. Level set segmentation

The level set (LS) is a method based on partial differential equations (PDE), via moving curves and surfaces with curvature-based velocities. Given a moving speed \( v \), and time \( t \), the level-set function \( \Phi \) satisfies

\[
\frac{\partial \Phi}{\partial t} = v |\nabla \Phi|,
\]

where \( |\cdot| \) represents the Euclidean distance. In image world, given the pixel \((i,j)\), the motion equation can be presented as

\[
\Phi(i,j,t + \Delta t) - \Phi(i,j,t) = \max[F, 0] |\nabla^+ x(i,j)| + \min[F, 0] |\nabla^- x(i,j)| = 0,
\]

where \( F \) indicates the force normal to the surface, and

\[
\nabla^+ x(i,j) = \max[0, \Delta^+ x \Phi(i,j)]^2 + \min[0, \Delta^+ x \Phi(i,j)]^2, F > 0
\]

\[
\nabla^- x(i,j) = \max[0, \Delta^- x \Phi(i,j)]^2 + \min[0, \Delta^- x \Phi(i,j)]^2, F < 0
\]

It is of note that \( \Delta^+ x \Phi \) and \( \Delta^- x \Phi \) are the left and right-side finite difference for the given pixel. Specifically, level-set segmentation approaches are widely used in medical images with intensity inhomogeneity. By incorporating image-guided restrictions and prior clinical knowledge, level-set segmentation approaches are well-suited for segmenting tubular structures with complex architecture and size variability. The level set segmentation can be classified into edge-based \([43, 44]\) and region-based \([45]\).

#### Edge-based LS:

Edge-based level set approaches are mainly driven by intensity or gradient-derived external forces. For instance, Law et al. \([43]\) determined the brain vessel border orientation using gradient-flux symmetry along the vessel centerlines and gradient-flux asymmetry along vessel borders, which was then utilized to fit an active contour model by minimizing the weighted local variance. Lorigo et al. \([44]\) introduced a curve evolution-based vessel segmentation approach that models the vessel walls as manifolds that evolve iteratively to minimize an energy criterion based on image intensity and local smoothness. However, the E-LS becomes unstable and is prone to leakage when the initialized curve is out of the boundary of interest.

#### Region-based LS:

The region-based model was introduced to address some of the difficulties associated with edge-based deformable models, such as the leakage problem that typically occurs when coping with noise or non-uniform intensity distributions. The deformable surface was first thought to be moving under the image foreground and background regions in the region-based model. Both regions are considered statistically homogeneous, and the main distinctions in this type of algorithm pertain to how the statistics of the two regions are defined. Despite the benefits of adaptive topology, region-based deformable models have a higher computation complexity than edge-based deformable models. Klepaczko et al. \([45]\) presented a method for validating vessel segmentation algorithms using a customed Magnetic Resonance Angiography (MRA) simulation framework. A reference model of a cerebral artery tree was created for this purpose using realistic Time-of-Flight (TOF) MRA data.

Blood flow was simulated using this specified geometry, and a series of TOF pictures were created using various acquisition protocol parameters and signal-to-noise ratios. A level-set algorithm was then used to reconstruct the synthetic artery tree, which was next compared to the reference model. This validation procedure is especially beneficial for detecting minor vascular structures.

### 3.2. Machine Learning Based Segmentation

Most conventional methods reviewed above have numerous tuning parameters, which are hard to estimate manually. Machine learning (ML) algorithms are trained with specific features, usually based on data statistics, to automatically obtain an optimal set of parameters for a model or a classifier. There are two types of ML methods depending on their training process: unsupervised and supervised. Unsupervised learning is ideal for scenarios when gold standard (GS) labels are not available or exploratory data analysis is needed. On the other hand, supervised learning requires GS datasets to train the learning model or classifier.

However, since ML approaches typically rely on manually extracted low-level features and criteria, they have a limited ability to learn features automatically, and segmentation accuracy is relatively low compared to deep learning approaches.

#### Figure 15: Level set segmentation.

#### Figure 16: Machine learning based segmentation.
### Table 1
Conventional segmentation methods

| Author et al. | Year | Region | Modality | Method | Validation Dataset (number of cases) | Metrics | Method Class |
|---------------|------|--------|----------|--------|--------------------------------------|---------|--------------|
| Jiang et al.  | 2003 | RV     | CFP      | AT     | DRIVE (40), STARE (20)               | Acc: 0.9212, AUCROC: 0.9114 | Basic image processing operations |
| Pu et al.     | 2011 | LA     | CT       | AT     | CT scans (75), DRIVE (40)             | Acc: 0.9367, AUCROC: 0.8906 |                       |
| Zana & Klein  | 2001 | RV     | CTA      | MM-CCE | DRIVE (40)                           | Acc: 0.9407, AUCROC: 0.9324 |                       |
| Passat et al. | 2006 | BV     | MRA      | Knowledge-guided MM | PC MRA scans (30) | Acc: 0.77, Sensitivity: 0.99 |
| Bouraoui et al.| 2010 | CV     | CTA      | Knowledge-guided MM | DRIVE (40) | Acc: 0.9407, AUCROC: 0.9324 |                       |
| Nguyen et al. | 2013 | RV     | CFP      | MSLD   | STARE (20)                           | Acc: 0.73, Sensitivity: 0.71-0.10 |                       |
| Câmara Neto et al. | 2017 | RV     | CFP      | CTF AT-MM | DRIVE (40) | Acc: 0.9631, Sensitivity: 0.7695, Specificity: 0.9537 |                       |
| Hoover et al. | 2000 | RV     | CFP      | MF     | DRIVE (40)                           | Acc: 0.9382, Sensitivity: 0.7120, FPR: 0.0276 | Template matching |
| Zhang et al.  | 2010 | RV     | CFP      | MF-FDOG | DRIVE (40)                           | Acc: 0.9442, Sensitivity: 0.9407, Specificity: 0.9563 |                       |
| Azzopardi et al. | 2015 | RV     | CFP      | B-COSFIRE | DRIVE (40) | Acc: 0.9344, Sensitivity: 0.7246, Specificity: 0.9701, FPR: 0.0345 |                       |
| Krissian et al. | 2000 | BV     | X-ray    | 3D CMM | X-ray scans (10)                     | Acc: 0.9487, Sensitivity: 0.9567, Specificity: 0.9387 |                       |
| Worz & Rohr   | 2007 | TA     | MRA      | 3D CMM | MRA scans (n/a)                      | Acc: 0.9484, Sensitivity: 0.7177, Specificity: 0.0247 |                       |
| Kiraly et al. | 2002 | LA     | CTA      | Adaptive RG | CT scans (10) | Acc: 0.9704, Sensitivity: 0.9442, Specificity: 0.9614 |                       |
| Fetita et al. | 2004 | LA     | CT       | Energy-constrained RG | CT scans (30) | Acc: 0.91, Sensitivity: 0.91, Specificity: 0.67 | Region growing |
| Tschirren et al. | 2005 | LA     | CT       | MSFC   | LD CT scans (22)                     | Acc: 0.9344, Sensitivity: 0.7246, Specificity: 0.9701, FPR: 0.0345 |                       |
| Martinez-Perez et al. | 2007 | RV     | CFP      | Geometrical RG | DRIVE (40) | Acc: 0.9410, Sensitivity: 0.9401, Specificity: 0.7506, FPR: 0.0431 |                       |
| Fabijańska | 2009 | LA     | CT       | Geometrical RG | CT scans (10) | Acc: 0.99, Sensitivity: 0.9964, Specificity: 0.964 |                       |
| Cetin et al.  | 2013 | CV     | CTA      | Geometrical RG | ROTTERDAM (32) | Acc: 0.9286±0.03453 |                       |
| Cetin & Unal  | 2015 | BV     | MRA      | Geometrical RG | MRA scans (50) | Acc: 0.973 |                       |
| Bartz et al.  | 2003 | LA     | CT       | CTF RG  | CTA (24)                           | Acc: 0.985, Sensitivity: 0.986, Specificity: 0.86 |                       |
| Mayer et al.  | 2004 | LA     | CT       | CTF RG  | CT scans (22)                     | Acc: 0.86, Sensitivity: 0.86, Specificity: 0.90 |                       |
| Graham et al. | 2010 | LA     | CT       | CTF RG  | CT scans (23)                     | Acc: 0.86, Sensitivity: 0.86, Specificity: 0.90 |                       |
| Mendonca et al. | 2006 | RV     | CFP      | CTF RG  | DRIVE (40)                           | Acc: 0.9463±0.0065, Sensitivity: 0.7344, Specificity: 0.7246, FPR: 0.0270 |                       |
| Law et al.  | 2010 | BV     | MRA      | Edge-based LS | MRA scans (4) | Acc: 0.9479±0.0123, Sensitivity: 0.7123, Specificity: 0.9901 |                       |
| Lorigo et al. | 2001 | BV     | MRA      | Edge-based LS | MRA scans (n/a) | Acc: 0.9479±0.0123, Sensitivity: 0.7123, Specificity: 0.9901 |                       |
| Klepacko et al. | 2016 | BV     | MRA      | Region-based LS | MRA scans (n/a) | Acc: 0.9479±0.0123, Sensitivity: 0.7123, Specificity: 0.9901 |                       |

The regions include the retinal vessel (RV), lung airways (LA), lung vessel (LV), brain vessel (BV), coronary vessel (CV), and thoracic aorta (TA). All unnamed datasets are private and may have specific modalities such as phase-contrast (PC), low-dose (LD), standard-dose (SD), or time-of-flight (TOF). All metrics are shown as mean ± standard deviation or mean only.
3.2.1. Unsupervised

Unsupervised ML algorithms are trained without a gold standard or ground truth. These methods discover a model or a classifier capable of describing a hidden grouping rule of input features without prior knowledge or supervision. The lack of GS can limit unsupervised segmentation performance, which is often inferior to supervised approaches.

**GMM:** In the Gaussian mixture model (GMM), the image voxels are divided into backgrounds and tubular structures. Each class is characterized by a stochastic modelling process using a mixture of Gaussian or Gaussian-like distributions. Then the parameters of the mixture model \((\pi, \mu, \Sigma)\) are estimated by maximizing the log-likelihood.

\[
\theta(\pi, \mu, \Sigma; X) = \sum_{m=1}^{M} \ln \sum_{k=1}^{K} \pi_k N(x_m|\mu_k, \Sigma_k)).
\] (6)

An unsupervised statistical modelling strategy using two-level GMM for segmenting the 3D cerebrovascular system from MRA data is presented in [46]. Each voxel class is characterized by a low-level intensity histogram across the space and a high-level statistical dependence between adjacent voxels. The background’s low-level process is represented by two Gaussian distributions and one Rayleigh distribution, whereas the blood vessels are modelled by one Gaussian distribution. The parameters of the suggested GMM are automatically estimated utilizing an Expectation-Maximization (EM) algorithm [96] before the Maximum Likelihood Estimation (MLE) classification. Spatial contextual information has been incorporated through the 3D Markov random field (MRF) [97], which serves as a prior distribution of the true label of the class of interest to avoid misclassification caused by noise and signal loss, which frequently occurs for classifiers using the pixel intensity solely.

**GMM and RG (GMM-RG):** Oliveira et al. [47] also presented a method for segmenting nodules and vessels based on a GMM integrating the RG. The histogram combines three Gaussian models representing the nodules, liver parenchyma, and vasculature. The Levensen-Marquart (LM) optimizer is used to discover the parameter values that provide the best fit. The RG strategy allocates voxels with intensity lower than the lower threshold to nodules and is used as seeds in a process that aggregates neighbouring voxels with intensity lower than the upper threshold. The hepatic and portal veins are then identified using liver anatomical knowledge and a vein-tracking algorithm.

**Feature-based GMM:** Roychowdhury et al. [48] used a feature-based GMM to segment blood vessels accurately on pathological retinal images containing bright lesions, red lesions, and brightness/contrast variation while remaining unaffected by the gradient or any derivative of the objective. The GMM classifier is deployed to classify pixels as a vessel or non-vessel using a set of eight features collected from pixel neighbourhoods and first and second-order gradient images, which reduces the reliance on training data and improves the robustness. The number of pixels applied classification significantly decreases by removing major vessels overlapped by the high-pass filtered image and the morphologically (Top-hat) reconstructed image.

**Clustering:** Clustering is the process of discovering natural groupings in the feature space automatically. A cluster is frequently a dense area in the feature space where domain objects (e.g., image voxels) are closer to this cluster than others. Given the number of clusters \(K\), the objective function of \(N\) pixels within an image is presented as:

\[
J = \sum_{j=1}^{K} \sum_{i=1}^{N} |x_i - c_j|^2,
\] (7)

where \(c_j\) is the centroid for cluster \(j\). A centre vector represents each cluster in centroid-based clustering, which is unnecessarily a dataset member. The standard optimization objective is to locate cluster centres and assign objects with the minimized squared distance to the cluster centre.

**SWFCM:** Kande et al. [49] presented a four-step unsupervised fuzzy-based vascular segmentation method to minimize erroneous vessel detection. First, matched filtering is applied to enhance blood vessels. The responses are then segmented using spatially weighted fuzzy c-means clustering (SWFCM) based thresholding, which preserves the spatial structure of the vascular tree. To speed up the procedure, the grey-level histogram of the image is used to determine the parameters of the FCM algorithm instead of the entire image data.

**K-means clustering (KMC):** Goceri et al. [50] proposed an unsupervised and adaptive clustering approach for segmenting portal and hepatic veins from liver magnetic resonance images. K-means clustering is employed at the initial stage to adaptively identify tissue regions and build a marker image. The results are processed iteratively with a linear contrast stretching technique in the next stage, using both edge and region-based information to generate a mask. The markers and mask images are then used to refine vascular areas using binary morphological image reconstruction [98]. The suggested methodology outperformed previous thresholding-based methods because it automatically and adaptively calculates thresholds and weights.

**Swarm optimization:** Swarm optimization (SO) is a bio-inspired algorithm that uses a population of candidate solutions to search for an optimal solution in the solution space iteratively. It differs from other optimization techniques in that it simply requires the objective function and is unaffected by the gradient or any derivative of the objective.

**Ant colony optimization (ACO):** Because MF alone may not be sufficient to retrieve the capillaries, Ant colony optimization (ACO) segmentation results are incorporated in [51]. After basic preprocessing, the image is processed by the MF and the ant algorithm in parallel. In ant algorithms, several artificial ants build solutions and exchange information on the quality to find approximate solutions to the given optimization problem, using a communication mechanism.
similar to that used by actual ants [99]. The result is then merged with the MF response using a simple OR gate, followed by a length filtering to acquire the entire vasculature. However, this method mistook pathological areas for vessels, and the segmented vessels are slightly thicker than manual labelling.

Artificial bee colony (ABC): In [52], another swarm-based optimization algorithm is utilized to develop an unsupervised retinal blood vessel segmentation approach. After enhancing the brightness, an artificial bee colony (ABC) stochastic optimization on the fuzzy c-means (FCM) compactness fitness function is used to determine cluster centres representing vessels and backgrounds. However, vessels with small diameters are usually distorted. They cannot be correctly categorized at this first level of segmentation since they are always confused with surrounding pixels, especially in pathological images. Hence, the second optimization step is employed to update the acquired cluster centres using the generalized pattern search (GPS) method [100] to maximize the total thinness ratio of vessel segments. This method is proven to be robust to noise, exudates, haemorrhages, and variations in the pigment epithelium after deploying on DRIVE and STARE datasets. Su et al. [101] proposed CCABC which has faster convergence and can obtain higher quality solutions.

Whale optimization algorithm (WOA): Hassan & Has-sanien [53] further explored a novel swarm optimization method for extracting the vasculature from retinal fundus images. A hybrid model of multi-level thresholding is used in conjunction with the whale optimization algorithm (WOA) to segment the enhanced image. The WOA [102] is utilized as an optimizer to determine thresholds that maximize between-class variance. On the DRIVE test dataset, the proposed method achieves state-of-the-art sensitivity and the results are visually identical to the ground truth.

MCET-HHO: In [54], a novel unsupervised vessel segmentation approach is proposed based on optimized multi-level thresholding. The main processing stage is divided into two configurations after image smoothing. The new optimized top-hat, homomorphic filtering, and median filter were the first to segment thick vessels. The Minimum Cross-Entropy Thresholding - Harris Hawks Optimization (MCET-HHO) method [103], a state-of-the-art multi-level image segmentation algorithm, is applied in the second configuration to segment thin vessels. Finally, an OR logic combines the thick and thin vessel masks. The suggested method outperforms most state-of-the-art methods described in terms of specificity and accuracy, while the remaining computational costs are small. Another multilevel thresholding method named Salp Swarm Algorithm (SSA) is optimized by Zhang et al. [104] and has potential to be implemented in vessel segmentation.

3.2.2. Supervised

In contrast, supervised learning algorithms infer a classification rule from manually annotated training pairings. The introduction of prior knowledge in the training process makes these methods usually outperform the unsupervised ones. However, that means the training time and computational cost are also increased.

KNN: The K-nearest neighbour (KNN) method is a classical supervised classification algorithm. The idea of the K-nearest neighbour method is straightforward and intuitive: if most of the K most similar (i.e., most neighbouring) samples in the feature space belong to a class, then the sample also belongs to that class.

KNN and primitive-based method (KNN-PBM): Staal et al. [55] introduced a ridge-based retinal vessel segmentation methodology using a KNN classifier, taking advantage of the intrinsic feature that vessels are elongated structures rather than pixel representation. The primitive-based method (PBM) is developed to extract intrinsic features (image primitives) by grouping ridge pixels into sets that approximate straight-line elements. Each line element provides a local coordinate frame within each set, from which local characteristics are retrieved for each pixel. A sequential forward selection method is utilized to choose 27 features from convex sets and individual pixels, and then a KNN classifier is used for segmentation. This approach achieves high accuracy on the STARE dataset, but it is training data dependent, sensitive to false edges, and relatively slow due to extensive feature sets.

KNN and RG (KNN-RG): Lo et al. [56] employed KNN to classify the airway and non-airway voxels. The voxel-wise initial feature vector was constructed using the Hessian matrix and Gaussian kernel from the training data. The final airway tree segmentation is achieved using a 3D region growing method with a decision function that combines the probability map from the KNN classifier with the vessel orientation similarity evaluated by Hessian matrix analysis, which reveals how similar the orientation of a possible airway is to an adjacent vessel. By combining the KNN classifier with the vascular axial orientation constraint, the leakage is effectively limited, and the false positive rate of the segmentation results is significantly reduced. However, some fine tracheas are lost, reducing the total number of branches segmented.

SVM: A Support Vector Machine (SVM) is a class of generalized linear classifiers that performs binary classification of data in a supervised learning manner, where the decision boundary is the maximum-margin hyperplane solved for the learned samples. For a soft-margin SVM, the objective function \( J \) is minimized to find the hyperplane.

\[
J = \lambda ||w||^2 + \frac{1}{n} \sum_{i=1}^{n} \max(0, 1 - y_i(w^T x_i - b)) \tag{8}
\]

where \( y_i \) is the \( i \)-th target, \( w^T x_i - b \) is the \( i \)-th prediction.

SVM with Line detectors (SVM-LD): The application of line detectors as a feature vector for SVM to segment retinal vessels is proposed by Ricci & Perfetti [57]. The line strength response from three orthogonal line detectors
is thresholded to create a simple 3D feature vector for quick supervised classification using a support vector machine (SVM). In comparison to a convolution with a 2D kernel, the suggested line detector has a lower computing cost and is less sensitive to noise due to averaging. Also, because no hypothesis is based on the cross-section profile, this detector achieves reliable detection in the presence of vessels of various sizes. However, this approach is height dependent on the training dataset.

**SVM and graph-cut (SVM-GC):** Meng et al. [58] utilized the SVM to provide an accurate supervised extraction strategy for the complicated airway tree. First, Hessian analysis and adaptive multi-scale cavity enhancement filter (CEF) enhance the areas with tube-like and cavity-like structures. SVM is then used to classify voxel candidates and remove false positives using 3D feature vectors based on local pixel intensity and structures. Finally, the candidate voxel map is refined using the graph-cut method to create the integrated airway tree. The introduction of the SVM classifier improves the accuracy, but it is highly dependent on the training data and is computationally expensive. The experimental results also showed that the lung tracheal tree segmented by this method was prone to branch breakage, particularly in the case of small tracheal branches. Zhai et al. [59] proposed a graph-cuts objective function combined with a Hessian based filter to segment lung vessels. They validated the method on 20 CT scans in VESSEL12 challenge and achieved a competitive performance.

**SVM with multi-scale filter (SVM-MSF):** Lee et al. [60] proposed a hybrid method integrating multi-scale filtering and an SVM classifier to segment the airway tree on 3D chest CT scans automatically. First, a modified 2D Hough transform [105] was used to detect a starting seed point in the trachea. Then, a multi-scale tubular structure-based filter is employed to identify the initial candidate airway regions, including a grey-scale morphological operation and a tubular structure detection [106, 107]. An SVM classifier is then utilized to search for weak-contrast narrow airways and unconnected regions missed by the two enhancement filters. The fuzzy connectedness technique extracts 28-dimensional feature vectors to discover additional candidate airways while simultaneously suppressing FPR. However, the method also has shortcomings, including a high probability of false positives and the potential for leakage.

**Trainable GMM:** Soares et al. [61] proposed a supervised learning approach for segmenting the blood vessels in retinal images using 2D Gabor wavelet. To account for vessels of various widths, the feature vector comprises the pixel’s intensity and 2D Gabor wavelet responses collected at multiple scales. The wavelet helps detect and analyse localized characteristics and singularities [108], such as edges and blood vessels in this case. Moreover, because the Gabor wavelet can tune to specific frequencies, noise filtering, and vessel augmentation can be performed in one step. Then, a GMM and a Bayesian classifier with class-conditional probability density functions (likelihoods) are employed for fast classification. Probability distributions are computed using a training set of labelled pixels from manual segmentations. However, the training takes hours and is prone to overfit.

**Ensemble modules:** Based on an ensemble system of bagged and boosted decision trees, Fraz et al. [62] proposed an accurate vessel segmentation approach for colour fundus photography. The Ensemble classification [109] is the technique of strategically generating and combining multiple classifiers. The decision trees are used as the classification model in this approach, and the outcomes of these weak learners are pooled using bootstrap aggregation (bagging) and boosting algorithms. The 9D feature vector comprised the inverted pixel intensity, gradient orientation maps, top-hat transform responses, line detector responses, and Gabor filter responses, which achieve high robustness. However, because of the boosting strategy with 200 decision trees, this method has a significant computational complexity.

### 3.3. Deep Learning Based Segmentation

Deep learning (DL) approaches automatically extract features to explore raw data without using handmade features. They can automatically learn multiple-level patterns and are not constrained by a specific application. Therefore, they have a superior generalization and recognition capabilities. DL models are composed of layers organized in a hierarchal structure called artificial neural networks (ANN) and interpret input data into meaningful output. The network’s parameters, such as kernel coefficients and fully-connected (FC) layer weights, are learned in the training process to represent the input data comprehensively. In this review, all DL models applied for treelike tubular structure segmentation are trained under supervision.

**Figure 17:** Deep learning based segmentation.

#### 3.3.1. Classical CNNs

Convolutional neural networks (CNNs) are a class of DL models that use convolution operations in at least one of their layers. The convolutional layers are inspired by the receptive field structure found in the animal’s primary visual cortex. Hence, they are specifically designed to perform image processing tasks such as classification or pixel-wise segmentation.

Classical CNNs are typically comprised of multiple convolutional and pooling layers, followed by a set of fully connected layers. The convolutional layer extracts features from inputs by convolution operation, in which a convolution kernel slides throughout the input map, multiplying elementwise with a specified region called the receptive field.
### Table 2
Machine learning based segmentation methods

| Author                  | Year | Region | Modality | Method            | Validation Dataset (number of cases) | Metrics | Method Class |
|------------------------|------|--------|----------|-------------------|--------------------------------------|---------|--------------|
| Hassouna et al. [46]   | 2006 | BV     | MRA      | GMM               | TOF MRA scans (n/a)                  | Visual  | Unsupervised |
| Oliveira et al. [47]   | 2011 | LV     | CT       | GMM-RG            | LD CT scans (15)                     | Accuracy 0.9519±0.005 |
| Roychowdhury et al. [48] | 2014 | RV     | CFP      | Feature-based GMM | DRIVE (40)                           | Accuracy 0.983±0.0071 |
|                        |      |        |          |                   | STARE (20)                           | AUROC 0.962 |
|                        |      |        |          |                   | CHASE-DB1 (28)                       | Accuracy 0.953±0.005 |
|                        |      |        |          |                   |                                      | Accuracy 0.9532 |
| Kande et al. [49]      | 2010 | RV     | CFP      | SWFCM             | DRIVE (40)                           | Accuracy 0.8911 |
|                        |      |        |          |                   | STARE (20)                           | AUROC 0.9518 |
| Goceri et al. [50]     | 2017 | LV     | MRI      | KMC               | MRI scans (14)                        | Accuracy 0.896 |
| Cinsdikici & Aydin [51]| 2009 | RV     | CFP      | ACO               | DRIVE (40)                           | Accuracy 0.9203 |
| Hassanien et al. [52]  | 2015 | RV     | CFP      | ABC               | DRIVE (40)                           | Accuracy 0.9388 |
|                        |      |        |          |                   |                                      | Accuracy 0.9772 |
| Hassan & Hassanien [53]| 2018 | RV     | CFP      | WOA               | DRIVE (40)                           | Accuracy 0.8918 |
|                       |      |        |          |                   |                                      | Accuracy 0.983 |
| Ramos-Soto et al. [54] | 2021 | RV     | CFP      | MCET-HHO          | DRIVE (40)                           | Accuracy 0.9667 |
|                       |      |        |          |                   | STARE (20)                           | AUROC 0.9820 |
|                       |      |        |          |                   |                                      | Accuracy 0.9836 |
|                       |      |        |          |                   |                                      | Accuracy 0.9580 |
| Staal et al. [55]      | 2004 | RV     | CFP      | KNN-PBM           | DRIVE (40)                           | Accuracy 0.9441 |
|                       |      |        |          |                   |                                      | AUROC 0.9520 |
| Lo et al. [56]         | 2010 | LA     | CT       | KNN-RG            | STARE (20)                           | Accuracy 0.8958 |
|                       |      |        |          |                   |                                      | AUROC 0.9614 |
|                       |      |        |          |                   |                                      | F1 0.0011 |
|                       |      |        |          |                   |                                      | BD 0.598 |
|                       |      |        |          |                   |                                      | LE 1184 mm |
|                       |      |        |          |                   |                                      | ILD 0.540 |
|                       |      |        |          |                   |                                      | ACC 0.9509 |
|                       |      |        |          |                   |                                      | AUC 0.9633 |
| Ricci & Perfetti [57]  | 2007 | RV     | CFP      | SVM-LD            | DRIVE (40)                           | Accuracy 0.9646 |
|                       |      |        |          |                   | STARE (20)                           | AUROC 0.9680 |
|                       |      |        |          |                   |                                      | ACC 0.783 |
|                       |      |        |          |                   |                                      | FPR 0.10 |
|                       |      |        |          |                   |                                      | BD 122.2 |
|                       |      |        |          |                   |                                      | LE 63.14 |
|                       |      |        |          |                   |                                      | ILD 4.084 |
| Meng et al. [58]       | 2017 | LA     | CT       | SVM-GC            | SD CT scans (50)                     | Accuracy 0.9534 |
|                       |      |        |          |                   |                                      | Accuracy 0.9763 |
|                       |      |        |          |                   |                                      | Pre 0.7956 |
|                       |      |        |          |                   |                                      | AUC 0.9768 |
|                       |      |        |          |                   |                                      | Accuracy 0.9480 |
|                       |      |        |          |                   |                                      | Accuracy 0.9671 |
|                       |      |        |          |                   |                                      | Accuracy 0.9466 |
|                       |      |        |          |                   |                                      | Accuracy 0.9614 |
|                       |      |        |          |                   |                                      | FPR 0.10 |
|                       |      |        |          |                   |                                      | BD 122.2 |
|                       |      |        |          |                   |                                      | LE 63.14 |
|                       |      |        |          |                   |                                      | ILD 4.084 |
|                       |      |        |          |                   |                                      | ACC 0.783 |
|                       |      |        |          |                   |                                      | FPR 0.10 |
|                       |      |        |          |                   |                                      | BD 122.2 |
|                       |      |        |          |                   |                                      | LE 63.14 |
|                       |      |        |          |                   |                                      | ILD 4.084 |
|                       |      |        |          |                   |                                      | ACC 0.783 |
|                       |      |        |          |                   |                                      | FPR 0.10 |
|                       |      |        |          |                   |                                      | BD 122.2 |
|                       |      |        |          |                   |                                      | LE 63.14 |
|                       |      |        |          |                   |                                      | ILD 4.084 |
|                       |      |        |          |                   |                                      | ACC 0.783 |
|                       |      |        |          |                   |                                      | FPR 0.10 |
|                       |      |        |          |                   |                                      | BD 122.2 |
|                       |      |        |          |                   |                                      | LE 63.14 |
|                       |      |        |          |                   |                                      | ILD 4.084 |
|                       |      |        |          |                   |                                      | ACC 0.783 |
|                       |      |        |          |                   |                                      | FPR 0.10 |
|                       |      |        |          |                   |                                      | BD 122.2 |
|                       |      |        |          |                   |                                      | LE 63.14 |
|                       |      |        |          |                   |                                      | ILD 4.084 |
|                       |      |        |          |                   |                                      | ACC 0.783 |
|                       |      |        |          |                   |                                      | FPR 0.10 |
|                       |      |        |          |                   |                                      | BD 122.2 |
|                       |      |        |          |                   |                                      | LE 63.14 |
|                       |      |        |          |                   |                                      | ILD 4.084 |

H. Li, Z. Tang, Y. Nan, and G. Yang: Preprint submitted to Elsevier
The regions include the retinal vessel (RV), lung airways (LA), brain vessel (BV), coronary vessel (CV), and thoracic aorta (TA). All unnamed datasets are private and may have specific modalities such as phase-contrast (PC), low-dose (LD), standard-dose (SD), or time-of-flight (TOF). All metrics are shown as mean ± standard deviation or mean only.

Then, the activation function maps inputs non-linearly to increase the visual representation ability of convolutional layers. The pooling layers are usually employed after to reduce feature position dependency and thus improve CNNs’ robustness. Besides, down-sampling along with pooling can reduce computational costs. Fully connected layers are the last block that produces a specialized semantic output, for example, foreground or background, from the feature map.

Charbonnier et al. [63] presented a 2D CNN-based leakage detection approach that treats the removal of segmentation leakages as an independent classification method to improve the quality of tracheal segmentation on thoracic CT. Three 2D image patches were taken from each airway candidate, corresponding to the top, middle, and bottom layers. The airway candidate is then classified by passing three patches through three parallel stacks of convolutional and max-pooling layers separately and combining FC layers of three stacks in a final softmax layer. However, this method does not thoroughly analyse the 3D continuity of airway candidates and treelike structure. Similarly, Yun et al. [64] presented a voxel-by-voxel airway segmentation approach for chest CT utilizing a 2.5D CNN (a pseudo-3D CNN) which takes three adjacent slices in each orthogonal direction (axial, sagittal, and coronal) at each voxel.

UNet: U-Net is a CNN designed for medical image segmentation [110], which has a symmetric encoder-decoder architecture with skips connections between the encoding (down-sampling) and decoding (up-sampling) pathways. U-Net can process the entire image in one forward pass through its down-sampling/up-sampling architecture, resulting in a direct and faster segmentation than typical pixel-wise CNNs. Besides, using skip connections, UNets improve the accuracy of prediction by combining a low-level feature map that provides local information and a high-level feature map that represents global information. Based on the U-Net architecture, Meyer et al. [65] proposed a retinal vessel segmentation method from Scanning Laser Ophthalmoscopy (SLO) CFP images. The 3D U-Net [111] based method for airway segmentation is presented in [66].

Jin et al. [67] introduced a method capable of high-quality airways segmentation from incompletely labelled training datasets based on a shallow 3D U-Net. This approach reduces the previous deeper 3D U-Net to only two pooling layers to preserve small airways at distal sites and potentially benefit the learning of 3D structure variations. A domain-specific sampling scheme is applied to strategically select annotations from a highly-specific and moderately-sensitive conservative segmentation method [112]. In addition, fuzzy-connectedness RG and skeletonization-guided leakage removal are implemented to address local discontinuities in the output. Similarly, Garcia-Uceda et al. [68] proposed an optimized and efficient approach for segmenting the airways based on a simple and low-memory 3D U-Net architecture. This approach employs unpadded convolutions in the first three resolution levels of the U-Net, which provides a 30% reduction in memory footprint. Compared to the previous U-Net based airway segmentation algorithms, the proposed method can process larger 3D image patches, usually covering the entire lung, in a single run, resulting in an improved outcome and a faster converging of the training.

ResNet: The residual neural network (ResNet) is a CNN that utilizes shortcuts that skip one or more layers. Adding skip connections solves the problem of gradient disappearance and avoids the increasing training error resulting from stacking more layers for higher discriminatory ability. Based on a multi-pathway ResNet, a robust 3D hepatic vessel segmentation method is proposed by Kitrungrotsakul et al. [69]. The proposed network structure consists of three parallel DenseNets[113], i.e., extensions of ResNet, for extracting data in the sagittal, coronal, and transverse planes, which are concatenated in the FC layer to predict the class of central voxels. In addition, a multi-scale method [114] transforms raw images into vessel probability maps as network inputs to enhance robustness.

Generative adversarial network (GAN): Generative adversarial networks (GANs) [115] frame the segmentation task as a supervised learning problem with two components: a generator (G), which learns the mapping from a random input z with density p_z to x_g = G(z), and a discriminator (D), which either takes the ground truth x with density p_{gt} or the generated data x_g, and outputs the probability D(x) of x to be true. These two components are trained adversarially to drive the generator to generate more accurate segmentation labels that can fool the discriminator. In [90], a retinal vessel segmentation method is developed based on the GAN. The proposed GAN takes advantage of U-Net, combining high-resolution local features with up-sampled high-level features. Besides, the dilated convolutions employed by GAN effectively expand the receptive field to capture large-scale morphological vessel properties without raising the number of computations required.

3.3.2. Advanced architectures

In addition to the basic CNN models, a variety of new DL architectures are developed by applying or integrating one or more advanced modules including holistically-nested edge detector (HED) [70], graph neural network (GNN) [71, 83], mean-field network (MFN) [83], recurrent convolutional layer (RCL) [72], WingsNet [73], attention mechanism (AM) [74, 75, 93, 87], atrous convolution (AC) [77], back-propagation neural network (BPNN) [78], multi-scale feature aggregation (MFA) [79, 93], Inception [91], and multi-kernel pooling (MKP) [92]. Since the mentioned architec-
tural improvements are also present in other types of methods, in this subsection we only detail the methods based solely on advanced modules first.

**DSSRN**: Lin et al. [70] proposed a deeply supervised retinal vessel segmentation network, DSSRN, based on the holistically-nested edge detector (HED), incorporating the conditional random fields (CRFs). The proposed network begins with the HED architecture, a five-block single-stream network with multiple convolutional and ReLU layers in each block. The final predictions are formed by combining multi-scale side outputs from each hidden layer, eliminating the edge or boundary detection ambiguity. CRF provides global smoothness regularization into the system for robustness.

**UNet and GNN (UNet-GNN)**: Garcia-Uceda Juarez et al. [71] proposed an airway tree segmentation approach for chest CTs that combines a 3D UNet with graph neural network (GNN) based modules [116]. This method substitutes a GNN-based module for the two deepest convolutional layers in the baseline UNet. The introduction of graph convolutions provides nodes with corresponding feature vectors containing information about graph neighbourhood connectivity. The connectivity of nodes is described by the adjacency matrix with learnable GNN filter weights.

**UNet with RCL (UNet-RCL)**: In [72], Wang et al. incorporated 3D slice-by-slice convolutional layers called recurrent convolutional layers (RCLs) in a 3D UNet architecture for airway segmentation. Different from the conventional layer-by-layer CNN, the in-layer recurrent convolutions make the message-passing between neurons more efficient in the same layer, which helps in capturing the spatial information of elongated structures.

**WingsNet**: Zheng et al. [73] argued that the poor performance of CNN on segmenting distal small airways is attributed to the gradient erosion and dilation of the neighbourhood voxels, which is caused by the imbalance between foreground and background regions. Therefore, they proposed WingsNet with group supervision to boost the training of shallow layers. Moreover, they designed a new loss called General Union Loss to address the imbalance between large and small airways.

**UNet with AM (UNet-AM)**: In [74], Qin et al. presented a 3D-UNet based airway segmentation method with AMs that is highly sensitive to bronchioles in the periphery. The method initially employs a feature recalibration module based on [117, 118] to make the greatest use of learned features. Moreover, inspired by [119, 120], an attention distillation module is introduced to promote specific structure and distribution pattern learning of airways to improve the ability to distinguish small bronchioles. The proposed method retrieved significantly more branches while maintaining comparable overall segmentation accuracy. A similar method is applied in [75] for both airway and artery-vein segmentation on non-contrast computed tomography (CT).

**UNet with HRMF**: Fan et al. [76] proposed an unsupervised UNet with hidden random field to segment brain vessels in TOF MRA. The method first uses a HRMF module to classify each TOF-MRA image into three classes, background, brain tissue and blood vessels. Then the 3-D UNet takes in the output from HRMF module and segments brain vessels.

**TACNet**: Cheng et al. [77] presented the Tiny Atrous Convolutional Network (TACNet) to account for multi-scale changes in the airway. The principal convolution operation employed in the TACNet is the atrous convolution (AC), which enlarges the receptive field and captures multi-scale context information. In a conventional UNet, the feature map is downsampled numerous times, leading tiny bronchi to vanish from the feature map or become undetectable. The ACs maintain the original resolution between layers, ensuring that the feature map does not reduce in size while increasing the receptive field.

**Back-propagation neural network (BPNN)**: A back-propagation neural network (BPNN) for retinal vascular segmentation is proposed in [78] to increase the accuracy of retinal vessel segmentation. Four different green channel enhancement results are utilized to create 4D feature vectors: adaptive histogram equalization, morphological processing, Gaussian matched filtering, and Hessian matrix filtering. The feature vectors are fed into a typical three-layer BP neural network to segment blood vessels.

**MFA-Net**: In [79], a 3D multi-scale feature aggregation network (MFA-Net) is presented against the scale difference of various structures in airway tree segmentation. The MFA blocks apply parallel 3D atrous convolutions of different rates to capture multi-scale context information, which improves the sensitivity of MFA-Net to small bronchi segmentation to solve local discontinuities.

### 3.3.3. Task-based/Hybrid approaches

According to the morphological characteristics of human treelike tubular structures, the segmentation task of DL networks can be converted into voxel connectivity analysis [80], graph refinement [83], or bifurcation point detection (BPD) to support conventional RG [82].

Multi-task segmentation were implemented through cascaded or hybrid modules, including separate segmentation of vertical and horizontal tubular structures [84] or thick and thin tubular structures [85], and coarse segmentation followed by refinement [87].

Another strategy is to combine DL-based segmentation tasks with other tasks to improve performance using additional information, such as centreline tracking (CLT) [86, 88].

In addition, adversarial learning was also attempted, i.e., to constrain the segmentation network with discriminators [90, 91, 92, 93].

**AirwayNet**: Qin et al. [80] present the AirwayNet for accurate airway segmentation that focuses on the voxel con-
nectivity of airways. After modelling the voxel 3D connectivity using a 26-connected neighbour analysis, the binary segmentation task is converted into 26 tasks to determine whether the voxel is connected to its neighbours along a specific orientation. Therefore, the AirwayNet learns relationships between neighbouring airway voxels to distinguish the airway tree from the background. In addition, the coordinates of voxels and their distance from lung borders are fed into the AirwayNet as supplementary semantic information to take full advantage of context knowledge. The final airway tree is generated from the predicted connectivity map using fuzzy connectedness rules [67].

**DeepVesselNet:** Tetteh et al. [80] present the DeepVesselNet based on FCN architecture for accurate 3-D angiographic feature extraction. First, they designed a 2-D orthogonal cross-hair filters to relieve computational burden. Second, a class balancing cross-entropy loss function is introduced to reduce the high false-positive rate. Finally, they generated a synthetic dataset and used this for transfer learning.

**Bifurcation point detection and RG (BPD-RG):** Wang et al. [82] proposed an automatic airway tree labelling system using ResNets to detect bifurcation points. A hybrid of three different ResNets is constructed with cascaded residual blocks that enable the learning of large-scale spatial information. The three ResNets are trained to predict four critical bifurcation points of airway trees independently, and their predictions are fused to reduce the variance. The complete airway tree is then reconstructed using an adaptive RG approach, guided by the detected bifurcation points.

**MFN and GNN (MFN-GNN):** Selvan et al. [83] extracted airway trees from CT scans by first constructing a graph-based representation of the image data and then posing the tree extraction task as a graph refinement task using mean-field network (MFN) and GNN. In the first model, refinement is framed as an approximate Bayesian inference problem solved by mean-field approximation (MFA) [121, 122] to approximate the posterior density over various subgraphs. The optimal subgraph is represented by parameterized nodes and pair-wise probabilities. Performing the MFA updating iterations as layers in a feed-forward network enables parameter learning. The second GNN model [123] is designed as an edge connection predictor that learns edge embeddings from an over-complete graph to produce an edge probability map.

**2D+3D NN:** Zhao et al. [84] presented a 2-stage airway segmentation method based on a 2D+3D neural network (NN) and linear programming (LP) based tracking algorithm. The 2D and 3D NNs were developed to segment horizontal airways and tube-like vertical airways separately due to significant appearance differences. The 3D NN structure resembles the 3D UNet structure, and the 2D NN is a standard fully convolutional network (FCN) [124]. The first stage targets the trachea and main bronchus, whereas the second stage targets the distal bronchus. The outputs from the two stages are then connected using an LP-based object tracking algorithm, which simultaneously filters out false positives.

**Hybrid UNet:** A hybrid UNet-based approach for retinal vessel segmentation is proposed in [85], consisting of a multi-task segmentation network and a fusion network. Since the signal-to-noise ratio and contrast of thin vessels are lower than those of thick vessels, the segmentation network is built to segment thick and thin vessels separately from fundus images. The encoder extracts features corresponding to both thick and thin vessels and passes them to the two independent decoders. Afterwards, a fusion network is designed to fuse these two segmentation results to acquire the final blood vessel tree.

**UNet with CE (UNet-CE):** Meng et al. [86] developed an airway segmentation method that incorporates the 3D UNet into an adaptive volume of interest (VOI) based tracking scheme. The parent VOI is first assigned based on the coarsely segmented airway tree using the RG algorithm. The 3D extension of the UNet proposed in [110] is employed to segment the airway region inside each VOI. Meanwhile, the centreline of the bronchi is retrieved by applying the gradient vector flow (GVF) magnitude and the tubular-likeness function based on the GVF. After detecting the branching point from the centreline, child VOI/VOIs can be assigned adaptively depending on the size and the travelling orientation of the bronchi.

**CAG-Net:** Li et al. [87] presented a novel Cascaded Attention Guided Network (CAG-Net) for accurate retinal vascular segmentation. The CAG-Net consists of prediction and refinement modules, using Attention UNet++ (AUNet++) as the basic architecture. Because the simple 1x1 convolutional layer in [125] cannot fully fuse the decoder’s multi-scale information, attention-guided convolutional blocks (AC blocks) are employed instead to fuse the multi-scale features adaptively. In the prediction module, the AUNet++ generates an initial segmentation map that is then concatenated with the original image to create a new volume as the input of the refinement module. The final segment is formed by concatenating the outputs of the two modules in an FC layer.

**UNet with CE and LR (UNet-CE-LR):** Nadeem et al. [88] integrated another tracking algorithm called freeze-and-grow (FG) with 3D UNet for airway segmentation. The FG algorithm is a traditional CT intensity-based segmentation approach that begins with conservative segmentation parameters and then iteratively captures finer details by freezing leakage-roots and relaxing parameters. In each iteration, the centreline and possible leakages are identified for propagation. A modified 3D UNet was applied to derive probability maps of the airway as inputs from the FG algorithm. Finally, a multi-task 3D UNet that classifies each pixel as airway, leak, or background is employed to remove distal leakages.

**DD-CNN:** Zhang et al. [89] employed a semi-supervised mixture probability model to fit the cerebrovascular intensity distribution from the sparse manual annotations and
generate massive labeled pixels. They also proposed a dilated dense(DD) CNN which consists of DD blocks that use both high-level and low-level feature maps. The DD-CNN is trained on the previously generated labels.

**GAN and dense UNet (GAN-DUNet):** In [91], a dense UNet employing the Inception module is combined with GAN for accurate retinal vessel segmentation. First, the skip connections in a standard UNet are substituted with dense blocks to enable full fusing of features from shallow to deep layers, hence improving accuracy without increasing the network’s depth or width. The Inception module is then utilized to replace conventional convolutional layers to expand the receptive field and extract multi-scale vessel features from different-sized convolution kernels. Finally, a GAN is adopted in the training, where the dense UNet is considered as the generator and the loss value is backpropagated by a multilayer neural network as the discriminator.

**M-GAN:** Park et al. [92] presented a novel conditional GAN called M-GAN to perform precise retinal vessel segmentation by balancing losses of stacked deep FCNs. Based on a conditional GAN [126], the M-GAN comprises a newly developed M-generator for more robust segmentation and a deeper M-discriminator for more efficient GAN model training. The M-generator consists of two stacked deep FCNs with short-term skip connections and long-term residual connections, as well as a multi-kernel pooling (MKP) block between two FCNs that supports scale-invariance of vessel segmentation of various sizes. In addition, the Lanczos resampling method [127] is used to smooth out the segmented vessel branching and eliminate false.

**SEGAN:** Zhou et al. [93] proposed a symmetric equilibrium GAN (SEGAN) with multi-scale features refine blocks (MSFRBs) and AMs to improve retinal vessel segmentation. SEGAN creates a symmetric adversarial architecture that eliminates the imbalance in generator-discriminator capabilities to drive the generator to produce more realistic and detailed images. Second, MSFRB is designed to optimize feature merging while retaining high-resolution and high-semantic information. In addition, AMs make the network focus on distinguishing features rather than irrelevant information. These improvements allow the proposed network to excel in extracting detailed information by maximizing the multi-scale feature representation.

### 3.3.4. Training strategy based

The training strategy of a model determines how effectively and efficiently it learns, and therefore also has a significant impact on its segmentation performance. Hence, some segmentation methods for treelike tubular structures focus on training strategies including transfer learning (TL) [94] and cascaded training [95].

**FCN with TL (FCN-TL):** In [94], a supervised retinal vessel segmentation method is proposed based on a pre-trained FCN [128]. The new training dataset was utilized to fine-tune the pre-trained FCN. Because the semantic segmentation result of retinal vessels is wider than the ground truth, and noises are added in slice merging and overlapping, additional unsupervised post-processing strategies using local Otsu’s method are applied. This paper proves the utility of transfer learning when using deep learning techniques in medical imaging.

**OFF-eNET:** Nazir et al. [95] trained a three-stage Optimally Fused Fully end-to-end Network (OFF-eNET) in a cascade fashion for 3D cerebral vessel segmentation. The first stage employs up-skip connections to improve information flow between the encoder and decoder and dilated convolution to preserve the spatial resolution feature map designed for thin vessels. In the second stage, residual mapping with the inception module is utilized for faster converging and learning richer representations. The cascaded training strategy is applied in the third stage to progressively achieve concrete segmentation results over three sub-stages (basic, complete, and enhanced) by utilizing transferred knowledge. During the cascaded training, the knowledge learned by the basic sub-stage is transferred to the complete sub-stage for initiation, and the enhanced sub-stage is fine-tuned based on the well-trained complete stage.

### 3.3.5. Loss functions

The commonly used loss functions are based on cross-entropy or Dice Similarity Coefficient (DSC). Since the tree-like structures occupy only a small portion of the images due to their sparsity, the background pixels are usually dominant. Therefore, the global class-balancing weight is applied to the binary cross-entropy (BCE) loss function [67, 66, 86] as follows:

\[
\mathcal{L}_{\text{BCE}} = -w \sum_{x=N_f} \log p(x) - (1-w) \sum_{x=N_b} \log (1-p(x)),
\]

where \( p \) is the pixel-wise foreground probability. \( N_f \) and \( N_b \) are the number of foreground pixels and the number of background pixels, respectively. The weight \( w \) is usually equal to \( \frac{N_b}{N} \), where \( N \) is the total number of pixels \( x \).

As for the Dice loss function, a tolerance term \( \epsilon \) is commonly introduced [66, 71] to avoid dividing by zero in cases where there is no foreground pixel as defined below:

\[
\mathcal{L}_{\text{Dice}} = -\frac{2 \sum_{x=N} \log p(x)g(x)}{\sum_{x=N} \log p(x) + \sum_{x=N} \log g(x) + \epsilon},
\]

where \( g \) is the foreground ground-truth.

Furthermore, [74, 75] combined the Dice loss with the Focal loss [129] which reduces the contribution of easy examples in training and more focuses on the hard example, defined by:
\[ \mathcal{L}_{\text{DiceFocal}} = - \frac{2}{\sum_{x=N} \log p(x) g(x)} \frac{2 \sum_{x=N} \log p(x) + \sum_{x=N} \log g(x) + e}{N} \sum_{x=N} (1 - p_g(x))^2 \log p_g(x), \]  

(11)

where \( p_g(x) = p(x) \) if \( g(x) = 1 \). Otherwise, \( p_g(x) = 1 - p(x) \).

Wang et al. [72] proposed a novel loss function called radial distance loss that focuses on spatial consistency.

\[ \mathcal{L}_{\text{RDL}} = - \frac{1}{2} \sum_{k=0}^{1} W_k \frac{2 \sum_{N} \hat{p}_{i,k} d_{i,k}}{\sum_{i} \hat{p}_{i,k}^2 + \sum_{i} d_{i,k}^2} \]  

(12)

where \( \hat{p}_{i} \) denotes the \( i^{th} \) voxel predicted binary result, \( d_i \) denotes the radial distance map which is defined as:

\[ D = -\frac{1}{\max(F)} \| F \|_1 + 1 \]  

(13)

where \( F \) denotes the Euclidean distance map from the ground truth centerline.

In addition to the least square loss [130] that is utilized as GAN loss and standard binary cross-entropy (BCE) loss, the generator in [92] adopts a newly designed false-negative (FN) loss function, defined as

\[ \mathcal{L}_{\text{FN}} = \frac{1}{N_f} \sum_{x=N_f} (1 - p_{FN}(x))^2, \]  

(14)

where

\[ p_{FN}(x) = \begin{cases} 1, & \text{if } p(x) \geq 0.5 \\ p(x), & \text{otherwise} \end{cases} \]  

(15)

The inclusion of the FN loss function has been shown to reduce the FNR.

### 3.4. Preprocessing

Preprocessing techniques are commonly employed as the first steps of segmentation methods, which involve intensity normalization, noise reduction, and contrast enhancement. Different combinations of these techniques must be used according to variable imaging modalities, conditions, and targets.

The intensity normalization can reduce the variance of the image data caused by different imaging conditions and improve the computational efficiency using uniform intensity scales. Many segmentation methods preprocessed the images to a standard intensity scale using uniformity transformations. For deep learning methods, rescaling intensities also promotes a faster convergence in the optimization stage.

Besides, spacing resampling is often adopted to standardize the 3D image data.

After data normalization, outliers and noise can be eliminated using the intensity histogram or histogram-based function such as LogSig [51]. Filter-based techniques such as anisotropic diffusion [131] are also applied by some methods to reduce the noise. As for the image with nonuniform lighting conditions, filters calculating the local mean can be used for brightness correction [52].

Many techniques are available for tubular structure enhancement and are usually followed by data re-normalization. Nonlinear intensity transformations such as squaring intensity value [48] and gamma enhancement [51] can directly improve the contrast. Mathematical morphology operations, including Top-hat transform, can extract small tubular structures using structuring elements that fit the largest diameter [48]. Using a filter like the unsharp mask filter can significantly reduce the noise and enhance the walls of tubular structures [58]. For Coloured Fundus Photography (CFP) images, the green channel shows the highest contrast between vessel and background [21]. Hence, most methods only adopt and convert the green channel into greyscale. However, matching histograms of the green channel and the red channel can further enhance this contrast [49].

Furthermore, most deep learning methods adopt typical data augmentation techniques in order to manually enlarge the training dataset to improve its robustness, including rotation, mirroring, scaling, cropping, rigid transformations, and elastic deformations.

### 3.5. Postprocessing

A postprocessing step may be applied after the segmentation to rejoin disconnected segments or to remove small, segmented regions usually corresponding to noise or abnormalities in the image.

The connected component analysis (CCA) [132] is a common postprocessing technique to isolate individual components of the segmentation output using connected neighbourhood and label propagation. In order to filter unwanted and disconnected small segments under a threshold, some methods define a specific number of pixels [51, 49, 77, 68] and some use a thinness ratio [52]. Besides, some methods have developed new algorithms to refine the output using information from neighbour pixels [69]. Moreover, the morphological reconstruction is also adopted in the postprocessing to refine the output [26].

For the CNN-based method, the border effects due to zero-paddings make the output less reliable toward the boundaries [68]. Hence, a quadratic polynomial decrease towards the borders or a Gaussian-distributed weight map can be applied to the probability map, accompanied by an overlapping slide-window sampling strategy. The final segmentation can be obtained by placing together all patches and averaging the overlapped regions.
### Table 3
Deep learning based segmentation methods

| Author et al. | Year | Region | Modality | Method | Validation Dataset (number of cases) | Metrics | Method Class |
|---------------|------|--------|----------|--------|--------------------------------------|---------|--------------|
| Charbonnier et al. [63] | 2016 | LA | CT | CNN | EXACT’09 (20) | FPR: 0.0101 | Classical CNNs |
| Yun et al. [64] | 2018 | LA | CT | CNN | EXACT’09 (20) | FPR: 0.0456±0.0373 | | |
| Meyer et al. [65] | 2017 | RV | SLO | CFP | Unet | IOSTAR (10) | FPR: 0.0274 | | |
| Charbonnier et al. [63] | 2016 | LA | CT | CNN | EXACT’09 (20) | TLD: 0.518 | Classical CNNs |
| Yun et al. [64] | 2018 | LA | CT | CNN | EXACT’09 (20) | ±0.0373 | | |
| Meyer et al. [65] | 2017 | RV | SLO | CFP | Unet | IOSTAR (10) | ±0.0011 | | |
| Garcia-Uceda Juarez et al. [66] | 2018 | LA | CT | UNet | CT scans (6) | DSC: 0.8 | |
| Jin et al. [67] | 2017 | LA | CT | UNet | EXACT’09 (20) | DSC: 0.916 | Comparative |
| Garcia-Uceda et al. [68] | 2021 | LA | CT | UNet | CF-CT (24) | DSC: 0.876 | Advanced architectures |
| Kitrungrotsakul et al. [69] | 2019 | LV | MRI | ResNet | IRCAD (20) | DSC: 0.876 | |
| Lin et al. [70] | 2019 | RV | CFP | DSSRN | DRIVE (40) | DSC: 0.887 | |
| Garcia-Uceda Juarez et al. [71] | 2019 | LA | CT | UNet-GNN | LD CT scans (12) | DSC: 0.89 | |
| Wang et al. [72] | 2019 | LA | CT | UNet-RCL | SD CT scans (38) | DSC: 0.865±0.010 | |
| Zheng et al. [73] | 2020 | LA | CT | WingsNet | EXACT’09 and LIDC-IDRI (20+70) | DSC: 0.914±0.045 | |
| Qin et al. [74] | 2020 | LA | CT | UNet-AM | EXACT’09 and LIDC-IDRI (20+70) | DSC: 0.987±0.069 | |
| Qin et al. [75] | 2021 | LA | CT | UNet-AM | EXACT’09 and LIDC-IDRI (20+70) | DSC: 0.987±0.069 | |
| Fan et al. [76] | 2020 | BV | TOF-MRA | UNet-HMRF | TOF-MRA (100) | DSC: 0.987±0.069 | |
| Cheng et al. [77] | 2021 | LA | CT | TACNet | CT scans (100) | DSC: 0.987±0.069 | |
| Tang et al. [78] | 2021 | RV | CFP | BPNN | DRIVE (40) | DSC: 0.987±0.069 | |
| Zhou et al. [79] | 2021 | LA | CT | MFA-Net | CT scans (150) | DSC: 0.987±0.069 | |
| Author               | Year | Region | Modality | Method      | Validation Dataset | Metrics                     | Method Class |
|----------------------|------|--------|----------|-------------|--------------------|-----------------------------|--------------|
| Qin et al. [80]      | 2019 | LA     | CT       | AirwayNet   | CT scans (10)      | Se: 0.847±0.049             | Task-based/Hybrid |
| Tetteh et al. [81]   | 2020 | BV     | MRA      | DeepVesselNet | TOF-MRA (40)      | Pr: 0.011±0.008             |
| Wang et al. [82]     | 2020 | LA     | CT       | BPD-RG      | LUNA16 (345)       | Pr: 0.902±0.028             |
| Selvan et al.* [83]  | 2020 | LA     | CT       | MFN-GNN     | LD CT scans (32)   | Pr: 0.8668                  |
| Zhao et al. [84]     | 2019 | LA     | CT       | 2D+3D NN    | CT scans (22)      | Pr: 0.9785                  |
| Yang et al. [85]     | 2021 | RV     | CFP      | Hybrid UNet | DRIVE (40)         | Pr: 0.875                   |
| Meng et al. [86]     | 2017 | LA     | CT       | UNet-CE     | SD CT scans (50)   | Pr: 0.876                   |
| Li et al.* [87]      | 2020 | RV     | CFP      | CAG-Net     | DRIVE (40)         | Pr: 0.976                   |
| Nadeem et al. [88]   | 2021 | LA     | CT       | UNet-CE-LR  | SD CT scans (120)  | Pr: 0.976                   |
| Zhang et al. [89]    | 2019 | BV     | TOF-MRA  | DD-CNN      | MIDAS21            | Pr: 0.976                   |
| Zhao et al. [90]     | 2020 | RV     | CFP      | GAN         | DRIVE (40)         | Pr: 0.976                   |
| Park et al.* [92]    | 2020 | RV     | CFP      | M-GAN       | DRIVE (40)         | Pr: 0.976                   |

H. Li, Z. Tang, Y. Nan, and G. Yang: *Preprint submitted to Elsevier*
### Table - Human Treelike Tubular Structure Segmentation

| Author          | Year | Region | Modality | Method     | Validation Dataset (number of cases) | Metrics       | Method Class   |
|-----------------|------|--------|----------|------------|--------------------------------------|---------------|----------------|
| Guo et al.*     | 2020 | RV     | CFP      | GAN-DUNet  | DRIVE (40)                           | Acc: 0.9542   | Task-based/Hybrid |
|                 |      |        |          |            |                                      | Sens: 0.8283  |                |
|                 |      |        |          |            |                                      | Spec: 0.9726  |                |
|                 |      |        |          |            |                                      | DSC: 0.8215   |                |
|                 |      |        |          |            |                                      | AUCROC: 0.9772|                |
| Zhou et al.*    | 2021 | RV     | CFP      | SEGAN      | DRIVE (40)                           | Acc: 0.9563   | Task-based/Hybrid |
|                 |      |        |          |            |                                      | Sens: 0.8294  |                |
|                 |      |        |          |            |                                      | Spec: 0.9812  |                |
|                 |      |        |          |            |                                      | Pre: 0.8397   |                |
|                 |      |        |          |            |                                      | DSC: 0.8345   |                |
|                 |      |        |          |            |                                      | AUCROC: 0.9021|                |
|                 |      |        |          |            | STARE (20)                           | Acc: 0.9671   |                |
|                 |      |        |          |            |                                      | Sens: 0.8812  |                |
|                 |      |        |          |            |                                      | Spec: 0.9781  |                |
|                 |      |        |          |            |                                      | Pre: 0.7952   |                |
|                 |      |        |          |            |                                      | DSC: 0.8359   |                |
|                 |      |        |          |            |                                      | AUCROC: 0.9263|                |
|                 |      |        |          |            | CHASE-DB1 (28)                      | Acc: 0.9630   |                |
|                 |      |        |          |            |                                      | Sens: 0.8435  |                |
|                 |      |        |          |            |                                      | Spec: 0.9783  |                |
|                 |      |        |          |            |                                      | Pre: 0.8013   |                |
|                 |      |        |          |            |                                      | DSC: 0.8218   |                |
|                 |      |        |          |            |                                      | AUCROC: 0.9083|                |
|                 |      |        |          |            | HRF (45)                             | Acc: 0.9559   |                |
|                 |      |        |          |            |                                      | Sens: 0.8310  |                |
|                 |      |        |          |            |                                      | Spec: 0.9730  |                |
|                 |      |        |          |            |                                      | Pre: 0.8115   |                |
|                 |      |        |          |            |                                      | DSC: 0.8211   |                |
|                 |      |        |          |            |                                      | AUCROC: 0.8992|                |
| Jiang et al.    | 2018 | RV     | CFP      | FCN-TL     | DRIVE (40)                           | Acc: 0.9593±0.0073 |                |
|                 |      |        |          |            |                                      | Sens: 0.7121±0.0546 | Training strategy based |
|                 |      |        |          |            |                                      | Spec: 0.9832±0.0069 |                |
|                 |      |        |          |            |                                      | Pre: 0.9580   |                |
|                 |      |        |          |            | STARE (20)                           | Acc: 0.9653±0.0107 |                |
|                 |      |        |          |            |                                      | Sens: 0.7820±0.1085 |                |
|                 |      |        |          |            |                                      | Spec: 0.9798±0.0117 |                |
|                 |      |        |          |            |                                      | Pre: 0.9580   |                |
|                 |      |        |          |            | CHASE-DB1 (28)                      | Acc: 0.9951±0.0065 |                |
|                 |      |        |          |            |                                      | Sens: 0.7217±0.0821 |                |
|                 |      |        |          |            |                                      | Spec: 0.9770±0.0037 |                |
|                 |      |        |          |            |                                      | Pre: 0.9580   |                |
|                 |      |        |          |            |                                      | Sens: 0.7686±0.0378 |                |
|                 |      |        |          |            |                                      | Spec: 0.9682±0.0057 |                |
|                 |      |        |          |            |                                      | Pre: 0.9770   |                |
|                 |      |        |          |            |                                      | Sens: 0.8956±0.0020 |                |
|                 |      |        |          |            |                                      | Spec: 0.8905±0.0030 |                |
|                 |      |        |          |            |                                      | Pre: 5.01±1.05 mm |                |
|                 |      |        |          |            |                                      | Sens: 0.8966±0.0010 |                |
| Nazir et al.    | 2020 | BV     | CTA     | OFF-eNET   | CTA scans (70)                       | Acc: 0.8956±0.0020 |                |
|                 |      |        |          |            |                                      | Sens: 0.9075±0.0030 |                |
|                 |      |        |          |            |                                      | Spec: 6.01±1.05 |                |
|                 |      |        |          |            |                                      | Sens: 0.8966±0.0010 |                |

The regions include the retinal vessel (RV), lung airways (LA), brain vessel (BV), coronary vessel (CV), and thoracic aorta (TA). All unnamed datasets are private and may have specific modalities such as phase-contrast (PC), low-dose (LD), standard-dose (SD), or time-of-flight (TOF). *: methods that also use advanced architectures. All metrics are shown as mean ± standard deviation or mean only.

### 3.6. Open-source Implementations

To promote the usage of the state-of-the-art (SOTA) models and help future researchers generate comparable segmentation results, we recommend some open-source implementations of aforementioned algorithms here.

- **MONAI**: Project MONAI was originally started by NVIDIA and King’s College London. It is a freely available, PyTorch-based deep learning framework for healthcare imaging with comprehensive tutorials. The framework provides the user with SOTA, end-to-end training workflows and many other features such as interactive data labelling. More information can be found here: [https://monai.io/index.html](https://monai.io/index.html)

- **Alleviating Class-wise Gradient Imbalance for Pulmonary Airway Segmentation** [https://github.com/haozheng-sjtu/3d-airway-segmentation](https://github.com/haozheng-sjtu/3d-airway-segmentation) [73]

- **Automatic airway segmentation from computed tomography using robust and efficient 3-D convolutional neural networks** [https://github.com/antonioguj/bronchinet](https://github.com/antonioguj/bronchinet) [68]

- **Lung-Vessel-Segmentation-Using-Graph-cuts** [https://github.com/Zhiwei-Zhai/Lung-Vessel-Segmentation-Using-Graph-cuts](https://github.com/Zhiwei-Zhai/Lung-Vessel-Segmentation-Using-Graph-cuts) [59]
Human Treelike Tubular Structure Segmentation

Table 4
Airway Dataset

| Name               | Imaging Modality | Number of Scans | Image Dimension (h*w) | Year Published | Publicly Available | Link               |
|--------------------|------------------|-----------------|-----------------------|----------------|-------------------|--------------------|
| DLCST              | CT               | 32              | 512*512               | 2009           | No                | Publication        |
| COPDGene           | CT               | 400             | 512*512               | 2010           | No                | Website Publication |
| EXACT'09           | CT               | 1018            | 512*512               | 2015           | Yes               | Website Publication |
| LIDC-IDRI          | CT               | 40              | 512*512               | 2012           | Yes               | Website Publication |
| CF-CT              | CT               | 24              | 512*512               | 2017           | No                | Publication        |
| KOLD               | CT               | 65              | 512*512               | 2019           | No                | Publication        |
| Subset of LUNA16   | CT               | 345             | 512*512               | 2020           | No                | Publication        |

4. Dataset

To compare the performance of different segmentation models on the same scale, several datasets were made public and summarized in the table below.

4.1. Airways

Currently, there is a very limited number of public datasets for airway segmentation which hinders the development in this area of research. To our best, we summarized four commonly seen datasets in Table 4. EXACT’09 [2] dataset consists of 40 volumetric chest CT scans obtained using different acquisition protocols. The dataset has large intra-variation because it ranges from clinical dose to ultra-low dose scans, from healthy volunteers to patients with severe lung disease, and from full inspiration to full expiration. LIDC-IDRI dataset [133] was created by seven academic centres and eight medical imaging companies, containing 1018 lung cancer screening thoracic low-dose CT scans. It is worth noting that the original dataset does not contain airway annotations. Qin et al. [134] annotated the airway of 40 CT scans from LIDC-IDRI and can be download from here COPDGene study contains 400 CT scans taken from subjects who had a minimum of 10 pack-year smoking history. Volunteers suffered from a range of COPD. Detailed inclusion and exclusion criteria and acquisition protocols can be found in the publication. KOLD dataset contains 477 volumetric lung CT scans from patients with chronic obstructive pulmonary disease (COPD), asthma, or other unclassified obstructive lung diseases. Detailed inclusion criteria can be found in [135][136].

4.2. Blood Vessels

4.2.1. Retinal blood vessels

Fundus imaging refers to the projection of 3-D retinal tissues onto a 2-D imaging plane through reflected light, which encompasses a wide range of modalities. Detailed review on retinal imaging and fundus photography has been carefully done by [137] and [138] To stimulate the development of more and more sophisticated methods of fundus image analysis, many publicly available datasets with expert annotation for vessel segmentation have been established which greatly stimulates the community of research. Table 5 summarizes brief information of these datasets.

4.2.2. Others

Blood vessels have vital roles in many other anatomical regions as well. The segmentation of hepatic vessels, brain vessels and coronary vessels is often seen in the area of research and has great implications in many clinical fields such as oncology and surgery. Table 6 summarizes commonly used datasets.

5. Evaluation Metrics

Since the metric is an essential basis for model evaluation and selection, it is meaningful and necessary to investigate the appropriate metric. In general, pixels are classified as foreground pixels (positive) or background pixels (negative). On this basis, there are four basic pixel measures compared with GS segmentation, namely TP (true positive), FP (false positive), FN (false negative), and TN (true negative). Some general metrics to evaluate the performance of segmentation networks are presented in Table 7.

In addition, the Hausdorff distance (HD), across the result and the GT, measures the maximum distance between two label boundaries. The absolute volumetric difference (AVD) represents the percentage ratio of the absolute difference between the GT volume and the segmented volume, to the GT volume.

The receiver operating characteristic (ROC) curve is a graph that illustrates the trade-off between the Pre and FPR of the model at different thresholds. The closer the curve is to the upper left corner, the better the performance of the system. The most common performance metric extracted from the ROC curve is the area under the curve (AUROC), which is used to compare different models at the same threshold, with an AUROC of 1 for an optimal system. Similarly, the precision-recall (PR) curve demonstrates the trade-off between the Pre and Se, where the area under the curve (AUPR) is also considered a performance metric.
### Table 5
Retinal Blood Vessel Dataset

| Name          | Imaging Modality | Number of Scans | Image Dimension (h*w) | Year Published | Publicly Available | Link                |
|---------------|------------------|-----------------|-----------------------|----------------|--------------------|---------------------|
| STARE         | CFO              | 20              | 605*700               | 2000           | Yes                | Website Publication |
| DRIVE         | CFO              | 40              | 565*584               | 2004           | Yes                | Website Publication |
| ARIA          | CFO              | 143             | 768*576               | 2006           | Yes                | Website Publication |
| DIARETDB1     | CFO              | 89              | 1500*1152             | 2007           | Yes                | Website Publication |
| REVIEW        | CFO              | 16              | N/A                   | 2008           | Yes                | Website Publication |
| CHASE-DB1     | CFO              | 28              | 999*960               | 2009           | Yes                | Website Publication |
| ROC           | CFO              | 100             | 1389*1383             | 2010           | Yes                | Website Publication |
| VICA/VR       | CFO              | 58              | 768*584               | 2010           | Yes                | Website Publication |
| HRF           | CFO              | 45              | 3504*2336             | 2013           | Yes                | Website Publication |
| Messidor      | CFO              | 1748            | 1440*960              | 2014           | Yes                | Website Publication |
| RC-SLO        | SLO CFO          | 40              | 360*320               | 2015           | Yes                | Website Publication |
| IOSTAR        | SLO CFO          | 30              | 1024*1024             | 2016           | No                 | Publication         |
| PREVENT       | OCTA             | 11              | 91*91                 | 2020           | Yes                | Website Publication |

### Table 6
Others

| Name           | Anatomy | Imaging Modality | Number of Scans | Image Dimension (h*w) | Year Published | Publicly Available | Link                |
|----------------|---------|------------------|-----------------|-----------------------|----------------|--------------------|---------------------|
| IRCAD          | Liver   | CT               | 20              | 512*512               | 2009           | Yes                | Website Publication |
| VASCUSYNTH     | Liver   | Algorithm synthesized | 10 | 101*101               | 2010           | Yes                | Website Publication |
| OSMSC          | Coronary and lung | MRA and CT | 93              | N/A                   | 2013           | Yes                | Website Publication |
| ROTTERDAM      | Coronary | CTA             | 32              | N/A                   | 2009           | No                 | Publication         |
| CASDQEF        | Coronary | CTA             | 48              | N/A                   | 2013           | No                 | Publication         |
| VESSEL12       | Lung    | CT               | 20              | 512*512               | 2012           | Yes                | Website Publication |
| MSD            | Liver   | CT               | 443             | 512*512               | 2021           | Yes                | Website Publication |
| MIDAS21        | Brain   | TOF-MRA          | 109             | 448*448               | 2005           | Yes                | Website Publication |

### 6. Discussion

#### 6.1. Research Trends

The segmentation of human treelike tubular structures has been a heavily researched area in recent years. As shown in Figure 18, the number of publications shows an overall upward trend from 2000 to 2021. However, the recent development of interest in this research area is mainly attributed to the emergence of DL algorithms. According to the publication of CNNs and the UNet that specifically for medical image segmentation [110], the number of DL-based methods for human treelike tubular structure segmentation started to increase remarkably in 2017 and became dominant in the last 3 years. In contrast, the number of new conventional methods has been decreasing since the first decade and became zero in 2018. Although new ML-based methods have been proposed since 2004, they have hardly dominated due to the need to extract features manually and the higher computation...
Table 7
Evaluation metrics

| Metric                          | Definition                                                                 |
|--------------------------------|---------------------------------------------------------------------------|
| Accuracy (Acc)                  | \( \frac{TP+TN}{TP+TN+FP+FN} \)                                           |
| Sensitivity (Se)                | \( \frac{TP}{TP+FN} \)                                                    |
| Specificity (Sp)                | \( \frac{TN}{TN+FP} \)                                                    |
| Precision (Pre)                 | \( \frac{TP}{TP+FP} \)                                                    |
| False positive rate (FPR)       | \( \frac{FP}{TP+FP} \)                                                    |
| True negative rate (TNR)        | \( \frac{TN+FN}{TP+FN} \)                                                 |
| Dice similarity coefficient (DSC)| \( \frac{2TP}{2TP+TN+FN+FP} \)                                           |
| Jaccard index (IoU)             | \( \frac{TP}{TP+FN+FP} \)                                                 |
| Volumetric overlap error (VOE)  | \( 1 - \frac{TP}{TN+FN+FP} \)                                             |
| Geometric mean (G)              | \( \sqrt{Se \times Sp} \)                                                  |
| Matthews correlation coefficient (MCC)| \( \frac{(TP\cdot TN) - (FP\cdot FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \) |

Generation count (GC) Number of correctly detected generations
Branch count (BC) Number of correctly detected branches excluding the trachea
Branches detected (BD) Percentage of correctly detected branches
Named Branch Count (NBC) Number of correctly detected named branches
Tree length (TL) Length of correctly detected branches excluding the trachea
Tree length detected (TLD) Percentage of correctly detected tree length
Leakage count (LC) Number of disconnected leakage
Leakage volume (LV) Volume of leakage
Centreline overlap (CO) Percentage of centreline overlap

Figure 18: Number of publications and years in terms of segmentation methods.

According to the trend analysis, the utilization of the DL algorithm for human treelike tubular structure segmentation is the main research direction recently and in the future. Compared with the conventional method and ML-based method, the DL-based method has the following advantages. First, conventional and unsupervised ML-based methods may suffer from performance shortcomings because they do not benefit from the manually labelled ground truth. Moreover, traditional methods require special templates or rules designed for segmentation targets, and ML algorithms need to extract meaningful features from images, both requiring deep domain expertise. DL algorithms, on the other hand, can directly extract suitable internal representations of images for automatic learning. Meanwhile, manual feature selection is application-specific and thus lacks the ability to generalize and learn new features. DL algorithms, on the contrary, can automatically learn features at various levels without application-specific restrictions, and hence have higher generalizability and robustness.

6.2. Segmentation Methods
From the perspective of the method classification, 40.26% (31) of the 77 papers reviewed combined DL algorithms,
conventional methods accounted for 35.06%, while ML-based methods had the lowest percentage. A more detailed breakdown of the percentage of method categories is shown in Figure 19. For the conventional methods, the RG method was the most widely used, probably due to its efficiency in 3D tree segmentation. Supervised and unsupervised ML algorithms were equally studied for human treelike tubular structure segmentation.

As for DL-based segmentation methods, task-based/hybrid algorithms were the most dominant, followed by architecture-based improvements. These methods had made targeted modifications to the segmentation task or model architecture for tree-like tubular structures. For example, some methods performed multi-task segmentation of thick and thin tubular structures or introduced GNN for tree-like connectivity. Classical CNNs were not popular because they are more generic and therefore did not achieve optimal segmentation performance on tree-like structures. Practically, for most task-based/hybrid segmentation, advanced architecture was a must. Therefore, a large portion of task-based methods also developed novel DL models at the same time, such as combining GNN in order to transform segmentation into graph refinement [83].

Quantitative analysis of segmentation performance for all reviewed methods is limited as they were evaluated on different datasets with various modalities and protocols. Therefore, the quantitative analysis was conducted representatively on two public datasets adopted by most methods, DRIVE and EXACT’09, for blood vessel segmentation and airway segmentation, respectively.

6.3. Blood vessel segmentation
27 of the 77 reviewed papers validated their performance for segmenting blood vessels on the DRIVE dataset. We quantitatively compared means of the four most used metrics (Acc, Se, AUROC, and Sp), but excluded a method using Visual evaluation [21], as shown in Figure 21. It’s worth noting that not every method used all four metrics, so our mean comparisons were only based on methods that used a particular metric.

Overall, DL-based methods have the best segmentation performance, followed by ML-based methods, and finally Conventional methods. For the Conventional method, basic image processing operations are most sensitive to vascular structure, but template matching has higher segmentation accuracy and specificity with fewer false positives according to AUROC. RG has moderate segmentation accuracy, but the sensitivity is not satisfactory. Similarly, unsupervised ML-based methods are more sensitive than supervised methods but slightly inferior in other segmentation metrics. On the other hand, task-based/hybrid DL-based methods...
achieve the best results on all metrics. DL-based methods based solely on advanced architecture or training strategies are lower in segmentation sensitivity even than Conventional methods, which is an obvious problem. Specifically, as shown in Figure 21c, all task-based/hybrid DL-based methods can achieve good segmentation performance. However, among all unsupervised ML-based methods, Hassan & Hassanien’s WOA method performs exceptionally well, reaching the state-of-the-art.

6.3.2. Airway segmentation

For airway segmentation, a total of 10 papers were evaluated in EXACT’09. After excluding one paper using duplicate methods [75] and one paper doing comparative analysis [67], the remaining 8 methods only covered part of the subcategories of ML-based and DL-based segmentation algorithms, as shown in Figure 22. We used mean values of the three most commonly applied metrics for quantitative comparison (TLD, BD, and FPR). Similarly, not every method...
used all three metrics.

Overall, the segmentation performance of DL-based methods is significantly higher than that of ML-based methods. Supervised ML-based methods are limited by the difficulty of finding suitable features and rules to maximize airway detection rate while minimizing segmentation leakage. Clearly, the method of Charbonnier et al. [63] has the lowest detection rate but the least leakage, while the method of Yun et al. [64] has a higher detection rate but also increased FPR. On the other hand, the advent of CNN enabled more suitable CT image features to be extracted, thus achieving a better balance between detection rate and FPR. We can see that by improving the classic CNN architecture, the detection rate of tracheal segmentation is further improved by more than 14% while maintaining a low FPR.

6.4. Open Challenges
Conventional segmentation methods suffer from incomplete prediction and leakage (shown in Fig. 23). EXACT’09 challenge has summarized performance of them on airway segmentation. On average, no traditional method achieves over 74% in TLR. Despite the recent sprout of deep learning based methods, the problem of missing thin terminal bronchioles in the final prediction still persists as shown in Fig. 24 and Fig. 25. Commonly used cost functions such as Binary cross entropy loss and Dice loss failed to address this problem. Results with high Dice coefficient can still have many disconnected peripheral branches which can be reflected by the low tree length detected ratio and branch detected ratio (often less than 70%). This problem is more predominant in the 3D cases than the 2D cases. In addition, model performance on pathological cases should be addressed in the future, morphological changes in structure can pose a further challenge to the robustness of the model.

Another problem is the lack of metrics to quantify the morphological shape of the airway which often contains useful information for lung-related disease diagnosis and prognosis. [140, 141, 142, 143] measure the airway thickness and show a strong correlation between it and reduced air flow of the lung. Apart from the airway wall thickness, very few studies focused on other morphological parameters such as branching patterns. This is partly because there is no mathematical way to quantify the spatial patterns of the human airway. [144] proposed a new metric named Airway Fractal Dimension (AFD) to help predict respiratory morbidity and mortality in COPD. AFD is calculated using the Minkowski-Bougliand box-counting dimension which essentially measures the self-similarity of a graph. In the case of airway structures, the more branch it has the more detailed it is considered. A formal mathematical definition of the box-counting dimension is shown below:

$$D = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{-\log \epsilon}$$

where $D$ is the fractal dimension, $N(\epsilon)$ is the number of
They analysed 8,135 participants enrolled in the COPDGene cohort and found a strong correlation \( p < 0.001 \) between AFD and forced expiratory volume. They also did a multivariate analysis and confirmed that AFD can be used in addition to airway thickness to give better prognosis. An intuitive visual representation is also provided by them as shown in Fig. 26. A review on computing the fractal dimension in medical imaging has been done by Konatar I. et al. [145]

There hasn’t been any release of new datasets with airway annotation in the community for years. Only 90 CT scans (20 from EXACT’09 and 70 from LIDC-IDRI) are widely publicly available, and they come from patients with a wide range of physical conditions.

Most proposed models cannot extend to more complex clinical settings because the algorithms would fail to predict airway regions when encountering pathological cases such as fibrosis and COVID-19. Moreover, the variations between CT scans acquired by different protocols also pose a challenge to the generalization of the model.

To summarize, there are still some open challenges for accurate and robust human treelike tubular structure segmentation, which are as follows:

- DL-based approaches through unsupervised or semi-supervised learning is a topic that has not been thoroughly investigated yet. Unsupervised training can overcome the current dilemma of lacking properly annotated datasets by domain experts.
- In terms of training strategies, the DL algorithm may have great room for improvement. For instance, applying hard sampling mining to improve the segmentation performance in pathological cases.
- There is a lack of a standard set of evaluation metrics and a systematic evaluation process for human tree tubular structure segmentation methods. The inconsistent metrics and the fact that most of the datasets are not publicly available make the performance comparison between algorithms very difficult. As mentioned above, using sole metrics such as accuracy, IoU and Dice cannot fully access the model on a spatial scale. Therefore, an open evaluation scheme across datasets is urgently needed.
- Binary cross entropy loss and Dice loss are not able to capture the discontinuity in peripheral bronchioles. New loss functions, such as cIDice [139], that emphasize connectivity should also be incorporated in training the model.
- For conventional image processing methods, their results are fully explainable. However, deep neural networks are normally considered as ‘black boxes’, whereas their interpret ability for tubular structure segmentation remains to be explored. Researchers should focus more on explainable AI solutions [146]. This could be achieved by sensitivity analysis [147], gradient-based methods [148], and data harmonisa-
tion strategies for multi-centre studies [149].

7. Conclusion
This review summarizes the algorithms, datasets, and evaluation metrics for human treelike structure segmentation methods in the literature. We present a systematic classification of the different algorithms, supported by tables reporting the anatomical regions of interest, the datasets used, and performance metrics, which can help researchers to better understand the available options and methods. Based on the analysis of the literature, deep learning based segmentation methods have become dominant with their advantages of capturing hidden information in complex structures. In this regard, this review proposes feasible research directions based on deep-learning algorithms, evaluation metrics, and loss functions to accelerate the development and improvement of human treelike tubular structure segmentation methods.

CRediT authorship contribution statement
Hao Li: Methodology, Formal analysis, Investigation, Writing - Original Draft, Visualization. Zeyu Tang: Methodology, Formal analysis, Investigation, Writing - Original Draft, Visualization. Yang Nan: Conceptualization, Methodology, Writing - Review and Editing. Supervision. Guang Yang: Conceptualization, Methodology, Writing - Review and Editing, Supervision, Funding acquisition.

References
[1] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, M. van Ginneken, Ridge-based vessel segmentation in color images of the retina, IEEE Transactions on Medical Imaging 23 (2004) 501–509.
[2] P. Lo, B. van Ginneken, J. M. Reinhardt, T. Yavarna, P. A. de Jong, B. Irving, C. Fettia, M. Ortnt, R. Pinho, J. Sijbers, M. Feuerstein, A. Fabijanska, C. Bauer, R. Beichel, C. S. Mendoza, R. Wiemker, J. Lee, A. P. Reeves, S. Born, O. Weinheimer, E. M. van Rikvoort, J. Tschirren, K. Mori, B. Ory, D. P. Naidich, I. Hartmann, E. A. Hoffmann, M. Prokop, J. H. Pedersen, M. de Bruijne, Extraction of Airways From CT (EXACT’09), IEEE Transactions on Medical Imaging 31 (2012) 2093–2107.
[3] P. Namburi, P. Zuo, G. Augustine, Neural mapping: Understanding how the cerebellum is wired, IBRO-KIST Summer School, Korea Institute of Science and Technology Courtesy Koch Institute Image Awards at MIT (2012).
[4] G. Rubin, Data explosion: The challenge of multidetector-row ct, European Journal of Radiology 36 (2000) 78–80.
[6] E. T. D. R. S. R. Group, Fundus photographic risk factors for progression of diabetic retinopathy: Etdrs report number 12, Ophthalmology 98 (1991) 823–833.
[6] J. W. Yau, S. L. Rogers, R. Kawasaki, E. L. Lamoureux, J. W. Kowalski, T. Bek, S.-J. Chen, J. M. Dekker, A. Fletcher, J. Grauslund, S. Haffner, R. F. Hamman, M. K. Ikrar, T. Kayama, B. E. Klein, R. Klein, S. Krishnaiah, K. Mayurasakorn, J. P. O’Hare, T. J. Orchard, M. Porta, M. Rema, M. S. Roy, T. Sharma, J. Shaw, H. Taylor, J. M. Tielisch, R. Varma, J. J. Wang, N. Wang, S. West, L. Xu, M. Yasuda, X. Zhang, P. Mitchell, T. Y. Wong, Global prevalence and major risk factors of diabetic retinopathy, Diabetes Care 35 (2012) 556–564.
[7] J. Pu, S. Gu, S. Liu, S. Zhu, D. Wilson, J. M. Siegfried, D. Gur, Ct based computerized identification and analysis of human airways: A review, Medical Physics 39 (2012) 2603–2616.
[8] S. Moccia, E. De Momi, S. El Hadji, L. S. Mattos, Blood vessel segmentation algorithms — Review of methods, datasets and evaluation metrics, Computer Methods and Programs in Biomedicine 158 (2018) 71–91.
[9] C. Chen, J. H. Chua, R. Ali, Y. Wang, Retinal vessel segmentation using deep learning: A review, IEEE Access 9 (2021) 111985–112004.
[10] M. Ciecholewski, M. Kassjański, Computational methods for liver vessel segmentation in medical imaging: A review, Sensors 21 (2021) 2027.
[11] A. A. Abdulshahid, M. A. Mahmoud, M. A. Mohammed, H. H. Rasheed, S. A. Mostafa, M. S. Maashi, Comprehensive review of retinal blood vessel segmentation and classification techniques: Intelligent solutions for green computing in medical images, current challenges, open issues, and knowledge gaps in fundus medical images, Network Modeling Analysis in Health Informatics and Bioinformatics 10 (2021) 20.
[12] M. R. K. Mookiah, S. Hogg, T. J. MacGillivray, V. Prathiba, R. Pradeepa, V. Mohan, R. M. Anjana, A. S. Doney, C. N. Palmer, E. Trucco, A review of machine learning methods for retinal blood vessel segmentation and artery/vein classification, Medical Image Analysis 68 (2021) 101905.
[13] J. Wallyn, N. Anton, S. Akram, T. F. Vandamme, Biomedical imaging: Principles, technologies, clinical aspects, contrast agents, limitations and future trends in nanomedicines, Pharmaceutical Research 36 (2019) 78.
[14] T. D. Noto, G. Marie, S. Tourbier, Y. Alemán-Gómez, O. Esteban, G. Saliou, M. B. Cuadra, P. Hagmann, J. Richiardi, Lausanne tof mra aneurysm cohort, OpenNeuro (2022).
[15] A. Hoover, V. Kouznetsova, M. Goldbaum, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response, IEEE Transactions on Medical Imaging 19 (2000) 203–210.
[16] P. H. Tomlins, R. K. Wang, Theory, developments and applications of optical coherence tomography, Journal of Physics D 38 (2005) 2519–2535.
[17] R. F. Spaide, J. G. Fujimoto, N. K. Waheed, S. R. Sadda, G. Staurenghi, Optical coherence tomography angiography, Progress in Retinal and Eye Research 64 (2018) 1–55.
[18] Y. Giarratano, E. Bianchi, C. Gray, A. Morris, T. MacGillivray, B. Dhillon, M. O. Bernabeau, Automated Segmentation of Optical Coherence Tomography Angiography Images: Benchmark Data and Clinically Relevant Metrics, Translational Vision Science and Technology 9 (2020) 5–5.
[19] Xiaoyi Jiang, D. Mojon, Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2003) 131–137.
[20] Jianao Pu, C. Fuhrman, W. F. Good, F. C. Sciarba, D. Gur, A Differential Geometric Approach to Automated Segmentation of Human Airway Tree, IEEE Transactions on Medical Imaging 30 (2011) 266–278.
[21] F. Zana, J.-C. Klein, Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation, IEEE Transactions on Image Processing 10 (2001) 1010–1019.
[22] N. Passat, C. Ronse, J. Baruthio, J.-P. Armspach, C. Maillot, Mathematical morphology and curvature evaluation, IEEE Transactions on Image Processing 10 (2001) 256–274.
[23] B. Bouraoui, C. Ronse, J. Baruthio, N. Passat, P. Germain, 3D segmentation of coronary arteries based on advanced mathematical morphology techniques, Computerized Medical Imaging and Graphics 34 (2010) 377–387.
[24] U. T. Nguyen, A. Bhuiyan, L. A. Park, K. Ramamohanarao, An effective retinal blood vessel segmentation method using multi-scale line detection, Pattern Recognition 46 (2013) 703–715.
[25] D. Aykac, E. Hoffman, G. McLennan, J. Reinhardt, Segmentation...
and analysis of the human airway tree from three-dimensional X-ray CT images, IEEE Transactions on Medical Imaging 22 (2003) 940–950.

[26] L. Câmara Neto, G. L. Ramalho, J. F. Rocha Neto, R. M. Veras, F. N. Medeiros, An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images, Expert Systems with Applications 78 (2017) 182–192.

[27] A. Hoover, V. Kouznetsova, M. Goldbaum, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response, IEEE Transactions on Medical Imaging 19 (2000) 203–210.

[28] B. Zhang, L. Zhang, L. Zhang, F. Karray, Retinal vessel extraction by matched filter with first-order derivative of Gaussian, Computers in Biology and Medicine 40 (2010) 438–445.

[29] G. Azzopardi, N. Strisciuglio, M. Vento, N. Petkov, Trainable COSFIRE filters for vessel delineation with application to retinal images, Medical Image Analysis 19 (2015) 46–57.

[30] K. Kriissan, G. Malandain, N. Ayache, R. Vaillant, Y. Trousset, Model-Based Detection of Tubular Structures in 3D Images, Computer Vision and Image Understanding 80 (2000) 130–171.

[31] S. Woz, K. Rohr, Segmentation and Quantification of Human Vessels Using a 3-D Cylindrical Intensity Model, IEEE Transactions on Image Processing 16 (2007) 1994–2004.

[32] A. P. Kiraly, W. E. Higgins, G. McLennan, E. A. Hoffman, J. M. Reinhardt, Three-Dimensional Human Airway Segmentation Methods for Clinical Virtual Bronchoscopy, Academic Radiology 9 (2002) 1153–1168.

[33] C. Fettia, F. Preteux, C. Beigelman-Aubry, P. Grenier, Pulmonary Airways: 3-D Reconstruction From Multislice CT and Clinical Investigation, IEEE Transactions on Medical Imaging 23 (2004) 1353–1364.

[34] J. Tschirren, E. Hoffman, G. McLennan, M. Sonka, Intrathoracic airway trees: Segmentation and airway morphology analysis from low-dose CT scans, IEEE Transactions on Medical Imaging 24 (2005) 1529–1539.

[35] M. E. Martínez-Pérez, A. D. Hughes, S. A. Thom, A. A. Bharath, K. H. Parker, Segmentation of blood vessels from red-free and fluorescein retinal images, Medical Image Analysis 11 (2007) 47–61.

[36] A. Fabijanska, Two-pass region growing algorithm for segmenting airway tree from MDCT chest scans, Computerized Medical Imaging and Graphics 33 (2009) 537–546.

[37] S. Cetin, A. Demir, A. Yezzi, M. Degertekin, G. Unal, Vessel Tractography Using an Intensity Based Tensor Model With Branch Detection, IEEE Transactions on Medical Imaging 32 (2013) 348–363.

[38] S. Cetin, G. Unal, A Higher-Order Tensor Vessel Tractography for Segmentation of Vascular Structures, IEEE Transactions on Medical Imaging 34 (2015) 2172–2185.

[39] D. Bartz, D. Mayer, J. Fischer, S. Ley, A. del Río, S. Thust, C. Heusel, H.-U. Kauczor, W. Strasser, Hybrid segmentation and exploration of the human lungs, in: IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, IEEE, Seattle, WA, USA, 2003, pp. 177–184. doi: 10.1109/VIU. 2003.1258370.

[40] D. Mayer, D. Bartz, J. Fischer, S. Ley, A. del Rio, S. Thust, H.-U. Kauczor, C. P. Heusel, Hybrid segmentation and virtual bronchoscopy based on CT images,1. Academic Radiology 11 (2004) 551–565.

[41] M. Graham, J. Gibbs, D. Cornish, W. Higgins, Robust 3-D Airway Tree Segmentation for Image-Guided Peripheral Bronchoscopy, IEEE Transactions on Medical Imaging 29 (2010) 982–997.

[42] A. Mendonca, A. Campilho, Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction, IEEE Transactions on Medical Imaging 25 (2006) 1200–1213.

[43] D. Hutchison, T. Kanade, J. Kitterl, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, M. W. K. Law, A. C. S. Chung, An Oriented Flux Symmetry Based Active Contour Model for Three Dimensional Vessel Segmentation, in: K. Daniilidis, P. Maragos, N. Paragios (Eds.), Computer Vision – ECCV 2010, volume 6313, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 720–734. doi: 10.1007/978-3-642-15558-1_52.

[44] L. Lorigo, O. Faugeras, W. Grimson, R. Keriven, R. Kikinis, A. Nabavi, C.-F. Westin, CURVES: Curve evolution for vessel segmentation, Medical Image Analysis 5 (2001) 195–206.

[45] A. Klepaczko, P. Szczypinski, A. Deistung, J. R. Reichenbach, A. Materka, Simulation of MR angiography imaging for validation of cerebral arteries segmentation algorithms, Computer Methods and Programs in Biomedicine 137 (2016) 293–309.

[46] M. S. Hossanna, A. Farag, S. Hushke, T. Moriarty, Cerebrovascular segmentation from TOF using stochastic models, Medical Image Analysis 10 (2006) 2–18.

[47] D. A. Oliveira, R. Q. Feitosa, M. M. Correia, Segmentation of liver, its vessels and lesions from CT images for surgical planning, BioMedical Engineering OnLine 10 (2011) 30.

[48] S. Roychowdhury, D. Kozokanekan, K. Parhi, Blood Vessel Segmentation of Fundus Images by Major Vessel Extraction and Sub-Image Classification, IEEE Journal of Biomedical and Health Informatics (2014) 1–1.

[49] G. B. Kande, P. V. Subbaiah, T. S. Savithri, Unsupervised Fuzzy Based Vessel Segmentation In Pathological Digital Fundus Images, Journal of Medical Sciences 34 (2010) 849–858.

[50] E. Goceri, Z. K. Shah, M. N. Gurcan, Vessel segmentation from abdominal magnetic resonance images: Adaptive and reconstructive approach: Vessel Segmentation from Abdominal MR Images, International Journal for Numerical Methods in Biomedical Engineering 33 (2017) e2811.

[51] M. G. Cinsdikici, D. Aydn, Detection of blood vessels in ophthalmoscope images using MF/ant (matched filter/ant colony) algorithm, Computer Methods and Programs in Biomedicine 96 (2009) 85–95.

[52] A. E. Hassanien, E. Emary, H. M. Zawbaa, Retinal blood vessel localization approach based on bee colony swarm optimization, fuzzy c-means and pattern search, Journal of Visual Communication and Image Representation 31 (2015) 186–196.

[53] G. Hassan, A. E. Hassanien, Retinal fundus vasculature multilevel segmentation using whale optimization algorithm, Signal, Image and Video Processing 12 (2018) 263–270.

[54] O. Ramos-Soto, E. Rodriguez-Esparza, S. E. Balderas-Mata, D. Oliva, A. E. Hassanien, R. K. Meleppat, R. J. Wadzawki, An efficient retinal blood vessel segmentation in eye fundus images by using optimized top-hat and homomorphic filtering, Computer Methods and Programs in Biomedicine 201 (2021) 105949.

[55] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, B. van Ginneken, Ridge-Based Vessel Segmentation in Color Images of the Retina, IEEE Transactions on Medical Imaging 23 (2004) 501–509.

[56] P. Lo, J. Sporring, H. Ashraf, J. J. Pedersen, M. de Bruijne, Vessel-guided airway tree segmentation: A voxel classification approach, Medical Image Analysis 14 (2010) 527–538.

[57] E. Ricci, R. Perfetti, Retinal Blood Vessel Segmentation Using Line Operators and Support Vector Classification, IEEE Transactions on Medical Imaging 26 (2007) 1357–1365.

[58] Q. Meng, T. Kitasaka, Y. Nimura, M. Oda, J. Ueno, K. Mori, Automatic segmentation of airway tree based on local intensity filter and machine learning technique in 3D chest CT volume, International Journal of Computer Assisted Radiology and Surgery 12 (2017) 245–261.

[59] Z. Zhai, M. Staring, B. C. Stoel, Lung vessel segmentation in CT images using graph-cuts, in: M. A. Styner, E. D. Angelini (Eds.), Medical Image Analysis 2016: Image Processing, volume 9784, International Society for Optics and Photonics, SPIE, 2016, p. 9784K.

[60] M. F. N. Medeiros, An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images, Expert Systems with Applications 78 (2017) 182–192.

[61] M. Lee, J.-G. Lee, N. Kim, J. B. Seo, S. M. Lee, Hybrid Airway Segmentation Using Multi-Scale Tubular Structure Filters and Texture Analysis on 3D Chest CT Scans, Journal of Digital Imaging 32 (2019) 779–792.
[125] Z. Zhou, M. M. Rahman Siddiquiee, N. Tajbakhsh, J. Liang, UNet++: A Nested U-Net Architecture for Medical Image Segmentation, in: D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, A. Madabhushi (Eds.), Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, volume 11045, Springer International Publishing, Cham, 2018, pp. 3–11. doi: 10.1007/978-3-030-08889-5_1.

[126] M. Mirza, S. Osindero, Conditional Generative Adversarial Nets, arXiv:1411.1784 [cs, stat] (2014).

[127] C. E. Duchon, Lanzcos Filtering in One and Two Dimensions, Journal of Applied Meteorology 18 (1979) 1016–1022.

[128] E. Shehham, I. Shaked, T. Darrell, Fully Convolutional Networks for Semantic Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence (2017) 640–651.

[129] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 42 (2020) 318–327.

[130] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, S. P. Smolley, Least Squares Generative Adversarial Networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, 2017, pp. 2813–2821. doi: 10.1109/ICCV.2017.304.

[131] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 629–639.

[132] C. Fiorio, J. Grey, Two linear time Union-Find strategies for image processing, Theoretical Computer Science 154 (1996) 165–181.

[133] S. G. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. C. Fiorio, J. Gustedt, Two linear time Union-Find strategies for image processing, Theoretical Computer Science 154 (1996) 165–181.

[134] S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhev, A. Unger, A. Zhylka, J. P. W. Pluim, U. Bauer, B. H. Menze, cldice - a novel topology-preserving loss function for tubular structure segmentation, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 16555–16564. doi: 10.1109/CVPR46437.2021.01629.

[135] I. Orlandi, C. Moroni, G. Camiciotto, M. Bartolucci, M. Pistolesi, N. Villari, M. Mascalchi, Chronic obstructive pulmonary disease: Thin-section ct measurement of airway wall thickness and lung attenuation, Radiology 234 (2005) 604–610.

[136] T. Achenbach, O. Weinheimer, A. Biedermann, S. Schmitt, D. Freudentein, E. Goutham, R. P. Kunz, R. Buhl, C. Dueber, C. P. Heusel, Mdct assessment of airway wall thickness in copd patients using a new method: correlations with pulmonary function tests, European Radiology 18 (2008) 2731–2738.

[137] P. A. de Jong, Computed tomographic imaging of the airways: relationship to structure and function, European Respiratory Journal 26 (2005) 140–152.

[138] M. Hasegawa, Y. Nasuhara, Y. Onodera, H. Makita, K. Nagai, S. Fuke, Y. Ito, T. Betsuyaku, M. Nishimura, Airflow limitation and airway dimensions in chronic obstructive pulmonary disease, American Journal of Respiratory and Critical Care Medicine 173 (2006) 1309–1315.

[139] S. Bodduluri, A. S. K. Puliyakote, S. E. Gerard, J. M. Reinhardt, E. A. Hoffman, J. D. Newell, H. P. Nath, M. K. Han, G. R. Washko, R. J. S. Estépar, M. T. Dransfield, S. P. Bhatt, C. Investigators, Airway fractal dimension predicts respiratory morbidity and mortality in copd, Journal of Clinical Investigation 128 (2018) 5374–5382.

[139] I. Konatar, T. Popovic, N. Popovic, Box-counting method in python for fractal analysis of biomedical images, in: 2020 24th International Conference on Information Technology (IT), 2020, pp. 1–4. doi: 10.1109/IT48810.2920.3909454.

[140] G. Yang, Q. Ye, J. Xia, Unbox the black-box for the medical explainable ai via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond, Information Fusion 77 (2022) 29–52.

[141] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, B. Kim, Sanity checks for saliency maps, Advances in neural information processing systems 31 (2018).

[142] F. Doshi-Velez, B. Kim, Considerations for evaluation and generalization in interpretable machine learning, in: Explainable and interpretable models in computer vision and machine learning, Springer, 2018, pp. 3–17.

[143] Y. Nan, J. Del Ser, S. Walsh, C. Schönlieb, M. Roberts, I. Selby, K. Howard, J. Owen, J. Neville, J. Guiot, et al., Data harmonisation for informatics fusion in digital healthcare: A state-of-the-art systematic review, meta-analysis and future research directions, Information Fusion (2022).