



**Politecnico
di Torino**

Politecnico di Torino

Master's Degree in DATA SCIENCE AND ENGINEERING

A.a. 2024/2025

Graduation Session October 2025

Exploring Feature-Based and Data-Driven Techniques for Quality Control in Brain Vessel Segmentation

Supervisors:

Luigi Borzì
Maria A. Zuluaga

Candidate:

Lorenzo Suppa

Acknowledgements

A mamma Sabrina e papà Vittorio, che mi avete sempre sostenuto su ogni fronte, rendendo possibile tutto questo. Mi siete stati accanto con costanza, fiducia e amore, incoraggiandomi a uscire dalla mia zona di comfort e ad alimentare la mia curiosità sempre.

Un grazie profondo ai miei nonni, che mi hanno trasmesso fiducia in ogni passo, cercando di comprendere e supportare ogni mia scelta. Ho sempre sentito il vostro affetto e la vostra fiducia, e questo per me ha significato moltissimo.

Ad Arianna, persona solare con cui ho condiviso gran parte delle mie giornate negli ultimi anni, che le ha rese più belle e ha reso ogni fatica più leggera.

Infine, grazie agli amici di sempre e a quelli conosciuti lungo il percorso universitario: siete stati e siete tuttora preziosissimi. Elencarvi tutti sarebbe superfluo, tanto sapete bene quanto siete importanti e speciali.

Table of Contents

List of Figures	IV
1 Introduction	1
1.1 Chapter Overview	2
1.2 Anatomy and Clinical Significance of Brain Vasculature	3
1.2.1 Arterial System	4
1.2.2 Venous System	7
1.3 Image Modalities	8
1.3.1 Techniques	8
1.3.2 Preprocessing	9
1.4 Segmentation	12
1.4.1 Definition	12
1.4.2 Evaluation with Ground-Truth Annotations	12
1.4.3 Common Evaluation Metrics	13
1.4.4 State-of-the-art	15
1.4.5 Encountered models	18
1.4.6 Why Brain Segmentation Matters	21
1.5 Quality Control, the other medal's face	22
1.6 Objectives and Contributions	24
2 Tool metrics' computation	26
2.1 Chapter Overview	27
2.2 Processing Pipeline for Vessel Masks	28
2.2.1 Preliminary transformations	29
2.2.2 Connected Components Decomposition	31
2.2.3 Segment Extraction and Tortuosity Analysis	33
2.2.4 Final Aggregation	37
2.2.5 Metric Computation and Storage	38

3	Learning-Based Approaches for Quality Control	41
3.1	Chapter Overview	42
3.2	Regression-Based Method	43
3.2.1	Overview	43
3.2.2	Feature-based Regression	44
3.2.3	Conclusion	53
3.3	Reconstruction-Based Methods	55
3.3.1	Overview	55
3.3.2	General encoder-decoder	55
3.3.3	VAE with search in latent space	58
3.3.4	Conclusion	60
4	Discussion and Conclusion	62
	Bibliography	67

List of Figures

1.1	The intricate vessel's network.	3
1.2	Cerebral arteries.	5
1.3	Time-of-flight MRA at the level of the Circle of Willis.	6
2.1	Tool's pipeline.	28
2.2	Each row shows the segments from one policy splitting.	34
3.1	Overview of the regression-based dataset construction pipeline. . . .	45
3.2	Distributions of target quality scores across the degraded dataset. .	46
3.3	Examples of possible slice degradations.	47
3.4	Performances on trainset.	49
3.5	Test Performance	51
3.6	Performance per model.	52
3.7	Feature importance.	53
3.8	Reconstruction flow using an encoder–decoder model.	56
3.9	Successful reconstruction instance ($DSC = 0.89$).	57
3.10	Inference reconstruction instance.	58
3.11	Pipeline of the VAE with latent space search.	59
3.12	Example of unstable soft predictions for a reconstructed slice. . . .	60

Chapter 1

Introduction

1.1 Chapter Overview

This chapter introduces the essential background needed to understand the topic of quality control in brain vessel segmentation. It begins by outlining the anatomy and clinical relevance of the brain vasculature, describing the organization and function of the arterial and venous systems. This section highlights why accurate visualization and analysis of these structures are crucial for diagnosing and monitoring cerebrovascular diseases.

The chapter then moves on to the imaging modalities used to acquire vascular data, presenting the most common techniques (e.g, MRI, CT) and discussing their specific capabilities and limitations in capturing vascular detail. It also covers preprocessing steps, which are critical for ensuring that imaging data are suitable for further computational analysis.

Next, the concept of segmentation is introduced, not as the main focus of this work, but as an essential component for understanding the complex context in which quality control operates. The section defines segmentation, explains how it is evaluated through ground-truth annotations and performance metrics, and reviews current state-of-the-art approaches and available models (including nnU-Net, A2V, JoB-VS, VesselBoost, and SPOCKMIP) that have been encountered throughout the journey. Understanding these methods and their limitations provides the necessary foundation for appreciating why quality control is such a critical challenge.

The following section, Quality Control, the other medal’s face, addresses the central theme of this work: ensuring the reliability, robustness, and interpretability of segmentation outputs. It emphasizes that even the most advanced segmentation algorithms can fail or produce inconsistent results, underscoring the need for systematic quality assessment mechanisms.

Finally, the chapter concludes with the objectives, linking the anatomical, imaging, and methodological discussions to the overarching goal of this research, setting and testing the first basic approaches and ideas for quality control in brain vessel segmentation.

1.2 Anatomy and Clinical Significance of Brain Vasculature

The brain, the central organ of the nervous system, functions as the body's primary command center. Owing to its high metabolic demand, it requires a continuous and tightly regulated supply of oxygen and glucose, which are essential for the survival and activity of neurons and glial cells. To meet this demand, a dense and highly branched vascular network spreads across the pial surface. An approximate, yet still representative, illustration is provided in Figure 1.1.

From this superficial plexus, penetrating arteries descend into the cortical layers to reach neurovascular units [1]. Beyond nutrient delivery, cerebral arteries also facilitate clearance of interstitial fluid (ISF) and metabolic waste products, including amyloid-beta. Impairment of this perivascular drainage contributes to pathological protein accumulation, a mechanism strongly implicated in Alzheimer's disease [2]. At the microvascular level, capillaries play a central role. Their walls are composed of endothelial cells supported by contractile pericytes, collectively forming part of the blood-brain barrier (BBB). The BBB exerts strict regulation over molecular exchange, thereby preserving homeostasis of the neural microenvironment.

Capillaries also contribute to ISF production, further underscoring their importance in cerebral fluid dynamics. Venous drainage begins with post-capillary venules, which direct blood toward larger cortical veins before ultimately converging into the dural venous sinuses. In this way, the cerebral vasculature can be broadly categorized into three interconnected components: arterial, capillary, and venous networks.

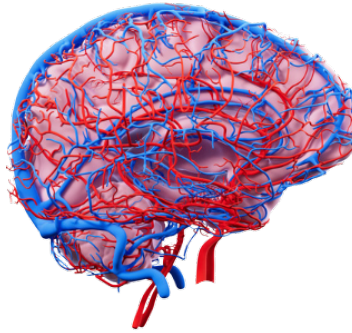


Figure 1.1: The intricate vessel's network.

A comprehensive understanding of the anatomy and vulnerabilities of the cerebral vasculature is indispensable for elucidating mechanisms of normal brain function as well as the pathogenesis of cerebrovascular and neurodegenerative diseases.

1.2.1 Arterial System

The cerebral blood supply is derived primarily from two systems: the internal carotid arteries (ICAs), which contribute approximately 70% of total cerebral blood flow, and the vertebrobasilar system, which accounts for the remaining 30%. The vertebral arteries originate from the subclavian arteries and unite to form the basilar artery. The basilar artery subsequently bifurcates into the posterior cerebral arteries (PCAs), which establish anastomotic connections with the ICAs via the posterior communicating arteries. These vessels together constitute the Circle of Willis, a critical arterial ring at the base of the brain. Anteriorly, each ICA bifurcates into the middle cerebral artery (MCA) and the anterior cerebral artery (ACA). The ACAs are joined by the anterior communicating artery, completing the anterior portion of the Circle of Willis. The anterior cerebral circulation comprises the ACAs, MCAs, and the anterior choroidal artery. The anterior choroidal artery typically arises proximal to the ACA–MCA bifurcation, although anatomical variations are common. Functionally, the ACAs supply the medial aspects of the cerebral hemispheres, the MCAs perfuse the lateral cortical surfaces, and the PCAs provide circulation to the occipital lobes and inferior temporal regions [3]. A precise illustration of the arterial system is presented in Figure 1.2.

The Circle of Willis is central both to this work and to the human brain; therefore, it is essential to provide deeper insight into its importance and clinical significance. Figure 1.3 clearly illustrates the considerable space it occupies. The Circle of Willis (CoW) is a vital arterial network located at the base of the brain, connecting the bilateral internal carotid arteries and the basilar artery. Structurally, it forms a ring composed of several key segments: the precommunicating portions of the anterior cerebral arteries (A1), the precommunicating portions of the posterior cerebral arteries (P1), the posterior communicating arteries (Pco), and the anterior communicating artery (Aco). The anterior section of the CoW consists of the A1 segments and the Aco, while the posterior section is formed by the P1 segments and the Pco.

Functionally, CoW, along with the ophthalmic artery and leptomeningeal vessels, plays a crucial role in maintaining cerebral blood flow. When the blood supply from primary arteries is compromised, a complete and well-formed CoW can redistribute blood from other brain regions to the affected areas, helping to prevent potentially severe outcomes.

As the primary source of arterial distribution in the brain, the CoW serves as the most important collateral circulation route.

Consequently, understanding its anatomy is essential for research into cerebral hemodynamics and is critically important in the prevention and treatment of stroke and several other cerebrovascular disorders.

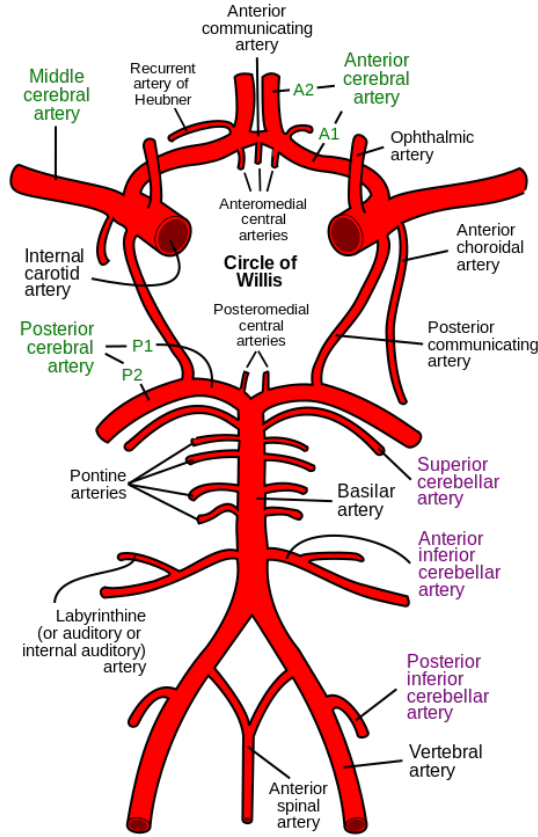


Figure 1.2: Cerebral arteries.

Despite its significance, CoW shows a high degree of anatomical variability between different populations. These variations often involve missing or underdeveloped segments.

Although the widespread occurrence of these variations is well documented, their underlying causes remain largely unknown [4]. The diseases directly linked with these variants are multiple:

- **Ischemic Stroke:** An intact Circle of Willis helps redistribute blood flow during arterial occlusion, reducing risk of brain infarction. In patients with $\geq 70\%$ internal carotid artery (ICA) stenosis, significantly more posterior CoW hypoplasia is seen in those with ischemic lesions [4]. A complete CoW is present in only about 11% of symptomatic carotid patients versus 55% in controls, indicating that incomplete variants correlate with higher stroke risk [5].



Figure 1.3: Time-of-flight MRA at the level of the Circle of Willis.

- **Intracranial Aneurysms:** About 90% of saccular aneurysms occur at Circle of Willis branch points, especially at the anterior communicating artery (ACoA), posterior communicating artery (PCoA), and ICA terminus. Variants like fetal-type PCA and A1 hypoplasia increase wall shear stress and promote aneurysm formation and rupture [6].
- **Leukoaraiosis (White Matter Lesions):** Incomplete CoW configurations, particularly in the posterior portion, are associated with up to 58% greater volume of white matter disease compared to complete circles [5, 4].
- **Carotid Surgery Risk:** Patients lacking both the ACoA and ipsilateral PCoA segments have a more than tenfold increased risk of stroke or TIA during carotid endarterectomy if no protective shunt is used [4].

1.2.2 Venous System

Cerebral venous drainage is mediated by an interconnected network of superficial and deep veins that ultimately converge into the dural venous sinuses. The superficial venous system, including the superior cerebral veins, primarily drains the cortical surfaces into the superior sagittal sinus. This major venous channel courses along the midline within the falx cerebri and constitutes the principal outflow pathway for much of the cerebral cortex. The deep venous system centers on the internal cerebral veins, which join with the basal veins of Rosenthal to form the great cerebral vein (vein of Galen). This short yet clinically significant midline vessel drains into the straight sinus, which then merges with the superior sagittal sinus at the confluence of sinuses. From this confluence, venous blood is directed laterally into the paired transverse sinuses, which continue as the sigmoid sinuses before ultimately exiting the cranial cavity through the internal jugular veins.

Additional venous pathways include the cavernous sinuses, situated bilaterally adjacent to the sella turcica. These sinuses receive venous return from the ophthalmic veins and communicate with the posterior venous system via the superior and inferior petrosal sinuses, thereby providing important collateral drainage routes. Functionally, the superficial venous system drains most of the cerebral cortex, while the deep venous system clears blood from the internal brain structures. Together, they ensure efficient removal of deoxygenated blood and metabolic by-products, thereby preserving intracranial homeostasis.

From a clinical perspective, the cerebral venous system is highly relevant to disease processes. Obstruction of venous outflow, as occurs in cerebral venous sinus thrombosis, can lead to increased intracranial pressure, venous infarction, and hemorrhage. The anatomical connections of the cavernous sinuses also predispose them to septic thrombophlebitis, since infections originating in the facial “danger triangle” can propagate intracranially via the ophthalmic veins. Moreover, malformations of the deep venous system, such as vein of Galen aneurysmal malformations, are associated with severe hemodynamic consequences in neonates and infants. Thus, an appreciation of both the anatomy and the vulnerabilities of cerebral venous drainage is fundamental for understanding normal physiology and a broad spectrum of neuropathological conditions [7].

1.3 Image Modalities

Imaging modalities form the foundation for brain vessel segmentation, as the accuracy and robustness of computational algorithms depend heavily on the quality, resolution, and contrast of the acquired data, since these constitute the input of the segmentation pipeline. In this section, we provide an overview of the most relevant imaging techniques for vascular analysis, each of which will be described in more detail in the following subsections. While the techniques differ in terms of physical principles, acquisition speed, invasiveness, and achievable resolution, they all play a complementary role in shaping the current landscape of vessel imaging. In particular, conventional Magnetic Resonance Imaging (MRI) provides structural and, in some cases, functional information about the brain and its vasculature; Time-of-Flight Magnetic Resonance Angiography (TOF-MRA) enables non-invasive visualization of blood flow with strong vessel-to-background contrast; and Computed Tomography (CT), particularly in its angiographic form (CTA), offers rapid, high-resolution volumetric data of both large and small cerebral vessels. The following subsections highlight the unique strengths and limitations of these approaches, laying the groundwork for understanding their impact on segmentation performance and clinical applicability.

1.3.1 Techniques

Among the most widely used techniques are conventional Magnetic Resonance Imaging (MRI), Time-of-Flight Magnetic Resonance Angiography (TOF-MRA), and Computed Tomography (CT), each with distinct physical principles, advantages, and limitations.

Magnetic Resonance Imaging (MRI) in a broader sense provides structural and, in some cases, functional information about the brain and its vasculature. MRI is free of ionizing radiation and offers excellent soft tissue contrast, which makes it valuable for integrating vessel segmentation into multi-modal brain analysis, such as combining vascular maps with lesion or tumor segmentation. Nevertheless, MRI's lower spatial resolution compared to CT and its susceptibility to motion artifacts can pose challenges in small vessel mapping.

Time-of-Flight Magnetic Resonance Angiography (TOF-MRA) is a non-invasive MRI technique specifically designed to visualize flowing blood without the need for contrast agents. It exploits flow-related enhancement, where unsaturated spins from inflowing blood generate higher signal intensity compared to stationary tissue. This method is particularly sensitive to medium-to-large arteries in the brain, offering high-contrast vessel delineation while avoiding ionizing radiation. However,

its sensitivity decreases for slow or turbulent flow, and very small distal vessels may appear attenuated or absent. Motion artifacts and long acquisition times can also affect segmentation quality.

Computed Tomography (CT), particularly when combined with iodinated contrast in CT angiography (CTA), delivers high-resolution volumetric data with excellent visualization of both large and small cerebral vessels. CTA is fast, widely available, and highly effective in detecting acute vascular pathologies such as aneurysms, stenoses, or occlusions. In segmentation tasks, CT's high spatial resolution allows accurate boundary definition of vessel lumens, even in tortuous or branching regions. However, the use of ionizing radiation and contrast agents introduces risks, especially for patients with renal impairment or allergies, and calcifications in vessel walls may complicate automated segmentation by appearing similar to contrast-filled lumens.

In summary, TOF-MRA excels in non-invasive, radiation-free vessel imaging with strong artery-to-background contrast; MRI offers complementary structural and tissue information that can enhance vessel context and multi-modal segmentation; and CT provides rapid, high-resolution vascular mapping ideal for acute settings. The choice among these modalities for brain vessel segmentation depends on the clinical scenario, target vessel size, patient safety considerations, and the specific requirements of the segmentation algorithm, such as tolerance to noise, resolution needs, and contrast characteristics.

1.3.2 Preprocessing

In medical imaging, preprocessing constitutes a critical initial step, as most computational algorithms require standardized input data to function reliably. Raw images acquired from scanners frequently exhibit variability in dimensions, voxel spacing, orientation, and intensity profiles. Such heterogeneity can significantly hinder downstream analysis, rendering preprocessing indispensable to ensure consistency, comparability, and robustness prior to the application of advanced image processing or machine learning algorithms.

Image dimension refers to the number of pixels or voxels along each axis. Since scans can vary in size depending on the field of view or acquisition protocol, resizing or cropping is typically required to standardize the input shape.

Resizing changes the number of pixels or voxels in an image to match a target shape, regardless of the original physical spacing. For example, a 256×256 MRI slice can be resized to 128×128 or 512×512 simply by scaling the pixel grid. This is usually done with interpolation methods such as nearest neighbor, bilinear, or bicubic

interpolation. Importantly, resizing does not preserve the physical dimensions of the anatomy, it just ensures that all images have the same array size, which many algorithms (especially deep learning models) require.

Cropping, in contrast, reduces the image size by cutting away unwanted parts of the image while keeping the original pixel or voxel resolution intact. This is useful when large parts of the scan contain background or irrelevant regions.

You can either combine the operations or just one of those, according to the task setup.

Image dimension refers to the number of pixels or voxels along each axis of an image. Since medical scans vary in size depending on acquisition protocols and field of view, dimension standardization is often required prior to analysis. Two common approaches are resizing and cropping.

Resizing adjusts the image matrix to a target shape by resampling the pixel or voxel grid, irrespective of the original physical spacing. For instance, a 256×256 MRI slice can be rescaled to 128×128 or 512×512 using interpolation methods such as nearest-neighbor, bilinear, or bicubic interpolation. While this ensures uniform input dimensions for algorithm, particularly deep learning models, resizing alters the physical resolution of the anatomy, potentially affecting geometric fidelity.

Cropping, by contrast, reduces image size by removing peripheral regions while maintaining the original pixel or voxel resolution. This approach is particularly useful when scans contain large background areas or anatomically irrelevant regions. In practice, resizing and cropping may be applied individually or in combination, depending on the specific requirements of the preprocessing pipeline and the downstream computational task.

Spacing, or *voxel size*, indicates the physical distance represented by each pixel or voxel. Different scanners and protocols may produce images with anisotropic (the voxel dimensions are not equal along all three axes) or varying resolutions, so resampling to isotropic (voxel spacing is equal across all axes) and uniform voxel spacing is often performed to make images comparable.

Resampling in medical imaging refers to changing the voxel spacing of an image so that the physical dimensions it represents become standardized. Every voxel in a medical image corresponds to a real world distance, such as $0.7 \text{ mm} \times 0.7 \text{ mm} \times 5 \text{ mm}$. When scans come from different machines or protocols, these voxel sizes can differ, leading to anisotropic images where one axis has much lower resolution than the others. This inconsistency can cause problems for algorithms that expect isotropic and comparable image data. To fix this, resampling interpolates the voxel grid to a new spacing, often chosen to be uniform, like $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$. During resampling, the image is recomputed so that the same anatomy is represented with the new voxel dimensions.

Orientation in medical imaging describes how the 3D volume is aligned with respect to the anatomical axes (e.g., left-right, anterior-posterior, superior-inferior). Different scanners and acquisition protocols can store images in varying orientations such as LPS (Left–Posterior–Superior) or RAS (Right–Anterior–Superior). If left unstandardized, these differences can lead to significant misinterpretation during both visualization and model training: for example, a brain magnetic resonance flipped left to right would confuse an algorithm. To address this, preprocessing pipelines often reorient all images to a consistent coordinate convention, such as RAS. This ensures that anatomical structures always appear in the same orientation across the dataset.

Intensity represents the signal values within the image. Because these intensities are not consistent across scans, normalization techniques such as z-score scaling are used to bring them to a comparable range. For example, in brain MRI one scan might have gray matter intensities centered around 600 while another has them around 1200 due to different scanner settings. After z-score normalization, both scans would be transformed so that the mean intensity becomes 0 and the standard deviation becomes 1, making the relative differences between tissues comparable across patients.

Standardizing image dimension, spacing, intensity and orientation ensures that medical images can be processed reliably by machine/deep learning algorithms. Often, operations are performed in a precise cascading order to obtain the desired standardized image.

Given the critical importance of establishing a robust preprocessing pipeline, this work extensively leverages MONAI, an open-source framework specifically developed for medical imaging. MONAI provides a comprehensive suite of tools that enable the implementation of preprocessing operations in a structured, reproducible, and modular manner. Beyond preprocessing, MONAI integrates seamlessly with PyTorch, offering specialized modules for model architectures, loss functions, and evaluation metrics that are tailored to the unique requirements of medical imaging tasks [8]. Its community-driven design, extensive functionality, and focus on reproducibility make MONAI a natural and effective choice for the development and deployment of advanced computational pipelines in this domain.

1.4 Segmentation

Segmentation is a central task in medical image analysis, as it enables the precise delineation of anatomical structures and provides the foundation for quantitative evaluation and computational modeling. In the context of cerebral vasculature, segmentation allows for the extraction of both global vessel trees and individual branches, which are subsequently analyzed for morphological, functional, or pathological insights. This section begins by introducing the formal definition of binary segmentation, followed by a discussion on how segmentation quality is evaluated in supervised settings, where ground-truth annotations are available. Finally, commonly used evaluation metrics are presented, with particular attention to those relevant for vascular imaging.

1.4.1 Definition

Binary segmentation aims to partition the image domain into vessel and background regions. Formally, let $\Omega \subset \mathbb{R}^2$ (or $\mathbb{R}^3$) denote the image domain, such as that of a CTA or MRA scan. Define a binary label function:

$$L : \Omega \rightarrow \{0,1\},$$

where

$$L(x) = \begin{cases} 1, & \text{if pixel/voxel } x \text{ belongs to a blood vessel,} \\ 0, & \text{otherwise.} \end{cases}$$

The vessel region is then

$$R_{\text{vessel}} = \{x \in \Omega : L(x) = 1\},$$

while the background is defined as

$$R_{\text{background}} = \Omega \setminus R_{\text{vessel}}.$$

In neurovascular imaging, segmentation models are trained to approximate L as accurately as possible, enabling the reconstruction of the vascular tree at multiple scales.

1.4.2 Evaluation with Ground-Truth Annotations

When ground-truth annotations are available, segmentation quality is assessed by comparing predicted masks with the reference labels. Let $I \in \mathbb{R}^{H \times W \times D}$ denote the input image (e.g., a 3D MRA volume), where H , W , and D represent height, width, and depth. A model f_θ , parameterized by θ , produces a probability map:

$$\hat{Y} = f_\theta(I), \quad \hat{Y} \in [0,1]^{H \times W \times D}.$$

A binary segmentation mask $\tilde{Y}$ is obtained by thresholding at level $\tau \in [0,1]$:

$$\tilde{Y}_i = \begin{cases} 1, & \text{if } \hat{Y}_i \geq \tau, \\ 0, & \text{otherwise.} \end{cases}$$

Given the ground-truth segmentation $Y \in \{0,1\}^{H \times W \times D}$, performance can be expressed in terms of voxel-level comparisons. For N total voxels, we define:

$$\begin{aligned} \text{TP} &= \sum_{i=1}^N \mathbb{1}(Y_i = 1 \wedge \tilde{Y}_i = 1), & \text{FP} &= \sum_{i=1}^N \mathbb{1}(Y_i = 0 \wedge \tilde{Y}_i = 1), \\ \text{FN} &= \sum_{i=1}^N \mathbb{1}(Y_i = 1 \wedge \tilde{Y}_i = 0), & \text{TN} &= \sum_{i=1}^N \mathbb{1}(Y_i = 0 \wedge \tilde{Y}_i = 0), \end{aligned}$$

where TP (true positives) are correctly segmented vessel voxels, FP (false positives) are background voxels misclassified as vessels, FN (false negatives) are missed vessel voxels, and TN (true negatives) are correctly identified background voxels.

1.4.3 Common Evaluation Metrics

Using TP, FP, FN, and TN, several broadly used metrics can be defined to evaluate segmentation performance.

It is very important to underline that each captures a different aspect of quality, and combining them provides a more comprehensive assessment.

- **Dice Similarity Coefficient (DSC):**

$$\text{DSC}(Y, \tilde{Y}) = \frac{2 \text{TP}}{2 \text{TP} + \text{FP} + \text{FN}}$$

Measures spatial overlap between prediction and ground truth; widely used in medical imaging as it balances false positives and false negatives.

- **Jaccard Index (Intersection-over-Union, IoU):**

$$\text{IoU}(Y, \tilde{Y}) = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

Similar to DSC but more conservative, penalizing errors more strongly when overlap is low.

- **Precision (Positive Predictive Value):**

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Indicates the fraction of predicted vessel voxels that are correct; high precision means few false positives.

- **Recall (Sensitivity, True Positive Rate):**

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Indicates the fraction of true vessel voxels correctly detected; high recall means few false negatives.

- **Accuracy:**

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Proportion of correctly classified voxels overall, though it may be misleading in highly imbalanced problems such as vessel segmentation.

- **F1-Score:**

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Equivalent to the Dice coefficient in binary segmentation; balances precision and recall.

- **Hausdorff Distance (HD):**

$$d(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|_2, \quad \text{HD}(Y, \tilde{Y}) = \max\{d(\partial Y, \partial \tilde{Y}), d(\partial \tilde{Y}, \partial Y)\}$$

Captures worst-case boundary errors by measuring the maximum surface-to-surface distance between predicted and ground-truth boundaries.

- **Centerline Dice Coefficient (clDice):**

$$\text{clDice}(Y, \tilde{Y}) = \frac{2 |S(Y) \cap \tilde{Y}| \cdot |S(\tilde{Y}) \cap Y|}{|S(Y)| + |S(\tilde{Y})|}$$

Tailored for vessel segmentation, it evaluates topological consistency by comparing skeletonized structures.

1.4.4 State-of-the-art

Vascular segmentation has evolved rapidly, from classical image processing to modern deep learning.

This section provides an overview of the main approaches, organized by supervision strategy: unsupervised, supervised, and semi-supervised methods. Each category highlights the key concepts, strengths, and limitations relevant for cerebrovascular segmentation.

Unsupervised segmentations methods

Before the rise of machine learning, numerous approaches were proposed for vascular segmentation. A classical review by Lesage et al. [9] categorized these methods into three complementary axes: appearance and geometric models, image features, and extraction algorithms.

Appearance and geometric models incorporate prior knowledge about vessel shape and intensity. Appearance models rely on intensity profiles specific to the imaging modality (e.g., CT angiography), while geometric models describe cross-sectional shapes, centerlines, or both [10]. Hybrid approaches combine both, for instance by modeling cross-sectional intensity with Gaussian or bar-like profiles. Image features are derived from intensity or its derivatives. Isotropic features treat vessels without directional priors [11], whereas anisotropic features leverage second-order derivatives to detect tubular structures. Several methods focus on cross-sectional edges, using ray projections [12] or active contours [13]. Extraction algorithms operationalize segmentation. Vessel enhancement filters highlight tubular structures before thresholding or further processing [14]. Region-growing methods iteratively expand from seed points [15], sometimes with directional constraints. In summary, pre-learning segmentation relied on a combination of prior modeling, image-derived features, and algorithmic extraction. While surpassed by modern deep learning methods in terms of accuracy and robustness, these classical approaches remain conceptually relevant and continue to inspire hybrid or initialization strategies in current research.

Other unsupervised methods aim to segment vascular structures without requiring annotated datasets. Most approaches rely on self-supervised learning by generating synthetic training data: artificial vessel masks are fused with angiographic images to create realistic training samples. For instance, Ma et al. (2021) [16] proposed an adversarial learning framework with two generators: one synthesizes angiographic images containing vessels at positions defined by binary masks, while the other learns to segment these images, trained to produce masks indistinguishable from ground truth. Kim et al. (2022) [17] adopted a similar strategy using diffusion models, later extending it with contrastive learning to improve vessel representations.

Shi et al. (2023) [18] explored fusing vessel masks with angiographic images in the Fourier domain. Overall, unsupervised vascular segmentation methods provide a promising avenue to reduce reliance on costly annotations, but their effectiveness strongly depends on the realism of the synthetic data and the generative models employed.

Supervised segmentations methods

We found ourself in the supervised setup setting, already explained in 1.4.2. Supervised learning has become the dominant paradigm for vascular segmentation, achieving the highest levels of accuracy when sufficient annotated data are available. These methods leverage deep neural networks trained on labeled angiographic images, and they can be grouped into general-purpose architectures and those specifically adapted to vascular data. The breakthrough came with U-Net [19], whose encoder-decoder structure with skip connections enabled precise localization while preserving contextual information. Many variants followed, such as U-Net++ [20] with redesigned skip pathways, residual U-Nets for deeper architectures, attention-gated U-Nets to focus on vascular regions, and 3D U-Nets that better capture volumetric continuity. Together, these models improved segmentation robustness across multiple scales, a crucial property for detecting both large arteries and fine peripheral vessels. Beyond convolutional networks, transformer-based models have recently been explored. Architectures like TransUNet [21] and Swin-UNet [22] integrate self-attention mechanisms, allowing the network to capture long-range dependencies between distant vascular segments. This is particularly relevant for cerebrovascular networks, where vessels form tortuous, branching structures that extend across wide regions. Several supervised methods have been designed explicitly for vascular applications. DeepVesselNet [23] introduced specialized blocks for analyzing tubular structures, while CS2-Net [24] improved the capture of multiscale connectivity. Other works embed topological priors directly into the learning process. For instance, the cLDice loss [25] leverages skeleton representations to encourage connectivity preservation. Some approaches adopt multitask frameworks, training networks to simultaneously predict vessel segmentations and skeletons, thereby reinforcing topological consistency. Despite these advances, supervised methods face two persistent limitations:

- **Data dependency:** Annotating vascular networks is extremely demanding, requiring voxel-level precision across 3D volumes. The scarcity of annotated datasets remains a major bottleneck for training large models and for generalization across modalities or patient populations.

- Topological errors: Classical loss functions (e.g., Dice, cross-entropy) optimize voxel-level accuracy but do not penalize clinically critical errors such as broken vessels or missing branches. Even high Dice scores can correspond to unusable segmentations in practice

In summary, supervised deep learning approaches define the state of the art in vascular segmentation, combining high segmentation accuracy with the flexibility to adapt to complex geometries. However, their reliance on costly annotations and their limited ability to guarantee topological correctness have spurred research into semi-supervised methods and topology-aware constraints, which seek to overcome these shortcomings.

Semi-supervised Segmentation Methods

Semi-supervised segmentation aims to overcome the scarcity of annotated medical data by combining a small labeled set $D_l = (X, Y)$ with a larger unlabeled set $D_u = X'$. Unlike purely supervised training, these methods exploit both sources to improve robustness and generalization. Jiao et al. [26] group them into three main families: Knowledge-based approaches, Pseudo-labeling approaches, Unsupervised regularization approaches.

Knowledge-based approaches inject anatomical priors (shape, topology, spatial coherence) into the learning process. Zheng et al. [27], for example, proposed an adversarial framework where a discriminator evaluates whether predicted segmentations remain anatomically plausible, guiding the model beyond the limited supervision.

Pseudo-labeling methods iteratively generate artificial labels for unlabeled data using a model trained on the labeled set D_l . The model predicts pseudo-labels Y_{pl} for D_u , forming an augmented dataset $D_{pl} = (X', Y_{pl})$, which is then combined with D_l for retraining until convergence. To reduce noise in pseudo-labels, strategies include confidence-based filtering [28], thresholding [29], or averaging predictions over iterations [30]. Some methods propagate annotations via prototypes, generating labels based on similarity to labeled examples [31].

Instead of producing labels, Unsupervised regularization approaches impose constraints directly on predictions for unlabeled images. A prominent example is the Mean Teacher framework [32], where a student network learns under supervision from D_l while being encouraged to remain consistent with a teacher network on D_u . Yu et al. (2019) [33] extended this with transformation consistency, ensuring stable predictions under perturbations.

Semi-supervised segmentation reduces annotation requirements by combining priors, pseudo-labeling, or regularization on unlabeled data. In vascular applications, these methods are particularly valuable for maintaining topology and connectivity, which are essential for downstream clinical tasks such as blood flow simulation or surgical planning.

1.4.5 Encountered models

However, segmentation remains a highly variable and architecture sensitive problem. Different models exhibit diverse behaviors depending on their design, training regimes, and the imaging modalities used.

Throughout the course of this work, several algorithms were explored: **nnU-Net**, **A2V**, **JoB-VS**, and **SP0CKMIP**. For each model, the architecture, main strengths, performance, and dataset are summarized.

nnU-Net

nnU-Net is a fully convolutional U-Net style encoder–decoder that automatically configures its architecture and preprocessing for a given dataset. It uses a four level downsampling encoder (stacked convolutional blocks) and a mirrored upsampling decoder with skip connections, transposed convolutions, and deep supervision at multiple decoder stages.

The framework auto adjusts network depth, feature channels and other hyperparameters per task, eliminating manual tuning [34].

A key advantage of nnU-Net is its “out-of-the-box” adaptability: it self-optimizes network architecture and learning parameters for diverse medical imaging datasets. This yields state-of-the-art segmentation accuracy without manual design, and robust performance across modalities. The deep supervision and large receptive fields ensure stable training and good boundary localization.

For vessel tasks, recall often remains high with minimal false negatives.

In the context of brain vessel segmentation, nnU-Net based architectures have consistently demonstrated strong baseline performance, even without modifications of the task specific architecture. For example, in the CAS2023 TOF-MRA cerebral artery segmentation challenge, **nnU-Net** achieved a Dice similarity coefficient in the range of 85–86%, reflecting a high degree of spatial overlap between predicted and reference vessel masks. Complementary metrics also indicated reliable segmentation quality, with a Hausdorff distance at the 95th percentile (HD95) of approximately 48.2 mm, a recall of around 83.5%, and a precision of about 87.8%.

These results, obtained in a multi-centre, multi-scanner setting, confirm the model’s robustness and its suitability as a strong reference method for cerebral vessel delineation tasks [35].

A2V (Angiography-to-Venography)

A2V is a two stage, semi-supervised framework based on generative adversarial models.

In Phase 1, a single generator (G) and discriminator (D) are trained in a StyleGAN2 architecture to model brain angiography/venography appearance.

In Phase 2, a single encoder network is trained to invert real images into the shared StyleGAN latent space, enabling image-to-image translation between angiograms and venograms. The generator’s “label-synthesis” branch (from DatasetGAN) directly predicts vessel masks from the latent features, avoiding a separate segmentation model.

Overall, A2V uses just one generator, one discriminator, and one encoder (no cycle of paired generators), with a disentangled latent code that separates vessel appearance (arteries vs veins) from spatial anatomy [36].

The model effectively bridges modality gaps with minimal labeled data in the target domain, and simplifies training by using only three networks instead of more complex cycle architectures.

Trained on OASIS-3 TOF-MRA angiograms and adapted to SWI venograms, A2V achieved high vessel segmentation accuracy in both domains, with only an $\approx 8.9\%$ Dice drop in the venography target domain compared to the angiography source domain [36].

JoB-VS (Joint Brain Vessel Segmentation)

JoB-VS is a 3D multitask network built on a lattice structured U-Net (ROG) backbone. It starts with an “initial module” followed by a triangular lattice of feature nodes that connects multi resolution paths, preserving high resolution details for small vessels while also capturing large scale context.

Critically, JoB-VS has a dual-head segmentation head: one branch predicts the brain mask and the other predicts the vessel mask simultaneously.

The loss is computed jointly on both outputs (sum of Dice+CE losses) to guide the network [37].

By segmenting brain tissue and vessels simultaneously, JoB-VS eliminates the need for skull stripping, resulting in an end-to-end pipeline.

Multitask learning improves sensitivity to small vessels.

On OASIS-3 TOF-MRA, JoB-VS achieved a mean average precision of 70.03%, maximum F1 (Dice) of 69.09%, and cIDice of 74.6% without brain mask preprocessing [37].

VesselBoost

VesselBoost is a modular Python toolbox for small vessel segmentation in ultra-high field MR angiography (TOF-MRA). It is based on a modified 3D U-Net with 4 encoder/decoder layers, lightweight convolutional blocks, and Tversky loss for imbalanced vessel data.

The framework is organized into three modules: Predict, Test-Time Adaptation (TTA), The Booster. Predict is a pre-trained network applied directly to MRA data with preprocessing (bias-field correction, denoising) and post-processing (thresholding, cluster removal).

Test-Time Adaptation (TTA) fine-tunes the pre-trained model using a proxy segmentation (either user-provided or generated via predict), improving small-vessel delineation.

The Booster trains a new segmentation model from scratch using imperfect labels, leveraging a data augmentation strategy that creates small-vessel training examples by zooming out large vessels and applying rotation/flipping transformations.

VesselBoost addresses the scarcity of manually segmented training data by learning effectively from imperfect labels. Its augmentation strategy stabilizes training and enhances sensitivity to small vessels. The modular design allows flexibility, users can either apply pre-trained models, adapt them to new data, or build new models for different contrasts.

Trained on the SMILE-UHURA 7T TOF-MRA dataset (300 μm resolution), VesselBoost generalized well to unseen resolutions (150–400 μm). In quantitative evaluation, TTA improved Dice scores by ≈ 0.04 compared to coarse proxy segmentations. Booster outperformed simple threshold-based labels by recovering finer small vessels. An ablation study confirmed that zoom-based patch augmentation with rotation and flipping yielded the most stable results, achieving segmentation performance close to manually corrected ground truth [38].

SPOCKMIP

SPOCKMIP is built on a 3D UNet-MSS (multi-scale supervision) backbone (inspired by DS6). Concretely, it uses the same encoder–decoder U-Net with multiple output scales (MSS) but replaces ReLU with LeakyReLU activations.

The key innovation is the MIP loss: during patch-based training, SPOCKMIP computes the maximum intensity projections (MIPs) of the predicted and ground-truth vessel masks. A loss term is added that compares these MIPs along the axial (z) direction (single-axis MIP) and/or cumulatively across all three axes. This augments the standard Dice/CE loss to explicitly penalize discontinuities in vessel segments across slices.

Apart from the MIP loss, SPOCKMIP uses a standard patch-based UNet training pipeline (5-fold CV) [39].

The MIP loss penalizes segmentation discontinuities, resulting in smoother and more anatomically consistent vessel trees, especially in high-resolution data.

On the StudyForrest 7T TOF-MRA dataset, the multi-axis MIP variant achieved a median Dice score of 80.25%, outperforming the single-axis variant (79.75%).

1.4.6 Why Brain Segmentation Matters

Accurate segmentation of intracranial blood vessels is essential for diagnosing and treating cerebrovascular disorders.

Deshpande et al. note that high-fidelity vessel delineation “is critical to various diagnostic and therapeutic purposes” (and to studying brain health) [40]. By converting angiograms (CT/MRI) into quantitative 3D maps, clinicians can systematically analyze vessel geometry (length, diameter, tortuosity) and detect subtle pathologies. Modern deep learning algorithms achieve high accuracy in vessel segmentation [41], automating complex tasks like anomaly detection and enabling personalized assessment of cerebral blood flow.

- **Stroke and Ischemic Disease**

In acute stroke, vessel segmentation helps identify arterial occlusions and evaluate collateral flow. Quantitative vessel models reveal pathology: for example, Deshpande et al. found that stroke patients have significantly reduced vessel length/volume but increased tortuosity and fractal dimension compared to controls [40]. Precisely delineating major arteries (e.g., middle cerebral artery and its lenticulostriate branches) is essential for treatment decisions. Zhou et al. emphasize that segmenting these arteries “is key for effectively managing stroke ... and for improving patient neurological outcomes” [42]. Automated segmentation thus enables rapid detection of occlusions and supports planning of thrombectomy or thrombolysis.

- **Intracranial Aneurysms** Segmentation is crucial for detecting and characterizing aneurysms, which are often asymptomatic until rupture. Accurate delineation of the aneurysm sac and parent vessels allows precise measurement of aneurysm size, neck width, and shape. Manual identification and segmentation of intracranial aneurysms is known to be “labor intensive and prone to human errors” [43]. Such challenges underscore the need for automated detection and segmentation systems with the aim of providing reproducible measurements of aneurysm geometry, supporting rupture risk assessment.

- **Monitoring and Risk Stratification**

Automated vessel segmentation also enhances disease monitoring and risk

stratification.

Quantitative metrics derived from segmentation (vessel volume, tortuosity, branching patterns) can be tracked over time to assess progression or treatment response. For example, an increase in aneurysm volume or a change in arterial tortuosity can indicate growth or instability. As previously said, Deshpande et al. showed that vessel tortuosity and fractal dimension (from segmentation) differ in stroke, suggesting these features could serve as risk markers [40].

By providing consistent, quantitative measures, segmentation enables objective follow-up and stratification (e.g., predicting stroke risk based on arterial morphology).

- **Efficiency and Reproducibility in Clinical Workflows**

Finally, automated segmentation greatly improves efficiency and reproducibility. Manual tracing of brain vessels is tedious and subject to inter-operator variability.

Deep learning models can segment entire angiographic scans in minutes with expert level accuracy [41] [42]. Such automation reduces reporting time and eliminates subjective bias, leading to faster diagnosis and allowing large-scale analyses of cerebrovascular morphology.

1.5 Quality Control, the other medal’s face

Even state-of-the-art networks like nnU-Net can produce imperfect segmentations under challenging conditions (e.g., low image quality or unusual anatomy). In practice, factors such as image noise, motion artifacts, or scanner differences can degrade segmentation accuracy.

Likewise, extreme anatomical variability (major cause) or rare pathologies that were not well represented in the training data may lead to spurious contours or missing regions. A robust quality-control stage is therefore needed to catch such failures before they affect downstream analysis.

- **Image quality issues:** Noise, poor contrast, and artifacts can obscure anatomical structures, confusing the model and resulting in inaccurate or incomplete segmentations.
- **Anatomical variability:** The cerebral vasculature exhibits substantial natural variability across individuals. Also cases with pathological changes (e.g., aneurysms, arteriovenous malformations), or surgical alterations may deviate significantly from the training distribution, leading to inaccurate or incomplete segmentations.

- **Model limitations:** Any learned model has blind spots. Out-of-distribution cases, unusual intensities, or some anatomical variations can trigger gross errors (e.g., missing thin vessels or false positives).
- **Acquisition differences:** Variability in scanners, protocols, or preprocessing steps may degrade a model’s performance on new data.

Given the potential for errors in automated segmentation, in particular for the before cited reasons, it is essential for the pipeline to include mechanisms that automatically evaluate the reliability of each output. A quality-control (QC) module can implement both basic and advanced checks. Basic sanity checks may verify that the segmented volume falls within an expected range or exhibits a plausible anatomical shape, while more sophisticated approaches can leverage model-based uncertainty estimates or assess disagreements across ensemble predictions. Additionally, key statistical features of the segmentation can be compared against normative values derived from reference populations. Cases that deviate from these expected ranges can be flagged for further review, ensuring that anomalous results are identified promptly. A well-designed QC pipeline thus complements the neural network by detecting errors and anomalies early, providing confidence that downstream analyses, such as quantitative measurements, clinical evaluations, or diagnostic decisions, are based on accurate and reliable segmentations. In summary, robust quality control is indispensable for guaranteeing the validity of automated vessel segmentation outputs and for highlighting any results that require manual inspection.

1.6 Objectives and Contributions

Ensuring the quality of segmentation results is critical in medical imaging, but it is challenging to assess quality on new data without ground truth.

In brain MRI analysis, manual quality control remains the gold standard but is impractical for large studies; automated QC tools (e.g: MRIQC [44], Qoala-T [45]) have been proposed. However, none of these strategies have been applied to cerebrovascular segmentation, which is especially challenging due to the thin, branching nature of brain vessels.

Our literature survey found virtually no dedicated QC workflow for brain vessel segmentation. In fact, only a few automated QC methods exist for medical image segmentation, and these have focused on organs like the prostate or heart. The lack of prior work on vascular QC underscores the need for the novel approaches we investigate.

The main contributions of this work are as follows:

- **Vessel metrics toolkit.** We developed a flexible software framework for computing topological and morphological descriptors of vessel segmentations, grounded in solid mathematical foundations. This toolkit can process any segmentation of the cerebral vasculature, whether manual or automatic, and serves both as the basis for quality control (QC) pipelines and as a general-purpose tool for vascular analysis.
- **Feature-based regression QC.** Predictive models are trained on the extracted segmentation features to estimate segmentation quality. This approach offers both interpretability and computational efficiency, as it relies on explicit, biologically meaningful descriptors. However, its effectiveness depends on how well these features capture relevant segmentation errors and may require model-specific examples for optimal training.
- **Generative reconstruction QC.** We employ generative models, such as three-dimensional variational autoencoders (VAEs), to learn the distribution of plausible vessel segmentations. By reconstructing input segmentations, the method can identify anomalous or low-quality segmentations. This unsupervised strategy potentially eliminates the need for labeled error examples and is capable of detecting novel failure modes, though it is computationally demanding and sensitive to model tuning.
- **Comparative analysis of QC strategies.** We systematically compare the two QC approaches in terms of practical applicability, computational cost, and

detection performance. This analysis elucidates each method’s strengths and limitations in the vascular context, providing guidance for their integration into automated QC pipelines.

To the best of our knowledge, this is the first systematic investigation of automated QC methods for cerebral vessel segmentation. By adapting and evaluating these two methodological pathways within the vascular domain, we establish a foundation for future research aimed at developing robust and scalable QC procedures for this challenging imaging task.

Chapter 2

Tool metrics' computation

2.1 Chapter Overview

Quantitative characterization of vascular networks is a cornerstone in biomedical imaging, pathology, and neuroscience. The morphology and geometry of vessels provide essential information about vascular health, pathological remodeling, and the efficiency of blood supply in tissues. Subtle alterations in vascular structure often precede clinical symptoms, making morphometric and topological analysis an important tool for both research and diagnostics.

With the advent of high-resolution three-dimensional (3D) imaging modalities such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Optical Coherence Tomography (OCT), it is now possible to obtain volumetric datasets that capture intricate vascular architectures at multiple scales. However, these datasets are typically represented as binary masks following segmentation, and their size and complexity necessitate automated, reproducible, and robust analysis pipelines. Manual inspection or simplistic measures are insufficient to capture the full range of structural and functional information encoded in vascular networks.

This chapter presents the methodology and mathematical underpinnings of **VESSEL_METRICS**, a computational framework designed to perform comprehensive morphometric and topological analysis of 3D vessel masks. The software converts volumetric vessel segmentations into skeleton-based graph representations, enabling systematic computation of segment-level, component-level, and global descriptors. It integrates several processing stages culminating in the extraction of a wide set of structural and geometric features. These include classical morphometrics such as vessel length, radius, volume, and bifurcation counts, as well as higher-order descriptors such as tortuosity indices, fractal dimension, and lacunarity.

On top of **VESSEL_METRICS**, **ATLAS_VESSEL_METRICS** software has been developed. Within this tool, the metrics computation is made at regional-level, according to the vessel tree partition defined by a given *ATLAS*.

By unifying these analyses into a reproducible pipeline, practically explained in Figure 2.1, **VESSEL_METRICS.py** facilitates quantitative evaluation of vascular structure across experimental conditions, disease models, and imaging modalities. The framework is particularly suited for large-scale studies where consistency, reproducibility, and automation are paramount.

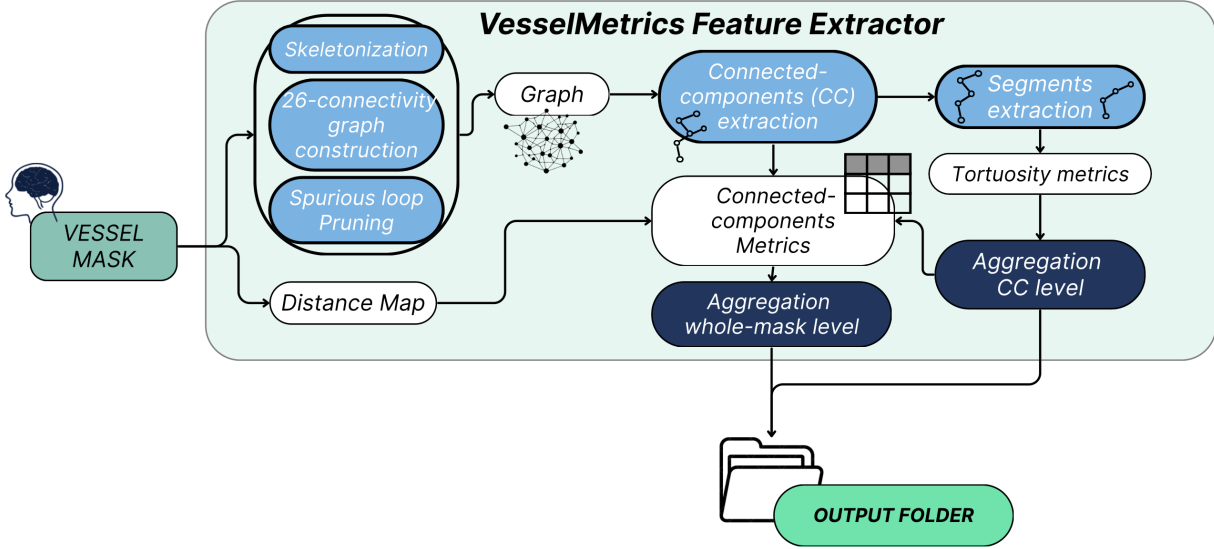


Figure 2.1: Tool's pipeline.

2.2 Processing Pipeline for Vessel Masks

The analysis pipeline processes a binary 3D vessel mask, where segmented vessels are represented by voxels labeled as 1 and the background as 0.

Notably, no additional cleaning operations (e.g., removal of minor disconnected islands) are applied.

This choice is deliberate, as the goal is to preserve the raw structure of the segmentation and analyze its intrinsic characteristics without introducing preprocessing biases.

Nevertheless, this strategy may not be universally appropriate. In certain applications, filtering out very small connected components can be beneficial, for example, to reduce the influence of segmentation noise, to improve computational efficiency, or when such small structures are known to be irrelevant for the biological question at hand.

The key stages are:

1. Skeletonization to extract centerlines
2. Distance transform for local radius estimation
3. Graph construction and pruning
4. Connected component decomposition

5. Extraction of vessel segments
6. Computation of morphometric and tortuosity metrics
7. Aggregation and output

2.2.1 Preliminary transformations

Skeletonization

Skeletonization reduces volumetric vessel data into a 1-voxel-thick centerline preserving topology. This step is essential to obtain vessel centerlines for subsequent graph construction.

Formally, given a binary mask $M : \Omega \subset \mathbb{Z}^3 \rightarrow \{0,1\}$, skeletonization computes a set $S \subseteq \Omega$ such that:

- S is connected,
- S is one-voxel thick (minimal thickness),
- The topology (connectivity and loops) of M is preserved,
- $S \subseteq M$.

This is achieved using algorithms like the thinning method implemented in *skimage.morphology.skeletonize*. The output S represents the medial axis of the vessel volume.

Estimation of Local Vessel Radius via Distance Transform

To quantify vessel thickness, a Euclidean distance transform is applied on the binary mask M , yielding a distance map $D : \Omega \rightarrow \mathbb{R}^+$:

$$D(x) = \min_{y \in \Omega \setminus M} \|x - y\|_2,$$

where $x \in M$. The value $D(x)$ approximates the radius of the vessel at voxel x assuming vessels are approximately circular in cross-section.

For skeleton voxels $s \in S$, their local radius is

$$r_s = D(s).$$

Graph Construction from Skeleton Voxels

Each skeleton voxel corresponds to a node in an undirected weighted graph $G = (V, E)$:

- Nodes $V = \{v_i\}$ represent skeleton points with 3D coordinates $\mathbf{x}_i \in \mathbb{Z}^3$.
- Edges E connect pairs of nodes (v_i, v_j) if their voxel coordinates are within a $3 \times 3 \times 3$ neighborhood (26-connectivity).
- Edge weights w_{ij} equal the Euclidean distance between nodes:

$$w_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2.$$

The graph captures vessel centerline topology and spatial relationships.

Pruning Triangular Loops

Skeleton graphs derived from 26-connectivity may introduce artificial triangular loops at bifurcations due to voxel adjacency patterns.

We observed that these spurious loops frequently arise at branching points as a direct consequence of the 26-connectivity scheme.

To preserve biologically meaningful loops while removing artifacts, the graph is pruned by eliminating the edge with the highest weight (i.e., the longest) in each 3-node cycle.

This procedure suppresses artificial loops while maintaining an accurate approximation of the true branching structure.

2.2.2 Connected Components Decomposition

The graph G may consist of multiple disconnected subgraphs, corresponding to separate vessel components. Connected components $\{G_k\}$ are identified via:

$$G = \bigsqcup_k G_k,$$

where each $G_k = (V_k, E_k)$ is maximal connected.

Total Length

The total length L_k of component G_k is the sum of all edge weights:

$$L_k = \sum_{(u,v) \in E_k} w_{uv}.$$

Number of Bifurcations

Nodes with degree $\deg(v) \geq 3$ are considered bifurcations. The count is:

$$B_k = |\{v \in V_k : \deg(v) \geq 3\}|.$$

Bifurcation Density

Density of bifurcations per unit length:

$$\rho_k = \frac{B_k}{L_k}.$$

Vessel Volume Estimation

Volume V_k approximated by treating each edge as a truncated cylinder with radius interpolated from node radii r_u, r_v :

$$V_k = \sum_{(u,v) \in E_k} \pi \left(\frac{r_u + r_v}{2} \right)^2 w_{uv}.$$

Number of Loops

Loops correspond to cycles in the graph. The cycle basis provides the number of independent loops:

$$C_k = |\text{cycle_basis}(G_k)|.$$

Abnormal Degree Nodes

Nodes with degree > 3 are considered abnormal, possibly representing artifacts:

$$A_k = |\{v \in V_k : \deg(v) > 3\}|.$$

Fractal Dimension Estimation

Vessel networks often exhibit fractal-like branching patterns. Fractal dimension D quantifies complexity beyond Euclidean dimensions.

Using the **box-counting method**:

1. Cover the skeleton voxels with cubic boxes of side length ϵ .
2. Count $N(\epsilon)$, the number of boxes containing any skeleton points.
3. Repeat for multiple scales ϵ .

Plotting:

$$\log N(\epsilon) \quad \text{vs} \quad \log \frac{1}{\epsilon},$$

the fractal dimension D is the slope of the linear fit:

$$\log N(\epsilon) = D \log \frac{1}{\epsilon} + c.$$

Lacunarity Calculation

Lacunarity Λ measures spatial heterogeneity or texture gaps in the vessel pattern, complementing fractal dimension.

1. Define box size

$$L = \frac{\max(\Delta x, \Delta y, \Delta z)}{10}.$$

2. Partition the vessel points into boxes of size L .
3. Compute the distribution of point counts n per box.
4. Calculate:

$$\Lambda = \frac{\text{Var}(n)}{\text{Mean}(n)^2} + 1.$$

A higher lacunarity indicates more heterogeneous spatial distribution.

2.2.3 Segment Extraction and Tortuosity Analysis

Vessel segments are meaningful subunits defined by paths between key nodes (roots).

Selection of Roots

Three root nodes per component G_k are selected:

- R_1 : Endpoint (degree=1) with largest diameter.
- R_2 : Endpoint with second-largest diameter.
- R_3 : Bifurcation node (degree ≥ 3) with largest diameter. If none, fallback to R_1 .

Segment Definition

Segments are shortest paths from each root to all reachable endpoints:

$$P_{r \rightarrow e} = \arg \min_{\text{paths } p} \sum_{(u,v) \in p} w_{uv}.$$

Segments correspond to vessel branches emanating from key nodes.

Tortuosity Metrics via Cubic B-Spline Fitting

Tortuosity serves as a key descriptor of the winding, bending, and overall geometric complexity of vascular structures.

Direct measurements on voxelized skeletons can be sensitive to noise and discretization artifacts; therefore, a smooth parametric representation is required.

In this work, which is inspired by [46], vessel centerlines are approximated using cubic B-spline fitting, which provides a continuous and differentiable curve representation. This formulation enables stable computation of curvature along the vessel segments' trajectory and facilitates the extraction of tortuosity metrics at multiple scales, including local curvature descriptors and aggregated root-level measures.

Input and Spline Fitting

Given N sample points

$$\mathbf{p}_i = (x_i, y_i, z_i) \in \mathbb{R}^3, \quad i = 0, \dots, N-1,$$

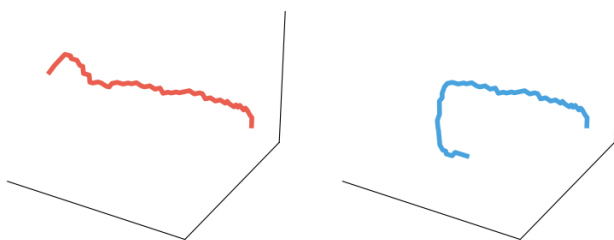
a cubic B-spline curve is fitted:

$$\mathbf{C}(u) = (x(u), y(u), z(u)), \quad u \in [0,1],$$

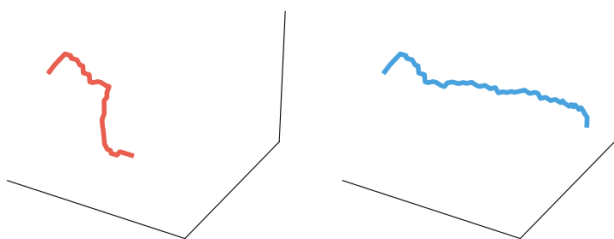
with smoothing factor $s \geq 0$. The parameter values are u_i with $u_0 = 0$, $u_{N-1} = 1$.



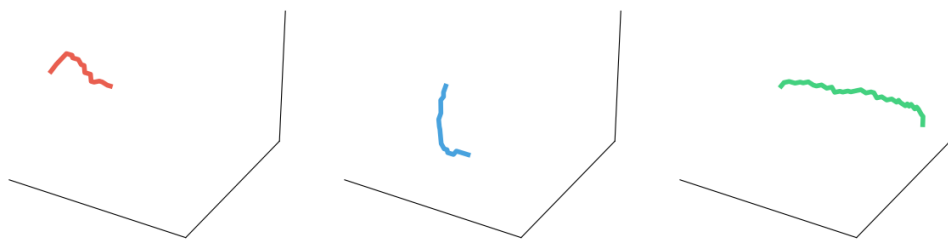
Trivial mask example.



a) Largest Endpoint as Root.



b) 2nd Largest Endpoint as Root.



c) Largest Bifurcation as Root.

Figure 2.2: Each row shows the segments from one policy splitting.

Arc-Length Reparameterization

The arc length as a function of u is

$$\frac{ds}{du} = \left\| \frac{d\mathbf{C}}{du} \right\|, \quad s(0) = 0,$$

and the total arc length is

$$L = s(1) = \int_0^1 \|\mathbf{C}'(u)\| \, du.$$

Let $s \in [0, L]$ and define the inverse map $u = u(s)$. Then we define uniformly sampled arc-length points

$$s_j = \frac{j}{n-1}L, \quad j = 0, \dots, n-1,$$

and corresponding points on the curve:

$$\mathbf{r}_j = \mathbf{C}(u(s_j)).$$

Derivatives

Using finite differences, compute the first and second derivatives with respect to arc length:

$$\mathbf{T}(s) = \frac{d\mathbf{C}}{ds}, \quad \mathbf{N}(s) = \frac{d^2\mathbf{C}}{ds^2}.$$

Curvature

The curvature at each point is given by:

$$\kappa(s) = \frac{\|\mathbf{T}(s) \times \mathbf{N}(s)\|}{\|\mathbf{T}(s)\|^3}.$$

Weighted Curvature

In some cases, certain points along the vessel skeleton may appear in multiple segments due to shared branches or overlapping paths. To avoid overemphasizing these points in the tortuosity computation, thereby reducing redundancy and bias, we introduce a weighting scheme based on the point's occurrence frequency.

Let $n_i = n(\mathbf{p}_i)$ be the number of segments containing the point $\mathbf{p}_i$. By interpolating these counts along the continuous arc-length parameter s , we obtain a smooth function $n(s) \geq 1$.

The weighted curvature is then defined as:

$$\tilde{\kappa}(s) = \frac{\kappa(s)}{n(s)}, \quad \tilde{\kappa}^2(s) = \frac{\kappa^2(s)}{n^2(s)}.$$

This weighting effectively reduces the contribution of points that are shared across multiple segments, ensuring that repeated structures do not disproportionately influence global tortuosity metrics. Consequently, the curvature measurements better reflect the overall geometric complexity of the vascular network without redundant inflation from overlapping segments.

Tortuosity Metrics

Define the following tortuosity metrics:

- **Spline arc length:**

$$\text{spline_arc_length} = L$$

- **Chord length:**

$$\text{spline_chord_length} = D = \|\mathbf{C}(1) - \mathbf{C}(0)\|$$

- **Mean weighted curvature:**

$$\text{spline_mean_curvature} = \bar{\kappa} = \int_0^L \tilde{\kappa}(s) ds$$

- **Mean square weighted curvature:**

$$\text{spline_mean_square_curvature} = \overline{\kappa^2} = \int_0^L \tilde{\kappa}^2(s) ds$$

- **RMS weighted curvature:**

$$\text{spline_rms_curvature} = \kappa_{\text{rms}} = \sqrt{\frac{1}{L} \int_0^L \tilde{\kappa}^2(s) ds}$$

- **Arc-over-chord ratio:**

$$\text{arc_over_chord} = \frac{L}{D}$$

- **Spline fit RMSE:**

$$\text{fit_rmse} = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} \|\mathbf{C}(u_i) - \mathbf{p}_i\|^2}$$

2.2.4 Final Aggregation

While tortuosity metrics are computed at the segment level, not all are equally informative for global characterization. In practice, the two most robust and interpretable measures are retained for aggregation:

- *Mean weighted curvature* $\bar{\kappa}$
- *RMS weighted curvature* κ_{rms}

The first captures the overall level of bending along the vessel, while the second emphasizes localized variations in curvature and sharp turns.

Length-weighted pooling.

To combine segment-level metrics into a component-level descriptor, a length-weighted average is applied. For a set of segments $\{P_j\}_{j=1}^M$ with arc lengths L_j , the aggregated metric $\mathcal{M}$ is defined as

$$\mathcal{M} = \frac{\sum_{j=1}^M L_j \mathcal{M}_j}{\sum_{j=1}^M L_j}.$$

This ensures that longer vessel branches contribute proportionally more to the final measure, reflecting their structural importance within the vascular network.

Root-policy stratification.

Since segments are constructed according to three distinct root-selection policies (R_1, R_2, R_3), aggregation is performed independently for each policy. As a result, each connected component G_k yields three sets of tortuosity descriptors:

$$\left(\bar{\kappa}^{(R_1)}, \kappa_{\text{rms}}^{(R_1)}\right), \quad \left(\bar{\kappa}^{(R_2)}, \kappa_{\text{rms}}^{(R_2)}\right), \quad \left(\bar{\kappa}^{(R_3)}, \kappa_{\text{rms}}^{(R_3)}\right).$$

Interpretation.

This stratified aggregation allows for comparing how the choice of root influences the perception of tortuosity.

Together, these aggregated metrics provide a comprehensive yet parsimonious summary of vascular geometry for each connected component.

2.2.5 Metric Computation and Storage

Metric aggregation is performed in two distinct modes, depending on whether `VESSEL_METRICS` or `ATLAS_VESSEL_METRICS` is applied. Each mode follows a specific aggregation strategy tailored to the data organization and intended level of analysis. The following sections provide a detailed overview of how metrics aggregated and stored within each mode, highlighting the hierarchical structure of outputs and the relationships between mask-, component-, and segment-level measurements.

VESSEL_METRICS

- **Whole-mask level:** Global metrics are obtained by aggregating measurements across all connected components. Reported features include: total vessel length, number of bifurcations, bifurcation density, estimated volume, fractal dimension, lacunarity, number of loops, number of abnormal-degree nodes, mean and maximum loop length, and root-level tortuosity metrics. Length-dependent metrics are aggregated using length-weighted means. Optionally, statistics can also be reported for the K longest connected components (*Top-K*).

Results are exported to:

- `Whole_mask_metrics.csv` – aggregated statistics across all components.

- **Component level:** Each connected component is analyzed independently, reporting the same set of metrics (length, bifurcations, volume, loops, abnormal-degree nodes, fractal dimension, lacunarity, loop lengths, tortuosity). A binary mask is reconstructed from the skeleton and distance map, and saved as `Conn_comp_<index>_skeleton.nii.gz`.

Results are exported to:

- `all_components_metrics.csv` – table of metrics for every connected component, sorted by connected component length.

- **Segment level:** Within each component, root branches are identified (e.g., largest endpoint root, second-largest endpoint root, largest bifurcation root). Each segment is quantified by geodesic length (arc length), average diameter, and local tortuosity descriptors. For every segment, a CSV file (`Segment_metrics.csv`) is generated, along with an optional reconstructed binary mask (`Segment.nii.gz`).

After execution, the <output_folder> has the following hierarchy:

```
output_folder/
  Conn_comp_1/
    Conn_comp_1_skeleton.nii.gz      component 1
    Segments/
      Largest endpoint root/
        Segment_1/
          Segment_metrics.csv
          Segment.nii.gz
        ...
      Second largest endpoint root/
      Largest bifurcation root/
  Conn_comp_2/
    ...
  all_components_metrics.csv
  Whole_mask_metrics.csv
```

ATLAS_VESSEL_METRICS

- **Region level:** Metrics are aggregated across all connected components within each region.

Reported features include: total vessel length, number of bifurcations, bifurcation density, estimated volume, fractal dimension, lacunarity, number of loops, number of abnormal-degree nodes, mean and maximum loop length, and root-level tortuosity measures.

Length-dependent metrics are computed as length-weighted means.

Optionally, statistics can also be restricted to the K longest connected components (*Top-K*).

Two overall CSV files summarize results at this level:

- `all_components_by_region.csv` : detailed per-component metrics grouped by region.
- `region_summary.csv` : region-level aggregation of metrics.

- **Component level:** Each connected component is analyzed independently, reporting the same set of metrics (length, bifurcations, volume, loops, abnormal-degree nodes, fractal and lacunarity measures, loop lengths, tortuosity).

A binary mask is reconstructed from the skeleton and distance map, and saved as `Conn_comp_<index>_skeleton.nii.gz`.

- **Segment level:** Within each component, root branches are identified (e.g., largest endpoint root, second-largest endpoint root, largest bifurcation root). Segments belonging to these roots are quantified by geodesic length (arc length), average diameter, and local tortuosity descriptors. Each segment has an associated *Segment_metrics.csv* file, and optionally a reconstructed binary mask (*Segment.nii.gz*).

After execution, the specified output folder follows a hierarchical structure:

```

Region_1/
  Conn_comp_1/
    Conn_comp_1_skeleton.nii.gz
    Segments/
      Largest endpoint root/
        Segment_1/
          Segment_metrics.csv
          Segment.nii.gz
          ... (other segments)
      Second largest endpoint root/
      Largest bifurcation root/
    Conn_comp_2/
    ... (other components)
Region_2/
  ... (same structure as above)
all_components_by_region.csv
region_summary.csv

```

Chapter 3

Learning-Based Approaches for Quality Control

3.1 Chapter Overview

This chapter explores a subset of learning-based strategies for Quality Control (QC) in brain vessel segmentation, focusing on the most promising approaches from our perspective: regression- and reconstruction-based methods. The motivation is to complement traditional morphological and topological descriptors with predictive models capable of automatically estimating segmentation quality across diverse masks and segmentation algorithms.

Regression-based methods are presented first, highlighting their interpretability, efficiency, and robustness. Using the VESSEL_METRICS tool, handcrafted vascular descriptors are employed to predict quality metrics such as Dice Similarity Coefficient (DSC) and Centerline Dice Coefficient (Cl-DSC). Key insights include the value of topology-aware features, the challenges of model-agnostic QC, and the limitations of synthetic degradations in representing real-world errors.

Reconstruction-based methods, including encoder–decoder and Variational Autoencoder (VAE) architectures, are then discussed. These approaches learn latent manifolds of anatomically plausible vessel structures to detect implausible segmentations. While they show promise on clean masks, practical application to degraded or noisy predictions revealed limitations in generalization and robustness.

Finally, the chapter synthesizes findings from both approaches, highlighting their complementary strengths and weaknesses, and presents practical recommendations for further analysis.

3.2 Regression-Based Method

3.2.1 Overview

Regression-based methods have historically constituted a cornerstone of Quality Control (QC) frameworks, owing to their adaptability and their ability to model relationships between input features and segmentation quality metrics.

In the medical imaging domain, these approaches have typically relied on hand-crafted or extracted features (e.g, morphological descriptors, intensity statistics, or texture measures) to predict the reliability of segmentation outcomes.

Their interpretability, statistical rigor, and robustness have made them a widely recognized and trusted standard in the evaluation of brain vessel segmentation results. Moreover, regression-based approaches generally offer lower computational demands, with faster training and inference compared to deep learning models, an important factor when considering their integration into medical pipelines for large-scale inspections and routine inference. Nonetheless, the dependence on predefined feature sets constrains their capacity to fully capture the complexity of vascular anatomy and the variability inherent in imaging data.

With the advent of deep learning and representation learning, QC strategies can now leverage end-to-end architectures that jointly perform feature extraction and prediction, enabling the modeling of richer, hierarchical, and non-linear relationships.

As a result, while regression methods continue to serve as a valuable and transparent baseline, particularly in contexts demanding interpretability, efficiency, or limited data, contemporary research increasingly turns to deep learning approaches for improved predictive accuracy and generalization across diverse clinical scenarios.

3.2.2 Feature-based Regression

Our proposed approach builds upon the tool introduced in Chapter 2, extending its functionality within a supervised regression framework.

The full dataset construction pipeline is illustrated in Figure 3.1. The objective is to train a regression model capable of predicting two quality metrics for a given segmentation mask: the Dice Similarity Coefficient (DSC) and the Centerline Dice Coefficient (Cl-DSC). The choice of DSC is motivated by its well-established role as a straightforward and interpretable measure of spatial overlap between two segmentations. In parallel, Cl-DSC is incorporated as it is particularly well-suited to skeleton-based features, which form a large part of the descriptors extracted in our setting. Together, these metrics provide a complementary evaluation of segmentation quality, capturing both volumetric overlap and structural alignment.

In this supervised setup, each row of the training dataset corresponds to one segmentation instance, represented by its extracted features. The target values (DSC and Cl-DSC) are obtained by comparing the predicted segmentation with its corresponding ground truth annotation. This formulation enables the regressor to learn a mapping between feature space and quality metrics, thereby facilitating automated quality estimation.

To establish a meaningful baseline, the same pipeline is replicated using features extracted with PyRadiomics, a widely used state-of-the-art library for radiomics analysis. This parallel setup allows us to build a second dataset and train equivalent regression models, enabling a direct comparison of predictive performance between VESSEL_METRICS and PyRadiomics.

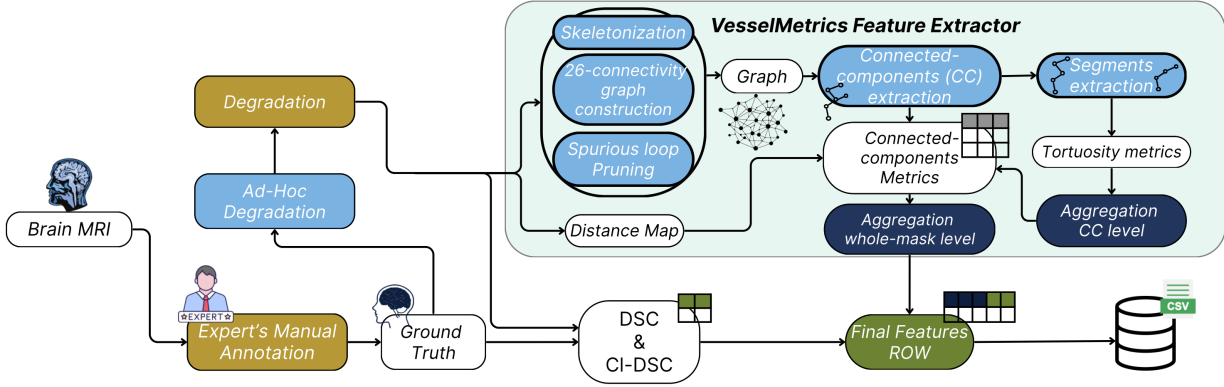


Figure 3.1: Overview of the regression-based dataset construction pipeline.

Input data

The training dataset is composed of features extracted from segmentation masks that were algorithmically degraded.

This strategy was deliberately chosen to ensure a model-agnostic setup, allowing the regressor to assess segmentation quality independently of the specific algorithm used to generate the masks. Such independence is particularly important in clinical contexts, where diverse segmentation methods may coexist across institutions, software pipelines, or imaging modalities. A model-agnostic QC framework thus enhances interoperability and supports more consistent integration into real-world medical workflows.

The construction of the training set was further guided by the objective of achieving as flat as possible distributions of quality scores (DSC and CI-DSC). From a clinical standpoint, this is crucial to guarantee reliable predictions not only for high- and low-quality cases but also for intermediate-quality segmentations, which are often the most challenging to detect and interpret.

However, automatically adjusting degradations to meet predefined target quality scores while preserving anatomical plausibility is inherently nontrivial.

In practice, degradations were directed toward target score ranges, even though the resulting masks may not correspond to fully realistic outputs. This compromise was considered acceptable, as it enabled effective regressor training and ensured sufficient representation across the spectrum of quality levels.

Ultimately, this design choice promotes both robustness and generalizability,

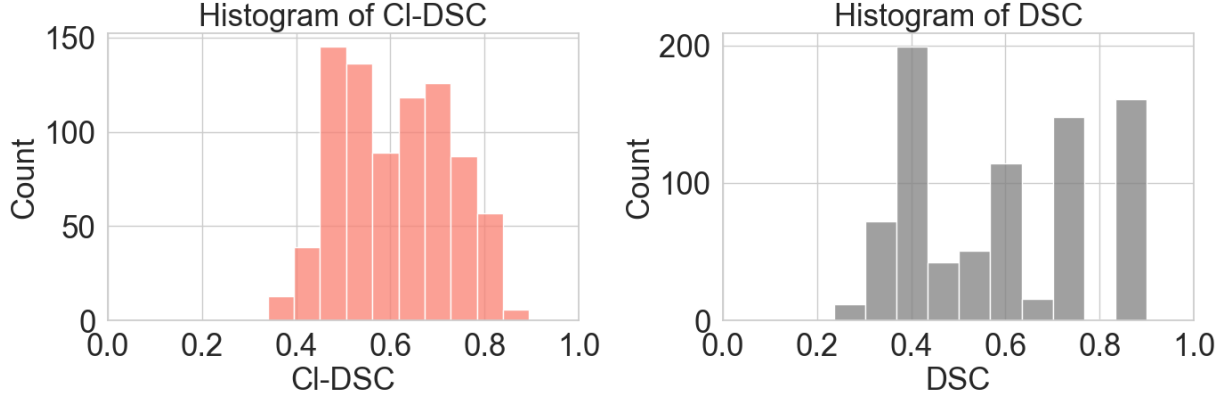


Figure 3.2: Distributions of target quality scores across the degraded dataset.

two properties that are essential for deploying QC systems in large-scale clinical pipelines.

The ground truth segmentation masks were sourced from the publicly available TopCoW Challenge dataset [47], a well-recognized benchmark in medical image analysis. Building upon these ground truths, approximately 800 degraded masks were algorithmically generated to construct the training set.

Despite the complexity of this task, the employed degradation algorithm succeeded in achieving balanced distributions of quality scores across the targeted ranges, as illustrated in Figure 3.2. This outcome provided decent coverage of the entire quality spectrum.

In practice, this was achieved by a degradation algorithm with several type of degradations, which will explained in the next section.

Degradation algorithm

The degradation pipeline was designed to algorithmically perturb ground-truth masks while preserving basic anatomical plausibility, with the explicit goal of generating masks spanning a broad spectrum of quality scores (DSC). Some examples are shown in Figure 3.3 This was achieved through controlled corruption routines applied in a probabilistic manner.

The core of the algorithm is a function which attempts to transform an input ground-truth mask into a degraded version that approximates a predefined target DSC. Target values were selected across the full quality range (0.85, 0.70, 0.55,

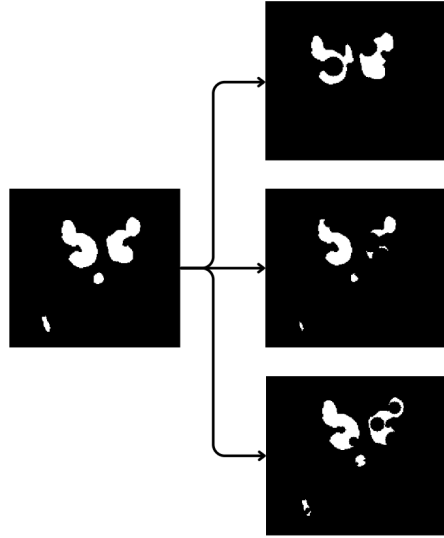


Figure 3.3: Examples of possible slice degradations.

0.35, 0.15), ensuring adequate coverage from nearly perfect to severely corrupted masks.

To reach a given target DSC, the algorithm applies different families of degradations via the `corrupt_one_masks` utility:

- Erosion: selectively shrinks the segmentation boundary, mimicking undersegmentation.
- False positives (fp): introduces spurious foreground regions disconnected from the true anatomy.
- Holes: removes internal parts of the structure, imitating local segmentation failures.
- Randomized corruption operations: each operation is applied with configurable probabilities, and up to a maximum number of iterations per mask.

The corruption parameters are conditioned on the target DSC range. For high DSC values (≥ 0.8), the algorithm applies minimal perturbations (e.g., low erosion and hole probability, few operations). For intermediate DSC values (0.4–0.7), more aggressive settings are used, combining erosion with false positive insertions and larger internal defects. For very low DSC values (< 0.4), strong corruptions with high probabilities of all operation types are applied, leading to heavily distorted and anatomically implausible masks.

Because corruption is stochastic, multiple attempts are made for each target DSC. After every attempt, the DSC is computed between the degraded mask and the ground truth. The attempt that achieves the closest DSC to the target is retained. If the achieved DSC falls within a tolerance of ± 0.05 of the target, the search terminates early; otherwise, the algorithm continues until the maximum number of attempts is reached. This strategy balances precision in hitting the target score with computational efficiency.

For each ground-truth mask, five degraded versions are produced, one per target DSC value, resulting in a diverse dataset of masks systematically covering the quality spectrum. This procedure was applied to all vessel annotations from the TopCoW dataset, ultimately yielding ~ 800 degraded masks for training.

Preliminary Analysis: 80-20 Split in Trainset

Within this setup a regression algorithm is trained on 80% of the training data, while the remaining 20% is set aside for validation. The model does not “see” this validation subset during training, which allows us to assess its predictive performance on data that is not directly used for fitting.

Formally, let the dataset be defined as

$$D = \{(x_i, y_i)\}_{i=1}^N,$$

where $x_i \in \mathbb{R}^p$ denotes the feature vector with p predictors and $y_i \in \mathbb{R}$ the corresponding target value. The dataset D is partitioned into:

$$D_{\text{train}} \subset D, \quad D_{\text{val}} \subset D, \quad D_{\text{train}} \cap D_{\text{val}} = \emptyset, \quad |D_{\text{train}}| = 0.8N, \quad |D_{\text{val}}| = 0.2N.$$

Model performance is then evaluated on the validation set D_{val} , providing an unbiased estimate of generalization.

This procedure is useful to understand how well the regression model generalizes within the same dataset. The validation split helps to:

- Evaluate performance: on unseen samples, giving an unbiased estimate of accuracy or error metrics.
- Identify overfitting: by comparing training and validation errors.
- Guide model selection: and hyperparameter tuning before testing on a truly independent test set.

In addition, this validation strategy can be employed as a analysis of the features available in the dataset.

By observing the model's performance on the 20% hold-out subset, we can infer whether the chosen features are sufficiently informative for predicting the target variable.

If the regression model achieves consistent accuracy between training and validation sets, this suggests that the features capture meaningful patterns. Conversely, if the validation error remains high despite reasonable model complexity, it may indicate that the features lack predictive power, suffer from redundancy, or require transformation.

Such an analysis provides an early diagnostic step before more elaborate techniques (e.g., feature selection, dimensionality reduction, or engineered variables) are introduced, ensuring that the dataset is rich enough to support robust modeling.

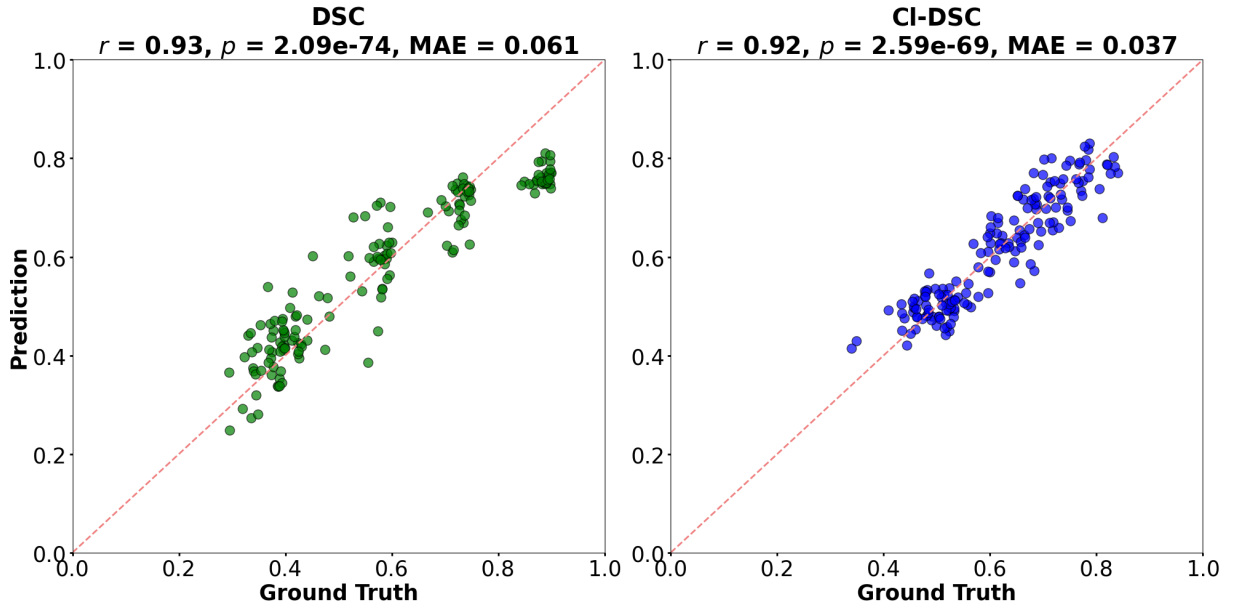


Figure 3.4: Performances on trainset.

As shown in Figure 3.5, the features were sufficient for the model to clearly distinguish between the different quality levels. It is worth noting that the two sets were constrained to share the same distribution across quality levels.

The reported metrics provide complementary perspectives on model performance. The Pearson Correlation (r) measures the strength and direction of the linear relationship between predicted and true quality levels, indicating how consistently the model follows the underlying trend. The Mean Absolute Error (MAE) captures the average magnitude of prediction errors, directly reflecting how close the predictions are to the actual values. Finally, the p-value (p) assesses the statistical significance of the observed correlation, ensuring that the results are unlikely to be due to

chance.

Together, these metrics demonstrate both the accuracy and the reliability of the model.

Test

In our test setup, as announced, we conducted a direct comparison with PyRadiomics, a widely adopted open-source Python library for deriving radiomic features from medical images and their corresponding masks [48]. PyRadiomics is particularly valued for its flexibility and reproducibility, making it an established baseline for this study.

Radiomic features are quantitative descriptors extracted from medical images that capture patterns, textures, shapes, and intensity distributions not easily perceivable by the human eye. They can represent properties such as tumor heterogeneity, vessel morphology, or tissue texture, providing a high-dimensional characterization of image regions. Radiomics has become broadly used because it transforms standard medical images (CT, MRI, etc.) into rich datasets of measurable features that can be linked to clinical outcomes, disease phenotypes, or molecular characteristics. This makes radiomics a powerful tool for diagnosis, prognosis, and treatment response assessment. Its popularity also stems from its reproducibility, open-source toolkits (such as PyRadiomics), and its ability to support data-driven, non-invasive biomarkers in precision medicine [49].

For both VESSEL_METRICS and PyRadiomics, features were extracted from the same set of degraded masks to construct training datasets. Two separate Random Forest regressors were then trained per feature set: one to predict the DSC and another to predict the CI-DSC. Each feature extractor was subsequently evaluated on a held-out test set comprising 125 predicted masks for each of the five segmentation models (A2V, JOB-VS, VesselBoost, nnUNet, and SPOCKMIP), described previously in Section 1.4.5.

As shown in Figure 3.5, VESSEL_METRICS demonstrates decent performance, achieving consistently low error values across the evaluations. In contrast, PyRadiomics struggles noticeably, yielding less accurate predictions. As expected, CI-DSC provides significantly better results, standing out as the most effective estimator for VESSEL_METRICS features, since the descriptors are directly linked with the vessel skeleton. Still, even in the CI-DSC plots, clusters of poorly predicted points can be observed, negatively impacting correlation values.

To better understand this behavior, we analyzed the results on a per-model basis.

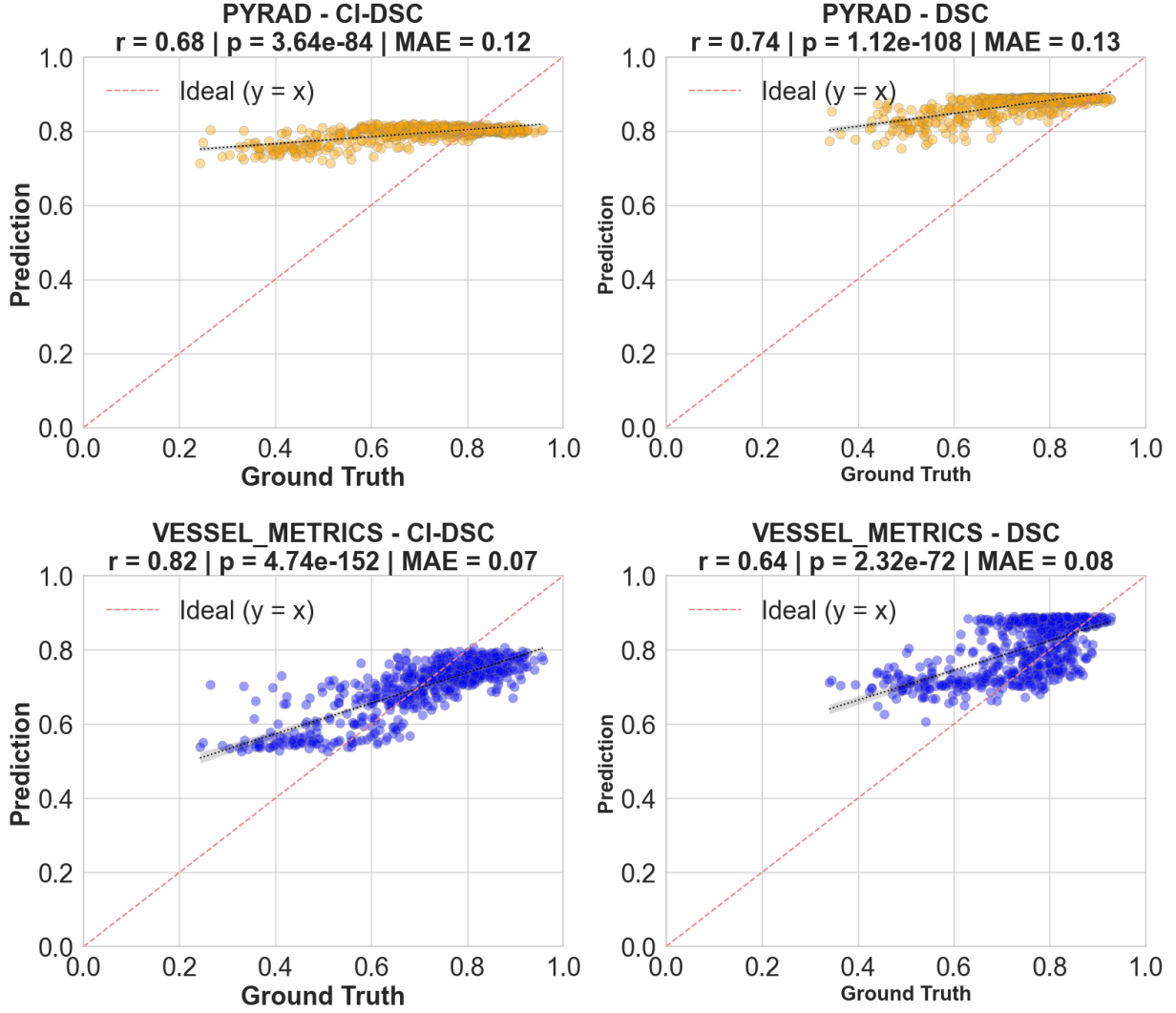


Figure 3.5: Test Performance

A compact visualization focusing on the superior target (CI-DSC) is presented in Figure 3.6. This analysis revealed that certain segmentation models were systematically harder to predict, even though the overall training performance appeared reasonable. A likely explanation is that the degradations used to construct the training dataset, while controlled and effective in simulating general errors, did not fully reflect the error patterns produced by some models in practice.

In other words, the variability introduced during training did not sufficiently encompass the broader spectrum of real-world prediction errors. As a result, the regressors struggled to generalize, exposing a limitation in the diversity of mask degradations employed during training.

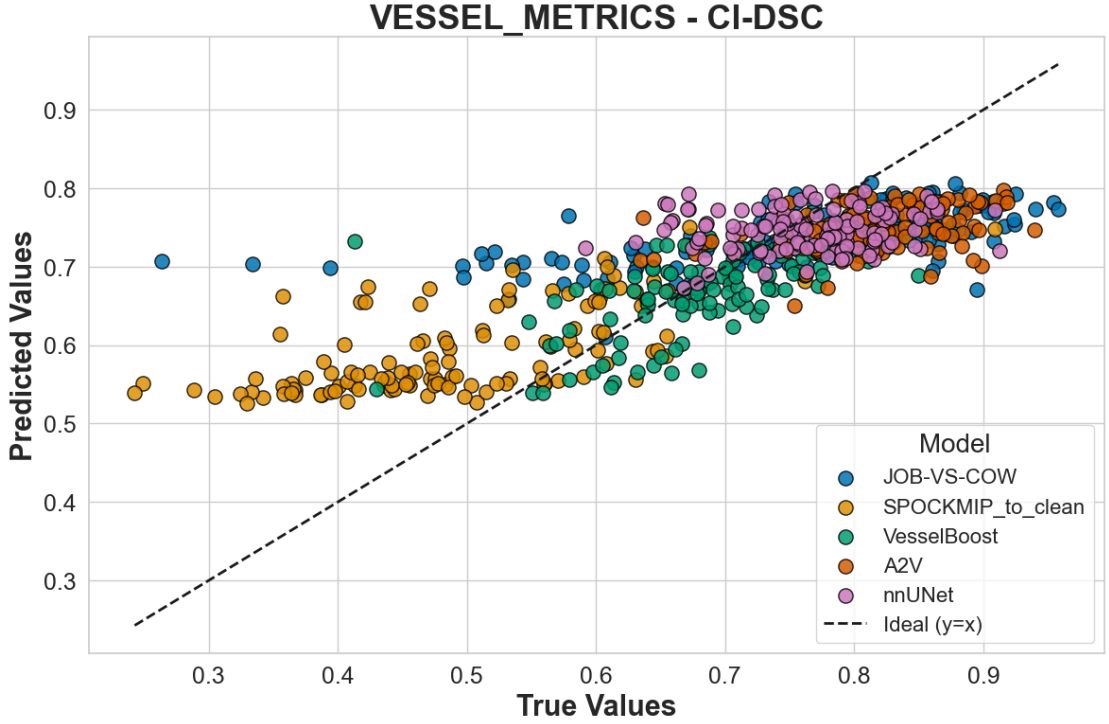


Figure 3.6: Performance per model.

A key point to emphasize is that, within this model-agnostic setup, the approach faces intrinsic challenges when applied to the complexity of real-world segmentations. Constructing an extensive and exhaustive training dataset that faithfully represents the full diversity of real segmentation errors is extremely difficult, if not infeasible. This highlights a fundamental limitation: while controlled degradations provide a valuable foundation for learning, they cannot fully capture the variability and unpredictability inherent to real-world outputs.

To fully uncover the weak points of the approach, it is not sufficient to rely solely on global performance metrics. A deeper inspection of the model's internal decision process is required. One effective strategy is analyzing the relative importance of features. Feature importance analysis reveals which descriptors the Random Forest relied on most heavily and which contributed little or no useful information. By studying this distribution, it becomes possible to identify potential biases, highlight redundancies or underutilized descriptors, and assess whether the chosen feature set captures the most relevant aspects of segmentation quality. Such an analysis offers valuable insight into both the strengths and limitations of the current

design, guiding improvements in feature selection and dataset construction.

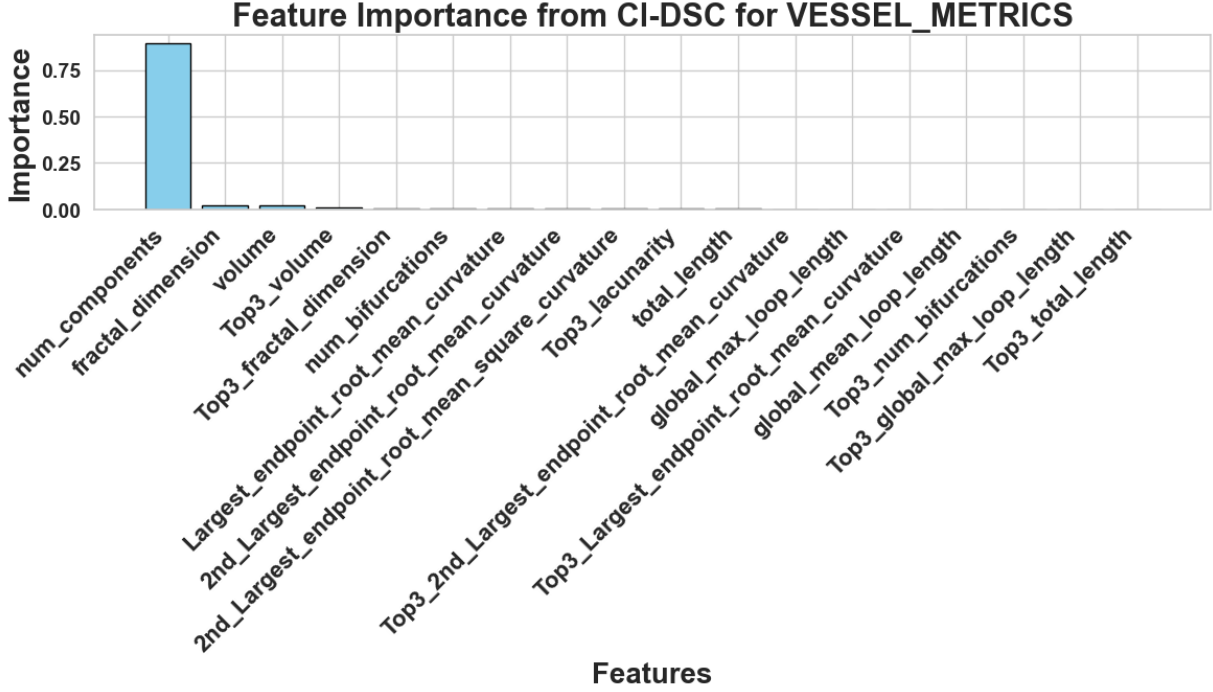


Figure 3.7: Feature importance.

As illustrated in Figure 3.7, the number of connected components emerged as a dominant feature for VESSEL_METRICS. However, this may simply be an artifact of how the training data was constructed. By degrading masks in a relatively uniform manner with a limited set of operations, the training procedure may have inadvertently emphasized fragmentation as the strongest signal of mask quality. The Random Forest effectively learned a shortcut: “fewer components $\rightarrow$ higher DSC; more components $\rightarrow$ lower DSC”. While this heuristic works well for synthetic degradations, it may fail to generalize to the more diverse and subtle error patterns present in real-world segmentation outputs.

3.2.3 Conclusion

In summary, VESSEL_METRICS consistently outperformed PyRadiomics in predicting segmentation quality metrics, with particularly strong performance on the CI-DSC, which directly reflects vessel skeleton fidelity. Beyond accuracy, these

results highlight the tool’s potential as an effective component of automated Quality Control (QC) pipelines for segmentation outputs, enabling rapid, objective, and reproducible evaluation of mask quality. Nevertheless, the analysis revealed systematic prediction challenges for certain segmentation models, indicating that training on controlled degradations alone may not fully capture the diversity of errors encountered in real-world data. This underscores the critical importance of incorporating realistic and varied degradations in training datasets to ensure that automated QC methods remain reliable and generalizable across heterogeneous segmentation outputs. By combining high predictive performance with interpretability through feature importance analysis, VESSEL_METRICS offers a promising pathway toward robust, data-driven QC practices that can help detect failures, guide model improvements, and ultimately enhance the reliability of medical image analysis workflows.

3.3 Reconstruction-Based Methods

3.3.1 Overview

Reconstruction-based methods have emerged as a powerful strategy for Quality Control (QC) in medical image segmentation, particularly in scenarios where anatomical plausibility and structural consistency are critical.

These approaches learn a model of high-quality, anatomically correct segmentations and use it to evaluate or refine new predictions. By projecting segmentations onto a latent manifold of plausible structures, reconstruction-based QC can identify deviations caused by artifacts, noise, or model errors.

Notable examples include Liu et al. [50], who proposed an alarm system for segmentation algorithms using a variational autoencoder trained on ground-truth masks. Their system detects implausible predictions by comparing reconstructions to the learned anatomical manifold, providing a model-agnostic mechanism for segmentation QC. Similarly, Painchaud et al. [51] applied a constrained VAE to cardiac segmentation, projecting predictions onto a latent space of anatomically valid shapes to correct implausible outputs automatically.

Compared to regression-based QC methods, reconstruction-based approaches are less dependent on handcrafted features and can capture complex, non-linear relationships inherent in anatomical structures. They are particularly advantageous when variability in imaging quality or acquisition protocols is high, offering a principled, automated, and model-agnostic way to assess and refine segmentation quality.

While these methods typically require larger datasets and higher computational resources than traditional regression approaches, they offer the potential to integrate anatomical priors directly into the QC process, improving robustness and reliability in both research and clinical pipelines.

3.3.2 General encoder-decoder

Core idea

A reconstruction-based unsupervised approach was investigated, designed to learn directly from vessel masks without paired image-segmentation annotations. The latent manifold is constructed using high-quality ground-truth segmentation masks only, exploiting their structural regularities to build an implicit anatomical prior.

An encoder–decoder network was trained to reconstruct vessel maps from themselves. Given a ground truth segmentation S , the encoder projects it into a latent representation $z \in \mathbb{R}^m$, which the decoder maps back to S' . The optimization relies on a reconstruction loss (binary cross-entropy), ensuring that S' remains close to S while regularizing the latent space to capture meaningful vascular structures. Figure 3.8 illustrates the pipeline.

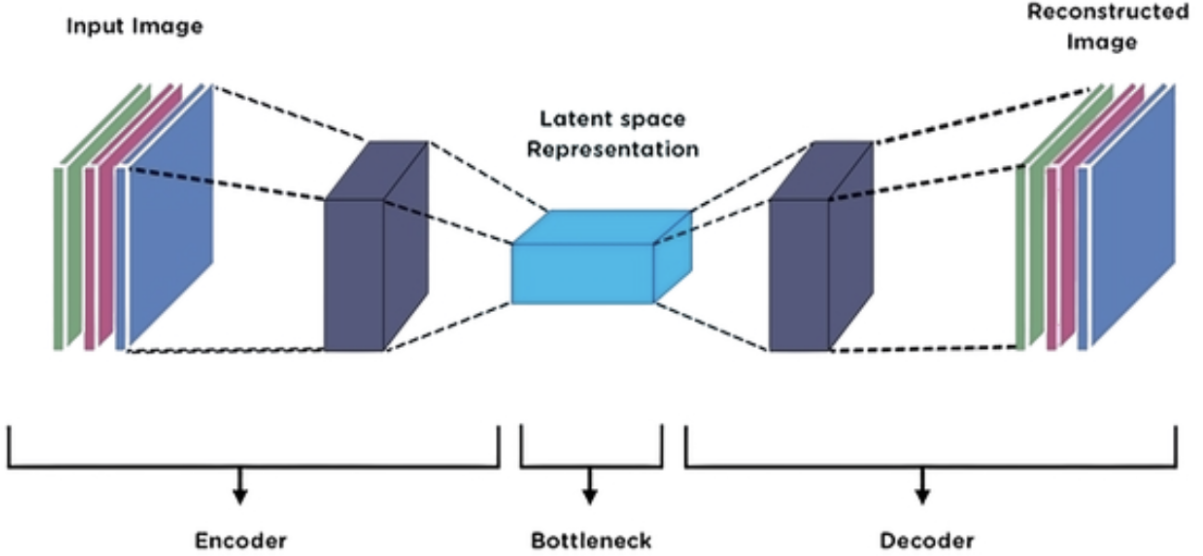


Figure 3.8: Reconstruction flow using an encoder–decoder model.

At inference, an external prediction $\hat{S}$ is processed through the model, yielding a surrogate reconstruction S^* that lies closer to the learned manifold of anatomically plausible vessels. Comparing $\hat{S}$ with S^* provides a plausibility estimate: reliable predictions reconstruct faithfully, whereas artifact-driven ones deviate. This mechanism is model-agnostic and robust to variability in imaging protocols.

Experimental setup

Careful consideration is required in selecting an appropriate model, particularly one that is balanced in terms of depth. The anatomical structures under consideration (thin, elongated, and highly branching vessels), demand a network capable of capturing fine-scale details while also maintaining global anatomical coherence. A model that is too shallow risks underfitting and missing subtle vascular patterns. Conversely, an excessively deep network increases the risk of overfitting, vanishing gradients, and unstable training dynamics, while also significantly raising computational cost.

To address this trade-off, several encoder–decoder architectures of varying depth were systematically evaluated. As an initial design, a deep 3D convolutional model was implemented, consisting of six to seven downsampling and upsampling stages. In parallel, different input configurations were tested, including variations in the spatial size of cropped volumes and modifications to the segmentation masks (e.g., enlarging or restricting the surrounding context).

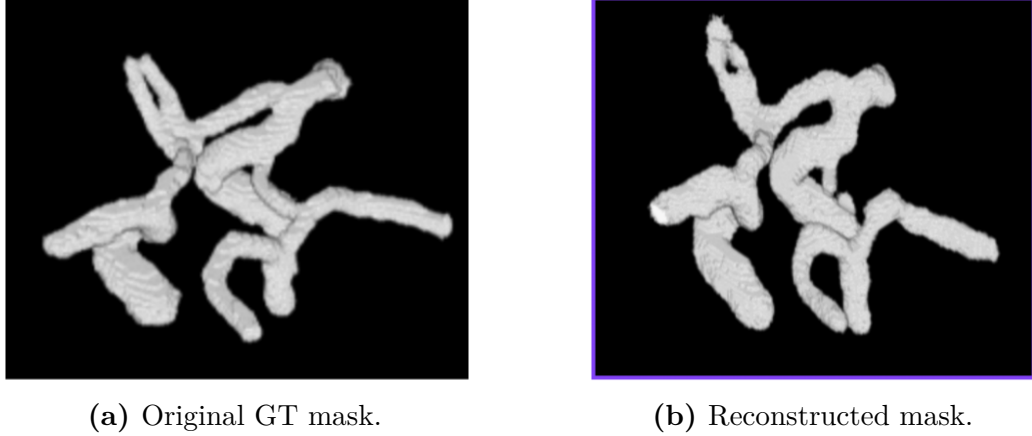


Figure 3.9: Successful reconstruction instance (DSC = 0.89).

These initial trials with the deep 3D model showed no signs of effective learning. The training loss remained stagnant throughout the epochs, and the model failed to converge. This outcome suggests that the network’s complexity was excessive relative to the available data and optimization setup: the large number of parameters, combined with the increased difficulty of training 3D convolutions, made gradient propagation unstable and prevented the model from entering a learning regime. The natural next step was a simplification of the model architecture. The depth of the encoder–decoder backbone was reduced to four downsampling and upsampling stages, thereby lowering representational complexity while retaining sufficient capacity to capture vessel structures at multiple scales. This shallower 3D model demonstrated clear improvements: the loss curve steadily decreased during training, and consistently high Dice scores were observed on the validation set. Figure 3.9 provides an illustrative example, showing that the reconstructed vessel segmentation obtained by the reduced-depth model closely aligns with the ground truth.

Inference phase

During inference, when the model was tasked with reconstructing degraded masks in order to recover the original, undegraded structures, it consistently failed to achieve the intended corrections. Rather than producing faithful restorations of

the ideal ground truth masks, the network instead generated alternative vessel segmentations that deviated substantially from the true anatomy.

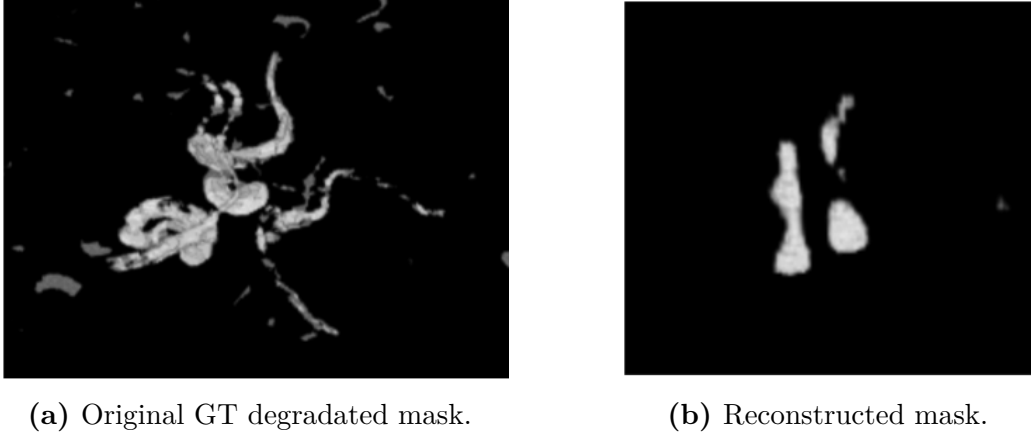


Figure 3.10: Inference reconstruction instance.

As illustrated in Figure 3.10, the reconstructed mask does not correspond to the original vascular anatomy.

The outputs remained incomplete and anatomically inconsistent, ultimately failing to meet the objective of reliably reversing the applied degradations.

3.3.3 VAE with search in latent space

Core idea

A generative model was designed to learn from image–segmentation pairs, drawing strong inspiration from [52]. The model provides a probabilistic latent embedding that jointly captures both intensity and anatomical domains. This latent manifold is constructed using high-quality ground-truth image–segmentation pairs, ensuring that the reconstructed segmentations remain anatomically consistent with true vessel structures. To approximate this manifold, a Variational Autoencoder (VAE) was employed..

Let $X = (I, S)$ denote an image–segmentation pair. The encoder maps X to a Gaussian-distributed latent vector $z \in \mathbb{R}^m$, and the decoder reconstructs (I', S') from z . The loss combines binary cross-entropy for masks, mean squared error for images, and a Kullback–Leibler divergence term to regularize the latent space. Figure 3.11 shows the pipeline.

At inference, an external prediction $\hat{S}$ is evaluated by performing an iterative search in latent space to identify the latent code z^* whose reconstruction (I^*, S^*) best

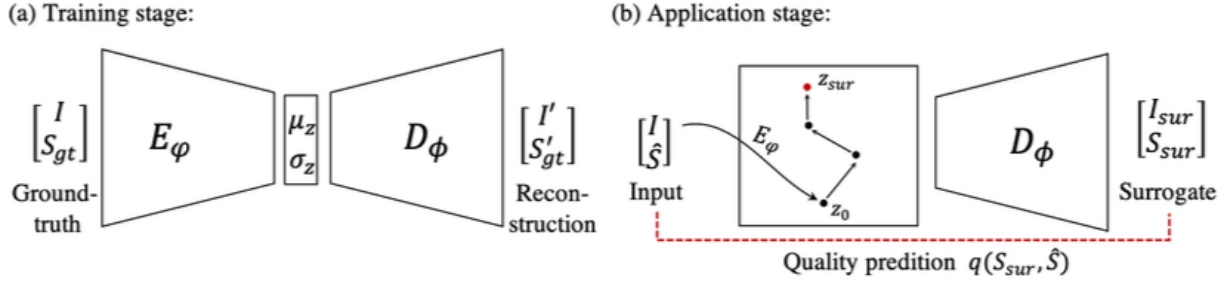


Figure 3.11: Pipeline of the VAE with latent space search.

matches $(I, \hat{S})$. The surrogate S^* serves as an anatomically plausible reference for comparison.

This formulation differs from the encoder–decoder baseline in two aspects:

1. the inclusion of image–segmentation pairs, allowing joint modeling of appearance and anatomy;
2. the latent search, which explicitly projects predictions back onto the learned manifold.

Experimental setup

The experimental setup followed the same general strategy as the encoder–decoder approach. Initially, a deeper 3D network was implemented to maximize representational capacity. However, training this architecture showed no effective learning, likely due to the increased difficulty of propagating gradients through multiple 3D convolutional layers and the limited size of the dataset. Consequently, a shallower network was adopted, reducing the number of downsampling and upsampling stages while retaining sufficient capacity to capture vessel structures at multiple scales.

To evaluate the impact of the latent space dimensionality on reconstruction quality, several bottleneck sizes were systematically tested. This allowed assessment of how the dimensionality of the latent representation influences the ability of the network to capture anatomical variations while maintaining reconstruction fidelity.

Inference phase

Within these setups, the model struggled to generate confident predictions, as illustrated in Figure 3.12, where the soft outputs rarely exceed the 0.5 threshold. This indicates a difficulty in producing segmentations that the network considers sufficiently certain. The underlying causes are likely multifactorial, including the

limited size of the training dataset, the complexity of the 3D anatomical structures, and the increased variability introduced by imaging artifacts and noise.

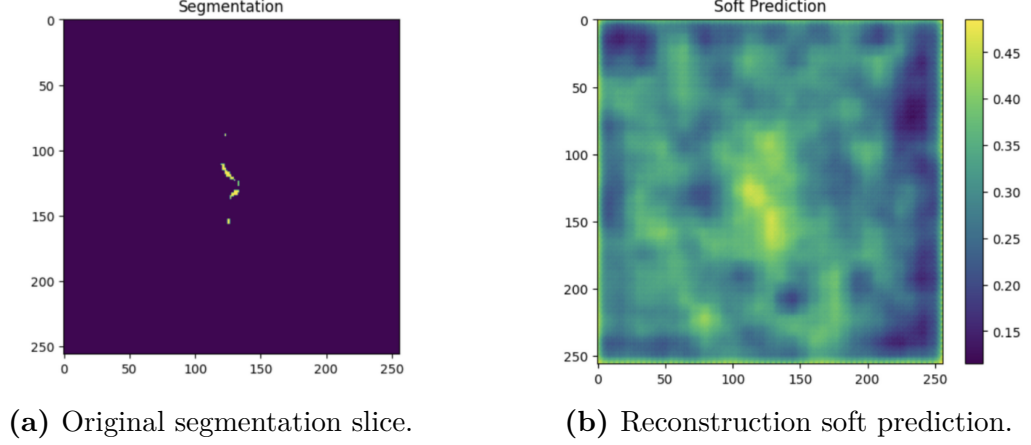


Figure 3.12: Example of unstable soft predictions for a reconstructed slice.

Additionally, the iterative latent space search may amplify these challenges. Small deviations in the predicted segmentation can result in large shifts within the latent space, leading the reconstructed outputs to deviate from the desired ground-truth anatomy. As a consequence, the surrogate segmentations sometimes fail to provide a reliable reference for assessing the plausibility of external predictions.

These observations suggest that, while the VAE framework provides a principled mechanism for learning anatomically plausible structures, careful tuning of the latent space dimensionality, network capacity, and training strategy is critical. Future refinements may include data augmentation, improved regularization, or alternative latent search strategies to enhance convergence and confidence in the reconstructed outputs.

3.3.4 Conclusion

Reconstruction-based QC methods offer an attractive, conceptually simple route: by learning a manifold of plausible vessel structures, they can in principle distinguish anatomically consistent outputs from artifact-driven or unstable predictions.

Our experiments confirm two important aspects of this promise. First, an encoder–decoder trained purely on high-quality masks is capable of learning to reconstruct clean vessel geometries, as evidenced by high validation Dice scores for reconstructions of ground-truth masks (see Figure 3.9).

Second, generative formulations that jointly model image and segmentation (the VAE) provide a principled probabilistic latent space that can, in theory, support projection of noisy predictions back onto the anatomical manifold.

However, the practical utility of these methods as corrective tools or stand-alone QC agents proved limited in our setting. Two recurrent failure modes emerged:

1. Inference mismatch: although reconstruction of clean masks succeeds under controlled conditions, passing degraded or partially incorrect masks through the trained reconstructor did not consistently recover the original anatomy.
2. Latent-space sensitivity and low confidence: the VAE-based latent search struggled to produce confident (high-probability) segmentations for many inputs; soft outputs frequently remained near the classification threshold. The iterative search in latent space amplified small input perturbations into large latent displacements, which in turn yielded reconstructions that could deviate markedly from the desired anatomy.

Analysis points to several root causes that are likely responsible for these behaviors: (i) limited dataset size relative to the complexity of 3D vascular anatomy, (ii) the increased optimization difficulty and parameter count of deep 3D convolutional models. Together, these factors reduce the models' ability to generalize from reconstructing ideal masks to reliably correcting or assessing imperfect predictions.

In summary, reconstruction-based approaches remain a promising component of segmentation QC for vascular tasks, but our results highlight that learning to reconstruct high-quality masks is not synonymous with reliably correcting or unequivocally validating arbitrary external predictions. With targeted changes, reconstruction-based methods can become much more effective and practically useful in both research and clinical pipelines.

Chapter 4

Discussion and Conclusion

The work presented in this study addresses a timely and practically significant gap in medical image analysis: the development of automated quality control (QC) mechanisms for cerebral vessel segmentation. Accurate delineation of the cerebral vascular network is essential for both diagnostic and research applications in cerebrovascular and neurodegenerative diseases.

Yet, the brain’s vascular anatomy presents exceptional technical challenges, its vessels are extremely thin, highly tortuous, and subject to pronounced anatomical variability across subjects, scanners, and pathologies. These characteristics make both segmentation and objective quality assessment particularly demanding. Even state-of-the-art neural networks such as nnU-Net achieve good accuracy but remain prone to subtle and sometimes clinically significant errors that can compromise downstream analyses or decision-making. Despite the evident importance of this problem, the existing literature reveals no prior systematic investigation of QC methodologies specifically designed for brain vessel segmentation.

Motivated by this gap, the present work introduces a systematic comparative study between two conceptually distinct paradigms for segmentation quality assessment. The first approach employs supervised regression on handcrafted, topology- and morphology-aware features derived from the VESSEL_METRICS framework, which quantifies vessel length, diameter, tortuosity, branching, and connectivity. The second explores self-supervised reconstruction using encoder–decoder and variational autoencoder (VAE) architectures trained to model the manifold of anatomically plausible vessel masks, thereby identifying or correcting implausible segmentations through reconstruction behaviour.

As part of this work, a dedicated software toolkit, VESSEL_METRICS, was developed to enable systematic and interpretable vascular analysis. This toolkit computes detailed topological and morphological descriptors from any cerebral vessel segmentation, whether generated manually or by automated methods. The extracted features encompass a broad range of biologically and geometrically meaningful properties, including vessel length, diameter, branching patterns, tortuosity, volume, fractal dimension, and connectivity. Together, these descriptors provide a quantitative and anatomically interpretable representation of vascular structure, forming the foundation for both explanatory analyses and data-driven modelling. In this sense, VESSEL_METRICS serves simultaneously as a general-purpose framework for vascular characterization and as a core component of the proposed automated quality control pipeline.

The overarching methodological insight of this study is that the two QC paradigms offer complementary approaches: feature-based models provide supervised, interpretable predictions of segmentation quality, while reconstruction-based models enable unsupervised estimation of the same. The following discussion synthesises

these empirical contrasts, interprets their implications for the broader field of vascular imaging, and outlines practical recommendations for developing robust, interpretable, and scalable QC systems for cerebral vessel segmentation.

In our experiments, Random Forest regressors built on VESSEL_METRICS descriptors, trained on a synthetically degraded dataset, achieved the most accurate and stable predictions when evaluated on a test set comprising segmentation outputs from multiple real-world models. Predictions of the centerline Dice (Cl-DSC) were consistently more reliable than those of the volumetric Dice, reflecting the skeleton- and topology-oriented nature of the feature set. Under identical conditions, PyRadiomics-based models underperformed, indicating that generic radiomic descriptors are less attuned to the fine-grained topological cues critical for vascular fidelity. However, feature-importance analyses exposed a notable limitation: model performance depended disproportionately on a small subset of variables, particularly the number of connected components, suggesting that part of the observed advantage may stem from exploiting strong synthetic artifacts rather than genuinely learning robust, generalisable topological relations.

Thus, while these findings confirm the potential of topology-aware descriptors for quality control in vascular segmentation, they also highlight the need to complement synthetic degradations with broad empirically observed errors and to conduct systematic ablation and importance analyses to identify and mitigate spurious shortcuts. In summary, feature-based regressors represent a powerful and pragmatic approach, yet their real-world generalisability ultimately depends on the realism and diversity of the degradations used for training.

Reconstruction-based QC, learning a manifold of clean, ground truth vessel segmentations via encoder-decoder and VAE architectures, offers a contrasting strategy: determine plausibility by how well an input mask projects onto or is reconstructed from a learned anatomical latent space. This approach has appealing theoretical properties. A sufficiently expressive generative model, trained on a diverse set of high-quality vessel masks, could, in principle, reconstruct anatomically plausible structures from incorrect, partially incorrect, or implausible inputs. In doing so, it would both flag and potentially correct segmentation errors without the need for explicitly labelled corruptions, while also providing a quantitative means of assessing segmentation accuracy by measuring the discrepancy between an input mask and its optimal reconstruction.

Empirically, however, the reconstruction methods exhibited a pattern of behaviour that tempers that theoretical optimism. When presented with clean masks, the models reconstructed anatomy with high fidelity, confirming that the architectures can learn useful generative priors for vascular structure. When presented with degraded or even partially incorrect masks, however, plain reconstructor inference

and even latent-search variants of the VAE often failed to produce reliable corrections. Reconstructions tended to be overly smooth, lost fine branches, clustered outputs near the decision threshold (indicative of low confidence), and were highly sensitive to small input perturbations. The farther a degraded mask lay from the training manifold, the less plausible the reconstruction; in many cases the output diverged substantially from the true anatomy rather than providing a conservative or corrective estimate. These observations point to a crucial moderating variable: dataset scale and diversity. VAE-type methods appear especially sensitive to the breadth of anatomical examples seen during training. Sparse coverage of the true manifold, inevitable when only a few hundred high-quality vessel maps are available, produces latent spaces that lack the local structure necessary to map incorrect inputs back to plausible anatomies. Consequently, while generative reconstructions are conceptually attractive, their utility as standalone certifiers or automatic correctors is limited under current data regimes. They may still be valuable as a component of a hybrid system or as a diagnostic tool when trained on far larger and more heterogeneous corpora.

Taken together, these results yield several actionable conclusions for researchers and practitioners aiming to deploy QC for cerebral vessel segmentation.

The study suggests a pragmatic roadmap of experiments and engineering steps that balance cost and expected benefit.

- 1) The highest-priority activity is empirical error collection: gather segmentation outputs and their failure modes from multiple state-of-the-art models and use these model-produced errors to re-train and validate regression models. To close the realism gap in training data, treat synthetic corruptions as a useful starting point but systematically augment them with these empirically sampled errors so that QC regressors learn to recognise real-world failure patterns rather than synthetic shortcuts.
- 2) Scale up the ground-truth corpus for generative training and measure learning curves for reconstruction fidelity and robustness; this experiment will clarify whether architecture or data scarcity is the limiting factor.
- 3) Implement and evaluate topology-aware losses for the reconstructor and quantify their effect on downstream QC metrics such as Cl-DSC and clinically relevant connectivity measures.
- 4) Prototype hybrid approaches that augment the interpretable descriptor set with lightweight learned modules: small CNNs applied to local patches around centerlines or graph neural networks operating on extracted centerline graphs can capture relational and contextual cues while preserving overall transparency.
- 5) Embed routine diagnostic procedures (e.g., feature ablations, importance tracking, and bias audits) into model development pipelines to detect and correct spurious shortcuts (e.g., overreliance on component counts) early.

The work also exposes persistent limitations and practical constraints. High-quality vessel annotations are expensive to produce and subject to inter-rater variability; assembling large, diverse, expert-curated corpora will require coordination and resources. Even with more data, some errors remain intrinsically hard to detect with global metrics: losses of clinically critical small branches may have modest influence on volumetric DSC but large clinical implications.

The dependence of regressors on the fidelity of corruption models and the sensitivity of generative models to latent coverage both argue for human-in-the-loop systems in the near term. Finally, generalisability across patient populations must be validated explicitly; QC systems tuned to a narrow set of conditions risk brittleness when deployed in broader clinical settings.

In conclusion, and to briefly sum up, this work presents the first systematic investigation of automated quality control for cerebral vessel segmentation and introduces `VESSEL_METRICS`, a comprehensive framework for quantifying morphometric and topological properties of vascular masks. The study demonstrates that feature-based regression models can potentially predict successfully segmentation quality using interpretable, biologically meaningful descriptors, while also exploring generative reconstruction as an unsupervised alternative for detecting corrupted segmentations. By comparing these two paradigms, the work delineates their complementary strengths and limitations. Collectively, these contributions establish a foundation for developing robust, scalable, and interpretable QC systems in medical image segmentation, particularly within the anatomically complex domain of cerebral vasculature, and point toward future work focused on integrating complementary methodologies, expanding dataset diversity, and advancing more generalisable and trustworthy quality control frameworks across medical imaging domains.

Bibliography

- [1] Nivedita Agarwal and Roxana Octavia Carare. «Cerebral Vessels: An Overview of Anatomy, Physiology, and Role in the Drainage of Fluids and Solutes». In: *Frontiers in Neurology* Volume 11 - 2020 (). ISSN: 1664-2295. DOI: 10.3389/fneur.2020.611485. URL: <https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2020.611485> (cit. on p. 3).
- [2] Saeed Sadigh-Eteghad, Babak Sabermarouf, Alireza Majdi, Mahnaz Talebi, Mehdi Farhoudi, and Javad Mahmoudi. «Amyloid-beta: a crucial factor in Alzheimer's disease». In: *Medical principles and practice* 24.1 (2015), pp. 1–10 (cit. on p. 3).
- [3] Audrey Tan and David Roberts. «Cerebral circulation 1: anatomy». In: *BJA Education* (2021) (cit. on p. 4).
- [4] Wei Wang, Qiudian Wang, Guanying Yang, Mingrui Qiu, and Aizhen Pan. «A Review of the *Circle of Willis*: Investigative Methods, Anatomical Variations and Correlated Ischemic Brain Diseases». In: *Science Insights* 33.2 (2020), pp. 159–166. DOI: 10.15354/si.20.re069. URL: <https://bonoi.org/index.php/si/article/view/103/81> (cit. on pp. 5, 6).
- [5] Mingrui Qiu and Aizhen Pan. «Variations in Circle of Willis Anatomy and Clinical Stroke Risk: A Meta-Review». In: *Frontiers in Neurology* 12 (2021), p. 1098950. DOI: 10.3389/fneur.2021.1098950. URL: <https://www.frontiersin.org/articles/10.3389/fneur.2021.1098950/full> (cit. on pp. 5, 6).
- [6] Mustafa Demir, Elif Ozturk, and Uner Ture. «Hemodynamic Risk Factors in Aneurysm Formation at Circle of Willis Junctions». In: *Turkish Neurosurgery* 34.1 (2024), pp. 21–29. DOI: 10.5137/1019-5149.JTN.38100-22.2. URL: https://www.turkishneurosurgery.org.tr/pdf/JTNEPUB_21835_online.pdf (cit. on p. 6).
- [7] Muhammad Azeem Uddin, Tanveer Ul Haq, and Muhammad Zafar Rafique. «Cerebral venous system anatomy». In: *Journal of Pakistan Medical Association* 56.11 (2006), p. 516 (cit. on p. 7).

- [8] M. Jorge Cardoso et al. *MONAI: An open-source framework for deep learning in healthcare*. 2022. arXiv: 2211.02701 [cs.LG]. URL: <https://arxiv.org/abs/2211.02701> (cit. on p. 11).
- [9] David Lesage, Elsa D Angelini, Isabelle Bloch, and Gareth Funka-Lea. «A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes». In: *Medical image analysis* 13.6 (2009), pp. 819–845 (cit. on p. 15).
- [10] Alejandro F Frangi, Wiro J Niessen, Romhild M Hoogeveen, Theo Van Walsum, and Max A Viergever. «Model-based quantitation of 3-D magnetic resonance angiographic images». In: *IEEE Transactions on medical imaging* 18.10 (1999), pp. 946–956 (cit. on p. 15).
- [11] Stephen Aylward, Elizabeth Bullitt, S Pizer, and David Eberly. «Intensity ridge and widths for tubular object segmentation and description». In: *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*. IEEE. 1996, pp. 131–138 (cit. on p. 15).
- [12] Onno Wink, Wiro J Niessen, and Max A Viergever. «Fast delineation and visualization of vessels in 3-D angiographic images». In: *IEEE transactions on medical imaging* 19.4 (2002), pp. 337–346 (cit. on p. 15).
- [13] Daniel Rueckert, Peter Burger, SM Forbat, RD Mohiaddin, and Guang-Zhong Yang. «Automatic tracking of the aorta in cardiovascular MR images using deformable models». In: *IEEE Transactions on medical imaging* 16.5 (2002), pp. 581–590 (cit. on p. 15).
- [14] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. «Multiscale vessel enhancement filtering». In: *International conference on medical image computing and computer-assisted intervention*. Springer. 1998, pp. 130–137 (cit. on p. 15).
- [15] Tobias Boskamp, Daniel Rinck, Florian Link, Bernd Kummerlen, Georg Stamm, and Peter Mildenerger. «New vessel analysis tool for morphometric quantification and visualization of vessels in CT and MR imaging data sets». In: *Radiographics* 24.1 (2004), pp. 287–297 (cit. on p. 15).
- [16] Yuxin Ma, Yang Hua, Hanming Deng, Tao Song, Hao Wang, Zhengui Xue, Heng Cao, Ruhui Ma, and Haibing Guan. «Self-supervised vessel segmentation via adversarial learning». In: *proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 7536–7545 (cit. on p. 15).
- [17] Boah Kim, Yujin Oh, and Jong Chul Ye. «Diffusion adversarial representation learning for self-supervised vessel segmentation». In: *arXiv preprint arXiv:2209.14566* (2022) (cit. on p. 15).

- [18] Tianyi Shi, Xiaohuan Ding, Liang Zhang, and Xin Yang. «FreeCOS: self-supervised learning from fractals and unlabeled images for curvilinear object segmentation». In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2023, pp. 876–886 (cit. on p. 16).
- [19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. «U-net: Convolutional networks for biomedical image segmentation». In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241 (cit. on p. 16).
- [20] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. «Unet++: A nested u-net architecture for medical image segmentation». In: *International workshop on deep learning in medical image analysis*. Springer. 2018, pp. 3–11 (cit. on p. 16).
- [21] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. «Transunet: Transformers make strong encoders for medical image segmentation». In: *arXiv preprint arXiv:2102.04306* (2021) (cit. on p. 16).
- [22] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. «Swin-unet: Unet-like pure transformer for medical image segmentation». In: *European conference on computer vision*. Springer. 2022, pp. 205–218 (cit. on p. 16).
- [23] Giles Tetteh, Velizar Efremov, Nils D Forkert, Matthias Schneider, Jan Kirschke, Bruno Weber, Claus Zimmer, Marie Piraud, and Björn H Menze. «Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes». In: *Frontiers in Neuroscience* 14 (2020), p. 592352 (cit. on p. 16).
- [24] Lei Mou et al. «CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging». In: *Medical image analysis* 67 (2021), p. 101874 (cit. on p. 16).
- [25] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylyka, Josien PW Pluim, Ulrich Bauer, and Bjoern H Menze. «clDice-a novel topology-preserving loss function for tubular structure segmentation». In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pp. 16560–16569 (cit. on p. 16).
- [26] Rushi Jiao, Yichi Zhang, Le Ding, Bingsen Xue, Jicong Zhang, Rong Cai, and Cheng Jin. «Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation». In: *Computers in Biology and Medicine* 169 (2024), p. 107840 (cit. on p. 17).

- [27] Han Zheng et al. «Semi-supervised segmentation of liver using adversarial learning with deep atlas prior». In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2019, pp. 148–156 (cit. on p. 17).
- [28] Huifeng Yao, Xiaowei Hu, and Xiaomeng Li. «Enhancing pseudo label quality for semi-supervised domain-generalized medical image segmentation». In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 36. 3. 2022, pp. 3099–3107 (cit. on p. 17).
- [29] Xiaoyan Wang, Yiwen Yuan, Dongyan Guo, Xiaojie Huang, Ying Cui, Ming Xia, Zhenhua Wang, Cong Bai, and Shengyong Chen. «SSA-Net: Spatial self-attention network for COVID-19 pneumonia infection segmentation with semi-supervised few-shot learning». In: *Medical image analysis* 79 (2022), p. 102459 (cit. on p. 17).
- [30] Caizi Li et al. «Self-ensembling co-training framework for semi-supervised COVID-19 CT segmentation». In: *IEEE Journal of Biomedical and Health Informatics* 25.11 (2021), pp. 4140–4151 (cit. on p. 17).
- [31] Kai Han, Lu Liu, Yuqing Song, Yi Liu, Chengjian Qiu, Yangyang Tang, Qiaoying Teng, and Zhe Liu. «An effective semi-supervised approach for liver CT image segmentation». In: *IEEE Journal of Biomedical and Health Informatics* 26.8 (2022), pp. 3999–4007 (cit. on p. 17).
- [32] Antti Tarvainen and Harri Valpola. «Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results». In: *Advances in neural information processing systems* 30 (2017) (cit. on p. 17).
- [33] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, and Pheng-Ann Heng. «Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation». In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2019, pp. 605–613 (cit. on p. 17).
- [34] Fabian Isensee et al. «nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation». In: (2018). arXiv: 1809.10486 [cs.CV]. URL: <https://arxiv.org/abs/1809.10486> (cit. on p. 18).
- [35] Han Zhong, Jiatian Zhang, and Lingxiao Zhao. «MedSAM/MedSAM2 Feature Fusion: Enhancing nnUNet for 2D TOF-MRA Brain Vessel Segmentation». In: *Journal of Imaging* 11.6 (2025), p. 202 (cit. on p. 18).
- [36] Francesco Galati, Daniele Falcetta, Rosa Cortese, Barbara Casolla, Ferran Prados, Ninon Burgos, and Maria A. Zuluaga. *A2V: A Semi-Supervised Domain Adaptation Framework for Brain Vessel Segmentation via Two-Phase Training Angiography-to-Venography Translation*. 2024. arXiv: 2309.06075 [eess.IV]. URL: <https://arxiv.org/abs/2309.06075> (cit. on p. 19).

- [37] Natalia Valderrama, Ioannis Pitsiorlas, Luisa Vargas, Pablo Arbeláez, and Maria A. Zuluaga. «Job-VS: Joint Brain-Vessel Segmentation in TOF-MRA Images». In: (2023), pp. 1–5. DOI: 10.1109/ISBI53787.2023.10230406 (cit. on p. 19).
- [38] Marshall Xu et al. «VesselBoost: A python toolbox for small blood vessel segmentation in human magnetic resonance angiography data». In: *bioRxiv* (2024) (cit. on p. 20).
- [39] Chethan Radhakrishna, Karthikesh Varma Chintalapati, Sri Chandana Hudukula, Ram Kumar, Raviteja Sutrave, Hendrik Mattern, Oliver Speck, Andreas Nürnberger, and Soumick Chatterjee. «Spockmip: Segmentation of vessels in mras with enhanced continuity using maximum intensity projection as loss». In: *arXiv preprint arXiv:2407.08655* (2024) (cit. on p. 20).
- [40] Aditi Deshpande, Nima Jamilpour, Bin Jiang, Chelsea Kidwell, Max Wintermark, and Kaveh Laksari. *Automatic Segmentation, Feature Extraction and Comparison of Healthy and Stroke Cerebral Vasculature*. 2020. arXiv: 2002.11208 [q-bio.QM]. URL: <https://arxiv.org/abs/2002.11208> (cit. on pp. 21, 22).
- [41] Bilal Bayram, Ismail Kunduracioglu, Suat Ince, and Ishak Pacal. «A systematic review of deep learning in MRI-based cerebral vascular occlusion-based brain diseases». In: *Neuroscience* 568 (2025), pp. 76–94. ISSN: 0306-4522. DOI: <https://doi.org/10.1016/j.neuroscience.2025.01.020>. URL: <https://www.sciencedirect.com/science/article/pii/S0306452225000223> (cit. on pp. 21, 22).
- [42] Langtao Zhou, Huiting Wu, Guanghua Luo, and Hong Zhou. «Deep learning-based 3D cerebrovascular segmentation workflow on bright and black blood sequences magnetic resonance angiography». In: *Insights into Imaging* 15 (2024), p. 81. DOI: 10.1186/s13244-024-01657-0 (cit. on pp. 21, 22).
- [43] Guanghao Zhu, Xinjie Luo, Ting Yang, Liang Cai, Joon Hock Yeo, Guang Yan, and Jie Yang. «Deep learning-based recognition and segmentation of intracranial aneurysms under small sample size». In: *Frontiers in Physiology* 13 (2022), p. 1084202. DOI: 10.3389/fphys.2022.1084202 (cit. on p. 21).
- [44] Oscar Esteban, Daniel Birman, Marie Schaer, Oluwasanmi O Koyejo, Russell A Poldrack, and Krzysztof J Gorgolewski. «MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites». In: *PloS one* 12.9 (2017), e0184661 (cit. on p. 24).
- [45] Eduard T Klapwijk, Ferdi Van De Kamp, Mara Van Der Meulen, Sabine Peters, and Lara M Wierenga. «Qoala-T: A supervised-learning tool for quality control of FreeSurfer segmented MRI data». In: *Neuroimage* 189 (2019), pp. 116–129 (cit. on p. 24).

- [46] Yiyang Pan, Kevin Kahru, Emma Barinas-Mitchell, Tamer S Ibrahim, Carmen Andreescu, and Helmet Karim. «Measuring arterial tortuosity in the cerebrovascular system using Time-of-Flight MRI». In: *medRxiv* (2024) (cit. on p. 33).
- [47] Kaiyuan Yang et al. *Benchmarking the CoW with the TopCoW Challenge: Topology-Aware Anatomical Segmentation of the Circle of Willis for CTA and MRA*. 2025. arXiv: 2312.17670 [cs.CV]. URL: <https://arxiv.org/abs/2312.17670> (cit. on p. 46).
- [48] QingLin Liu, Peng Jiang, YuHua Jiang, HuiJian Ge, ShaoLin Li, HengWei Jin, and YouXiang Li. «Prediction of aneurysm stability using a machine learning model based on PyRadiomics-derived morphological features». In: *Stroke* 50.9 (2019), pp. 2314–2321 (cit. on p. 50).
- [49] Marius E Mayerhoefer, Andrzej Materka, Georg Langs, Ida Häggström, Piotr Szczypiński, Peter Gibbs, and Gary Cook. «Introduction to radiomics». In: *Journal of Nuclear Medicine* 61.4 (2020), pp. 488–495 (cit. on p. 50).
- [50] Fengze Liu, Yingda Xia, Dong Yang, Alan L Yuille, and Daguang Xu. «An alarm system for segmentation algorithm based on shape model». In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 10652–10661 (cit. on p. 55).
- [51] Nathan Painchaud, Youssef Skandarani, Thierry Judge, Olivier Bernard, Alain Lalande, and Pierre-Marc Jodoin. «Cardiac segmentation with strong anatomical guarantees». In: *IEEE transactions on medical imaging* 39.11 (2020), pp. 3703–3713 (cit. on p. 55).
- [52] Shuo Wang et al. «Deep generative model-based quality control for cardiac MRI segmentation». In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2020, pp. 88–97 (cit. on p. 58).