WHOLE-MOUSE BRAIN VASCULAR ANALYSIS FRAMEWORK: SYNTHETIC MODEL-BASED VALIDATION, INFORMATICS PLATFORM, AND QUERYABLE DATABASE

A Dissertation

by

MICHAEL RICHARD NOWAK

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Yoonsuck Choe
Committee Members,	Ricardo Gutierrez-Osuna
	John Keyser
	Jim Ji
Head of Department,	Scott Schaefer

December 2019

Major Subject: Computer Science

ABSTRACT

The past decade has seen innovative advancements in light microscopy instrumentation that have afforded the acquisition of whole-brain datasets at micrometer resolution. As the hardware and software used to automate the traditional neuroanatomical workflow become more accessible to researchers around the globe, so will the tools needed to analyze whole-brain datasets. Only recently has the focus begun to shift from the development of instrumentation towards platforms for data-driven quantitative analyses. As a consequence of this, the tools required for large-scale quantitative studies across the whole brain are few and far between. In this dissertation, we aim to change this through the development of a standardized, quantitative approach to the study of whole-brain, cerebrovasculature datasets.

Our standardized and quantitative approach has four components. The first is the construction of synthetic cerebrovasculature models that can be used in conjunction with the second component, a model-based validation system. Any cerebrovasculature study conducted using imaging data must first extract the filaments embedded within that dataset. The segmentation algorithms that are commonly used to do this are frequently validated on small-scale datasets that represent only a small selection of cerebrovasculature variability. The question is how do these algorithms perform when applied to large-scale datasets. Our model-based validation system uses biologically inspired, large-scale datasets that asses the accuracy of the segmentation algorithm output against ground truth data.

Once the data is segmented, we have implemented an informatics platform that calculates descriptive statistics across the entire volume. Attributes describing each vascular filament are also calculated. These include measures of vascular radius, length, surface area, volume, tortuosity, and others. The result is a massive amount of data for the cerebrovasculature segments. The question becomes how can this be analyzed sensibly.

Given that both cerebrovasculature topology and geometry can be capture in graph form, we construct the fourth component of our system: a graph database that stores the cerebrovasculature.

The graph model of cerebrovasculature that we have developed allows segments to be searched across the whole-brain based on their attributes and/or location. We also implemented a means to reconstruct the segments returned by a specific query for visualizations. This means that a simple text-based query can retrieve cerebrovasculature geometry and topology of the specified vasculature. For example, a query can return all vessels within the frontal cortex, those with specific attribute(s) value range(s), or any combination of attribute and location. Complex graph algorithms can also be applied, such as the shortest path between two bifurcation points or measures of centrality that are important in determining the robust and fragile aspects of blood flow through the cerebrovasculature system.

To illustrate the utility of our system, we construct a whole-brain database of vascular connectivity from the Knife-Edge Scanning Microscope India Ink dataset. Using our cerebrovasculature database, we were able to study the cerebrovasculature system by issuing text-based queries to extract the vessel segments that we were interested in. The outcome of our investigation was a wealth of information about the cerebrovasculature system as a whole, and about the different classifications of vessels comprising it. The results returned from these simple queries even generated some interesting and biologically relevant questions. For instance, the profound spikes in radius distribution for some classes of vessels that did not present in other classes.

We expect that the methods described in this dissertation will open the door for data-driven, quantitative investigation across the whole-brain. At the time of writing – and to the best of our knowledge that prior to this work – there was not a systemic way to assess segmentation algorithm performance, calculate attributes for each segment of vasculature extracted across the whole brain, and store those results in a queryable database that also stores geometry and topology of the entire cerebrovasculature system. We believe that our method can and will set the standard for large-scale cerebrovasculature research. Therefore, in conclusion, we state that our methods contribute a standardized, quantitative approach to the study of cerebrovasculature datasets acquired using modern imaging techniques.

DEDICATION

To my mother and father.

ACKNOWLEDGMENTS

First and foremost, I want to acknowledge the unwavering support of my advisor, Yoonsuck Choe, throughout my doctoral studies. I've learned a lot from you over the years as your student. I would also like to express my sincere gratitude to Ricardo Gutierrez-Osuna for the advice he's given me throughout my studies and for his sense of humor that brightens up the day. John Keyser, for hours he spent reviewing this dissertation with me and the many off-topic discussions we'd have in the process. Jim Ji, for the perspective that he provided as an external committee member. The guidance, honesty, and support that each of you provided were invaluable throughout my doctoral studies. To my committee members mentioned above, thank you for mentoring me during my studies and for your support, perspective, and constructive criticism.

Outside of my committee members, I would like to thank additional faculty members who have been supportive over the years. Michael Moore, I've learned so much from you about teaching and life in general. I sincerely appreciate the time you've taken to mentor me, for always having an open door, and for all the tea you've shared. Dilma da Silva, for supporting my interest in teaching while department head and taking the time to mentor me during our collaborative initiatives. I would also like to thank Karrie Bourquin in advising for helping navigate through the processes involved with the many milestones associated with my degree.

I would also like to thank my friends and lab mates at Texas A&M University. Chris Liberatore, for the thoughtful discussions that we've had about research and life over the years. Jaewook Yoo, for providing invaluable advice and perspective from his own experiences along the way towards his doctoral degree. I took your recommendation and have not updated my computer for years (since your story that an update broke necessary code). Junseok Lee, for all your help and support over the many years we were confined to shared office space. Once you graduated, I finally had space for myself! However, by day one, I was already missing our daily conversations; the office wasn't the same. To the individuals mentioned above, my other labmates (Khoung Nguyen, Wookyung An, Han Wang, Sungjun Lim, and Qing Wan), and those I may have forgotten: thank

you for everything.

Most importantly, I need to express my thanks and gratitude to my parents for their steadfast support during my studies. Without your love, inspiration, drive and help you've given me, I would not be where I am today. I can never thank you enough for always being there for me.

Michael R. Nowak

September 2019

College Station, TX

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Yoonsuck Choe (advisor), Professors John Keyser and Ricardo Gutierrez-Osuna of the Department of Department of Computer Science and Engineering and Professor Jim Ji of the Department of Electrical Engineering. All work for the dissertation was completed independently by the student. The whole-brain mouse dataset analyzed for Chapter 8 was provided by Junseok Lee; this dataset had been acquired using Knife-Edge Scanning Microscopy by David Mayerich and Jaerock Kwon.

Funding Sources

Graduate study was supported by graduate assistantships from the Department of Computer Science and Engineering at Texas A&M University. This research was also funded in part by the NIH/NINDS MSM grant and NSF (#0079874, #0905041, #1208174, #1256086).

NOMENCLATURE

3D	Three dimensional
AOH	All-optical histology
BBB	Blood brain barrier
CBF	Cerebral blood flow
ER	Experimental result
fMOST	Fluorescence micro-optical sectioning tomography
GT	Ground truth
inf	Infinity
KESM	Knife-edge scanning microscopy
KESMBA	Knife-edge scanning microscopy brain atlas
LEN:DIA	Length-to-diameter ratio
LM	Light microscopy
LSFM	Light sheet fluorescence microscopy
MCA	Middle cerebral artery
MOST	Micro-optical sectioning tomography
ROI	Region of interest
SA	Segmentation algorithm
SA:V	Surface-area-to-volume ratio
STPT	Serial two-photon tomography
TPTC	Two-photon tissue cytometry

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

The introduction of three-dimensional light microscopy (LM) instrumentation capable of imaging the whole-murine brain at sub-micrometer resolution has helped integrate our knowledge about cellular- and systemic-level aspects of the cerebrovasculature system. By bridging the critical gap in neuroimaging between large-scale, low-resolution methods (e.g., magnetic resonance imaging) and low-scale, high-resolution techniques (e.g., serial block-face electron microscopy), these cutting-edge imaging modalities have shifted neuroanatomy research away from investigator-driven studies of single anatomical pathways, and towards data-driven investigations across the entire brain. This change in perspective has been made possible by recent advancements that overcome the limited axial optical resolution of LM by either incorporating mechanical tissue sectioning with block-face microscopy or through the application of light-sheet fluorescence microscopy to chemically cleared tissue. By combining tissue sectioning (whether mechanically or optically) with image acquisition, a large part of the traditional neuroanatomical workflow is automated, making it possible to examine large volumes of tissue at LM resolution with high-throughput. The application of such technologies has resulted in the acquisition of complete whole-brain capillary-level datasets for mice, which are frequently disseminated online in atlas form. Although the atlas format provides an interactive environment for the multi-scale exploration of the cerebrovasculature system – from the major arteries and veins down to the individual capillaries of specific cerebral regions – a quantitative investigation into the characteristics and connectivity of the vascular elements requires significant processing to extract the embedded filaments and analyze the networks that they comprise. This processing requirement continues to impede quantitative investigations across the whole-brain; alleviating this barrier would allow researchers to study the variation of cerebrovasculature structure within and between healthy and diseased brains. Such research is fundamental towards our understanding of the changes of the cerebrovasculature system that occurs in many pathological conditions, such as cardiovascular disease, diabetes mellitus, sepsis syndrome, and pathologies hemorheologic in nature. In support of whole-brain vascular

studies, there is a critical need to foster a standardized graph encoding of cerebrovasculature connectivity data and cultivate an open-source informatics framework that facilitates the quantitative analysis of the encoded cerebrovascular system.

My *long-term goal* is to chart the local variation in vascular morphometry within the whole mouse brain, to document the relationship of such change to the composition of surrounding neurons, and to elucidate the differences in these relationships between healthy and diseased brains. In support of these pursuits, my *objective* in this dissertation is three-fold. The first is the design of a model-based validation framework capable of assessing the accuracy of vascular segmentation algorithms on synthetic datasets. Second is the development of an informatics platform for the quantitative analysis of cerebrovascular structural attributes and volumetric characteristics. Finally, the third is the quantification of the cerebrovasculature embedded in the Knife-Edge Scanning Microscope Brain Atlas India ink whole-brain dataset. My *rationale* for these objectives is that the development of data-driven, quantitative investigation at the whole-brain scale will enable the discovery of novel biomarkers of processes underlying healthy and diseased brains. The Brain Networks Laboratory at Texas A&M University has developed one of the first imaging modalities capable of imaging whole-murine brains at sub-micrometer resolution (i.e., Knife-Edge Scanning Microscopy): we have introduced amongst the first whole-brain cerebrovasculature atlases and have an understanding of the current state and limitations of quantitative investigation using whole-brain datasets. The following are the specific aims of this dissertation:

Specific Aim #1: Synthetic cerebrovasculature models. Construct biologically-grounded synthetic models for use in the model-based validation system.

Specific Aim #2: Model-based validation system. Design a validation system that assesses the accuracy of segmentation algorithm output (i.e., either centerline or volumetric) against ground truth data.

Specific Aim #3: Informatics platform. Develop an open-source informatics platform that: (1) converts centerline data to graph representation; (2) imports graph representation to Neo4j Graph Analytics; and, (3) calculates descriptive statistics describing vascular filament attributes across

every element in the model, while characterizing the encoded cerebrovascular volume as a whole.

Specific Aim #4: Whole-brain analysis. Perform quantitative analysis across the Knife-Edge Scanning Microscope Brain Atlas India ink whole-brain dataset: (1) assess algorithmic performance of the segmentation algorithm used to extract the vascular filaments using the model-based validation system; (2) use the informatics platform to quantitatively the describe specific attributes and characteristics of the mouse cerebrovascular system; and, (3) disseminate raw connectivity data through the graph database online.

These studies are *innovative* in that they provide the foundation for data-driven, quantitative investigation of the vasculature system across whole-organs of small animal models. The three primary objectives of this dissertation are as follows. Our first objective is to construct a validation system to help understand how the segmentation algorithms of yesterday scale across the larger-scale, higher-resolution, and more complete datasets of today. Secondly, an open-source informatics framework will be released to the research community for use in quantitative analyses of cerebrovasculature datasets. Finally, descriptive statistics detailing vascular morphometry attributes and characteristics will be reported for the entire mouse brain, with raw connectivity data made available online for the first time. The successful completion of these studies is expected to open the door to new data-driven discoveries in cerebrovasculature research, addressing critical scientific questions in the field about molecular transport, cerebral blood flow, and vascular dynamics.

2. BACKGROUND

2.1 Brain angioarchitecture

The close anatomical apposition between the cellular constituents of the central nervous system and the microstructures of the cerebrovascular system reflects an intimate interaction between neuronal activity and vascular dynamics. The structural and functional integrity of the brain critically depends on this interdependence, which ensures that adequate oxygen and nutrients are delivered to neurons by matching local blood flow to metabolic demand [14]. To understand which vessel attributes and cerebrovasculature characteristics are essential towards analysis of the angioarchitecture, one must be acquainted with the basics of blood flow and hemorheology.

2.1.1 How blood flows in arteries and veins

In general, the role of the macrocirculation can be taken primarily as a transport mechanism, while the microcirculation directly participates in fluid and molecular transference. Arteries of the macrocirculation transport oxygenated and nutrient-rich blood away from the heart and into the cerebrovasculature system, where arteries formulating a uniquely organized topology distribute the incoming blood towards the microvascular beds in the vicinity of neurons and cellular constituents in need. This choreography commences from the carotid and vertebral arteries, where the blood velocity and pressure are the greatest in the cerebrovasculature, and through junctions and offshoots, blood is distributed to proximal arteries. As the blood continues to make its way towards the capillaries, the divergent nature of the arteriole system begins to reveal itself at every branch point, where the diameter of the branching vessel(s) (i.e., daughter vessel[s]) and the width of the primary vessel is reduced [15].

The blood vessels comprising the vascular filaments of the arterial circulation are composed of three distinct layers [16]: a thick and relatively muscular wall called the tunica adventitia, which ensures an aptitude to transport blood under high pressure (approximately 40 mmHg to 80 mmHg [15]); the tunica media which are capable of dilating and constricting the vessel in response to

local humoral control and neuronal input; the tunica intima, the interior layer, which is composed of endothelial cells and a connective tissue layer, and in contact with the blood. Arteries are nearly uniformly cylindrical in their appearance by design: this helps distribute pressure across the entire volume equally [15].

After the blood passes through arteries, it enters smaller arterioles, the last branch before the microcirculation. Participation in the exchange of oxygen and nutrients for waste products in the microvascular beds then begins. Deoxygenated blood is subsequently accepted by the venous system return to the heart; venous transport is carried out under significantly less pressure (frequently less than five mmHg) [15]. The low hydrostatic pressure prevents veins from having a typical structure. While the venous filaments are comprised of the same three distinct layers as arteries are (tunica adventitia, tunica media, tunica intima), the composition of each layer is slightly different, which is responsible for the structural differences noted [16].

2.1.2 Blood rheology

Affecting the nature of flow is the composition of blood: a two-phase fluid formed by blood cells and plasma. In an average adult, about 40%-45% of blood is the cellular component (defined as hematocrit), the remainder is plasma. The cellular component of blood is comprised of approximately 99.7% erythrocytes ($5{,}000{,}000\,\mu\mathrm{L}^{-1}$), 0.2% leukocytes ($7{,}500\,\mu\mathrm{L}^{-1}$; i.e., neutrophils, lymphocytes, monocytes, eosinophils, basophils, and plasma cells), and 0.1% thrombocyte ($250{,}000\,\mu\mathrm{L}^{-1}$) [15]. Meanwhile, the plasma component is about 92% water, 7% plasma proteins (i.e., albumin, globulins, fibrinogen, and others), and 1% other solutes (electrolytes, nutrients, wastes) [id.].

Plasma functions as a Newtonian fluid and has a viscosity of around $1.2\,\mathrm{cP}$; meanwhile, whole blood has a variable viscosity concerning shear rate, hematocrit, temperature, and pathophysiology [15]. Under standard laminar, steady, fully developed flow conditions in a cylindrical tube, shear stress is zero at the centerline; blood at the midline moves as a rigid body; blood lateral to the midline will flow as a normal viscous fluid. This results in a blunted velocity profile for blood flow in the macrocirculation [id.].

2.1.3 Pressure, flow, and resistance

The simple approach to calculating the flow through a blood vessel (Q) is to express the relationship between the pressure difference (ΔP) across the vessel and the resistance (R) to flow through it in Ohm's Law form [15]:

$$Q = \frac{(\Delta P)}{R} = \frac{P_1 P_2}{R} \tag{2.1}$$

While this equation provides some insight into the flow, it oversimplifies the underlying biology [15]. A more helpful estimate for volumetric flow rate (Q) is taken using Hagen-Poiseuille's Law, which can be solved using the volumetric flow rate for laminar fluid flow within a cylinder. The Hagen-Poiseuille solution is derived from integrating the velocity of the fluid (as calculated using Navier-Stokes equations) concerning blood vessel area [id.]:

$$Q = \frac{\pi \Delta P R^4}{8 \mu L}. \tag{2.2}$$

The general equation assumes that the fluid in the tube has a constant viscosity μ, a uniform cross-sectional area over the entire length L, and the pressure gradient $\frac{\Delta P}{L}$ across the tube of radius r and length L [15]. It nonetheless states that flow rate (volume/time) is directly proportional to the product of the radius to the fourth power and the pressure drop across the vessel; is inversely proportional to its length. This approximation works well for arterioles but degrades slightly when applied to venules (due to their atypical shape) [id.].

The Hagen-Poiseuille solution to volumetric flow rate shows us that small changes in vessel radius can cause substantial changes in flow rate. This means that the blood vessel radius does not need to change significantly to increase (or decrease) flow in a given region of the brain, thereby limiting the extent to which dilation (or constriction) is required under hypoxic (or hyperoxic) conditions [15].

2.1.4 Flow separation at bifurcations

A bifurcation is a point where the flow rate in a vascular segment (i.e., the parent branch) is distributed amongst two child branches; the geometry of the bifurcation determines such division, controlling peripheral flow distribution and influencing vascular resistance [15]. A relationship between the parent and child vessels can be derived such that the work involved in the passage of blood through the bifurcation is minimized. Contributing towards this end is the observation that bifurcation branches frequently stay in one common plane under physiological conditions, helping to retain the momentum of blood flow across points of divergence [17]. Murray's law for the minimum work – a mathematical representation to optimize work at a vessel branch point – has been shown to fit biological data remarkably well in validation studies. The equations for Murray's law are represented as [15]:

$$\begin{aligned} r_1^3 &= r_2^3 + r_3^3 \\ \cos(\theta) &= \frac{r_1^4 + r_2^4 - r_4^3}{2r_1^2 r_2^2} \\ \cos(\phi) &= \frac{r_1^4 - r_2^4 + r_4^3}{2r_1^2 r_3^2} \\ \cos(\theta + \phi) &= \frac{r_1^4 - r_2^4 - r_4^3}{2r_2^2 r_3^2} \end{aligned} \tag{2.3}$$

where r_1 is the radius of the parent vessel, r_2 and r_3 are the radii of the two child vessels, θ is the angle between the midline of the parent and child vessel of radius r_2, ϕ is the angle between the midline of the parent and child vessel of radius r_3.

When blood flow arrives at a branch point, its velocity moves the flow towards the inner walls of the bifurcation and away from the outer walls, producing a skewed velocity profile: the highest speed is no longer at the centerline but positioned laterally adjacent the inner wall [15]. As the characteristics of the flow have changed, the flow is no longer entirely developed. The wall shear stress increases along the inner wall and decreases along the outer wall. Eventually, in each child branch, the pressure in the flow direction will balance the wall shear stress and the flow will become

fully developed again [id.].

2.1.5 Microvasculature

The regulation of blood flow through the arterioles that are on the order of $10\,\mu\mathrm{m}$ to $15\,\mu\mathrm{m}$ in diameter is critical in maintaining the capillary beds; a capillary is defined by divergent flow at its inlet and convergent flow at its outlet [15]. In the microcirculation, flow regimes are naturally characterized by low Reynolds numbers (RE), indicating that the inertial forces acting upon the flow are negligible; that the viscous forces dominate the flow [18].

In the microcirculation, the mean blood cell velocity becomes higher than the mean blood velocity; this is in contrast to the higher mean blood velocity to mean blood cell velocity observed in the macrocirculation [15]. This change is perceived as the blood vessel diameter decreases towards that of the suspended cells: red blood cells begin to congregate along the midline as the diameter of the vessel reduces to approximately $15\,\mu\mathrm{m}$ forcing the plasma to flow laterally at a slower velocity. This has a profound effect on the distribution of hematocrit within the microcirculation: given that red blood cells move into the high-velocity flow at midline, they will tend to follow the higher velocity pathway through a bifurcation, provided that the flow velocity is different in the child branches (and ignoring cell skimming into one of the branches) [id.].

Given the heterogeneous nature of blood in the microcirculation, blood may not be characterized as a continuum, and molecular effects must be considered when analyzing flow [15]. Therefore, calculating the bulk properties of fluid particles within the microcirculation is not advisable given that only a few molecules will present across the small sample, making quantification nearly impossible. Therefore, when calculating volumetric flow rate in the microcirculation, the 'apparent viscosity' of the blood is used in place of bulk viscosity [id.],

$$Q = \frac{\pi \Delta P R^4}{8 \mu_{ap} L}, \tag{2.4}$$

where μ_{ap} is the apparent viscosity of the blood. The apparent viscosity is dependent on the diameter of the blood vessel and the pressure drop across it; it decreases with vessel radius but

increases rapidly when blood vessel diameter diminishes to that of the red blood cells.

Once the capillaries are reached, the red blood cells enter single-file, while pushing the plasma steadily against the vessel walls [15]; when enough plasma collects, a plasma packet is forced into the line between the red blood cells. The integrity of the microenvironment is dependent on the highly selective permeability of the brain's blood-brain barrier (BBB); meanwhile, the acquisition of oxygen and nutrients by cerebral cells and for the removal of waste products from their environments [19]. The BBB, formed by specialized endothelial cells held together by tight junctions, imposes strict restrictions on the substances that may enter the brain. Transcapillary exchange by simple diffusion is limited to small inorganic molecules, whereas polar solutes and large insoluble compounds must be ferried across the vessel wall by carrier-mediated transport mechanisms [id.]. Together, these mechanisms afford the intravascular and interstitial spaces an ability to exchange water and specific nutrients; once these substances enter the interstitial space, they can be absorbed by nearby cells [15].

2.1.6 Cerebral blood flow regulation

Cerebral blood flow (CBF) to tissues is regulated by two mechanisms that differ concerning the time duration in which they operate. Chronic local control of CBF over an extended period is exerted through changes in the size of existing blood vessels and the number of vessels occupying a specific cerebral region (angiogenesis). Meanwhile, acute control rapidly redistributes CBF based on tissue oxygenation levels through modification of local arterioles diameter by vasoconstrictive and vasodilative [15].

2.1.7 Organizational principles of the angioarchitecture

The ability of the cerebrovasculature system to deliver adequate oxygen and nutrients to neurons in metabolic demand is highly dependent on systemic and local hemodynamics and hemorheology; the brain's angiome, or distribution of vascular elements, contributes directly to these aspects governing cerebral blood flow (CBF). Furthermore, the angioarchitecture influences angiogenesis, remodeling [20], and structural adaptation [21]. Therefore, elucidating the organizing

principles underlying the brain's angioarchitecture is fundamental towards the understanding of neurovascular coupling in normal and disease states.

Early research into the brain's angioarchitecture was accomplished through labor-intensive conventional histological workflows (and before that, gross anatomical studies) that significantly limited the throughput of data collection; this initially constrained the scope of available studies due to the limited ability of conventional light microscopy to resolve microscopic structures in optically thick samples. Confocal microscopy (the 1960s) and two-photon microscopy (1990s) instrumentation reduced the out-of-focus noise that blurred the image at the focal plane in conventional light microscopy. The advent of these imaging modalities addressed an early need for three-dimensional imaging [22] by providing increased imaging depth and reducing out-of-focus-blur to produce more precise and crisper images. Instrumentation using these techniques implemented in conjunction with laser scanning microscopy has been widely used in cerebrovasculature research.

2.1.7.1 Cerebral cortex

Investigations into the brain's angioarchitecture using confocal laser scanning microscopy, two-photon laser scanning microscopy, or modalities derived from there, have centered on the cerebral cortex, where the morphometry of vessels residing near the cortical surface and in the intracortical region have been studied. Notably, this has resulted in the report of morphometric measurements for diameter, length, and number of elements, as a function of vessel order number on the basis of diameter ranges (diameter-defined Strahler model, [23]), for the lateral aspect of the collateral sulcus on the fusiform gyrus in the human [10, 24] and the arterial network of the pial microcirculation in the rat [25]. In both cases, the frequency distributions of diameters and length were asymmetric with sizeable positive skewness and a leptokurtosis; however, taking the logarithm of the length and the inverse diameter showed linear distributions with similar mean and median, skewness close to zero, and kurtosis close to three [id.]. Moreover, the mean diameter and mean length were shown to grow as a geometric sequence with the order number (albeit some discontinuities) under Horton's Law [26]; this implies that the cerebrovasculature in these regions show self-similarity over many size scales and can be said to be 'fractal'.

For both the collateral sulcus [10, 24] and arterioles of the pial microcirculation [25], a connectivity matrix [23] was derived from the centripetal ordering scheme [i.d.] that reported the ratio of the total number of elements of order n from elements of order m, divided by the total number of elements in order m. Combining such connectivity information with the data describing the mean diameter, mean length, and the number of elements at each order, the total cross-section area and blood volume for each order can easily be calculated. Furthermore, if the inlet pressure and outlet pressure of each order n is known, the connectivity relationship can be used to calculate the flow in a vessel element of order m that springs from a larger vessel of order n; summing this for all vessels of order m yields the total flow in all vessels order m [23]. Given that the data acquired in these studies can be used to analyze pressure, flow, and vascular volume in analog to electrical circuit analysis, the effect of the reported morphometric data on these aspects of hemodynamics can be evaluated [id.].

The first comprehensive analysis of the angioarchitecture at the level of cerebral capillaries was put forward by Kleinfeld and colleagues in murine animal models [27]; much of our knowledge about the organization of the angioarchitecture in the cerebral cortex has been derived from these studies. The focus of these studies were within the territory of the middle cerebral artery (MCA), which includes all branches of the MCA, from the rhinal vein to the anastomoses with the anterior and posterior cerebral arteries (ACA and PCA respectively) [28]; it corresponds half of the total surface area of the cortical mantle for rats at approximately $150\,\mathrm{mm}^2$ and one-third that amount for mice [i.d.]. In this region, Kleinfeld's group has studied the two-dimensional pial network extensively [28] that sources the three-dimensional subsurface microvasculature beds [29], and the penetrating arterioles that connect the pial and subsurface systems [30].

Concerning blood flow, Kleinfeld's group has found that the pial surface network and the subsurface microvasculature network are primarily insensitive to an occlusion of a single vessel; that both systems were highly interconnected and capable of rebalancing flow after the blockage of a single vessel [28, 29]. The three-dimensional subsurface network is reported to form loops with an average of eight edges, while the two-dimensional surface network an average of four sides.

Meanwhile, the penetrating vessels are not interconnected with each other through anastomoses or direct collaterals, and serve as a weak link between the subsurface and surface networks [30]. The loss of flow to a single penetrating arteriole has been found to lead to the loss of perfusion in a columnar region of approximately $500\,\mu\mathrm{m}$ in diameter [31]. Due to the source-sink relationship of penetrating vessels, collateral flow cannot compensate for such an occlusion, meaning that the occlusion of a single penetrating vessel leads to infarction, possibly resulting in cognitive deficit [32].

2.1.7.2 Beyond the cerebral cortex

The highly redundant surface and subsurface vasculature of the cerebral cortex, as it presents within the territory of the MCA, has been shown to maintain perfusion in light of the occlusion of a single arteriole. Occlusion of the base of the MCA has shown that compensation brought forth through anastomoses with the anterior cerebral artery and intact regions of the MCA limits damage in the cerebral cortex / penumbral region [33]. Meanwhile, this same occlusion results in infarction and widespread death in the basal ganglia, apparently due to the lack of anastomoses in the basal ganglia [id.].

High-resolution light microscopy methods, such as confocal laser scanning microscopy and two-photon laser scanning microscopy, mostly focused on the cerebral cortex due to the limited imaging depth that they could resolve; therefore, the angioarchitecture of subcortical nuclei, such as the basal ganglia, have not been studied to the same extent as that of the cerebral cortex. This is set to change with the advent of new imaging techniques, such as STPT and KESM, that combine the physical or optical sectioning with the imaging of tissues [34] to exceed the depth offered by confocal or 2-photon microscopy [34, 2]. In fact, many whole-brain datasets are beginning to emerge. Visualizations techniques have been developed; currently, the focus has shifted from acquisition to the development of informatics methods capable of quantifying the unprecedented amount of data from whole-brain studies.

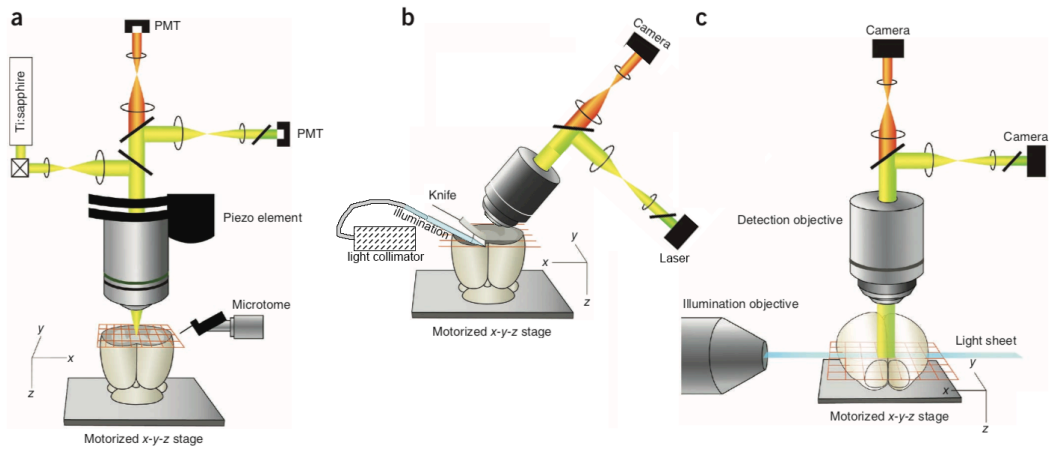


Figure 2.1: Whole-brain LM methods: (a) STPT, image brain in coronal plane in a mosaic grid pattern and microtome off imaged tissue; (b) KESM, image using line-scan camera of a $1\,\mu\mathrm{m}$ thin section illuminated through and cut by a diamond knife; (c) LSFM of chemically cleared brain, illuminate with a light sheet through illumination objective positioned at the side and image in mosaic grid pattern from above. Reprinted, with permission, from [1].

2.2 Whole-brain imaging with light microscopy

The introduction of three-dimensional light microscopy (LM) instrumentation capable of imaging the whole-brain at sub-micrometer resolution (Figure 2.1) has helped integrate our knowledge about the cellular- and systemic-level aspects of the cerebrovasculature system. By abridging the critical gap in neuroimaging between large-scale, low-resolution methods (e.g., magnetic resonance imaging) and low-scale, high-resolution techniques (e.g., serial block-face electron microscopy), neuroanatomy research has shifted away from investigator-driven studies of single anatomical pathways, and towards data-driven investigations across the entire brain [1]. This paradigm shift has been made possible by overcoming the limited axial optical resolution of LM, which has been accomplished by incorporating mechanical tissue sectioning with block-face microscopy and also through the application of light-sheet fluorescence microscopy (LSFM) to chemically cleared tissue [id.]. Furthermore, the combination of tissue sectioning (whether mechanically or optically) with image acquisition has resulted in large part of the traditional neuroanatomical workflow (Figure 2.2) to become automated (Figure 2.3). Automated LM modalities are capable

of whole-brain imaging, such as those shown in Figure 2.1, have made it possible to image large volumes of tissue at the sub-micrometer resolution with high-throughput.

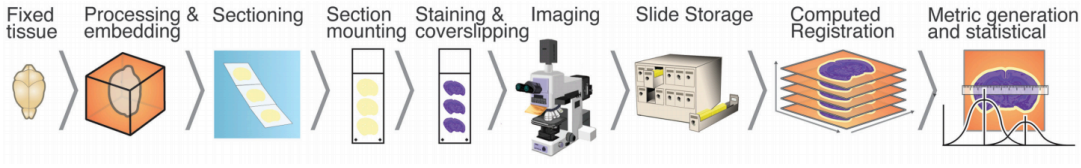


Figure 2.2: Traditional neuroanatomical workflow where tissue is embedded, sectioned, stained, imaged, and then registered to one another to produce three-dimensional image stacks. Reprinted, with permission, from [2].

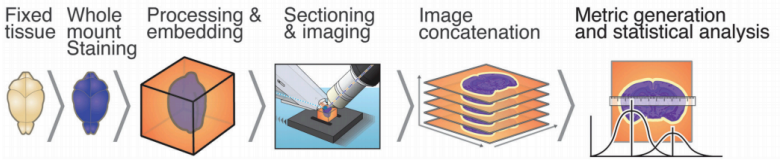


Figure 2.3: Automation of traditional anatomical process (Figure 2.2) by KESM imaging technology, requiring only image concatenation to produce three-dimensional image stacks. Reprinted, with permission, from [2].

2.2.1 Integration of block-face microscopy and tissue sectioning

This approach has been developed for wide-field imaging, line-scan imaging, confocal microscopy, and two-photon microscopy [1, 35, 36, 37, 38, 39, 40, 41, 42]. Common to these modalities is the motorized movement of a stage (upon which the sample is fixed) below the microscope objective, followed by mechanical removal of the imaged tissue. This process ensures that the sample being imaged is always directly below the objective, at a consistent depth during the acquisition process. Normally, the imaging depth of only a few hundred microns can be reached by an optical microscope. However, using tissue sectioning, the absorption and scattering of light in tissue that would otherwise limit the imaging depth in traditional optical microscopy are avoided [43]. The removal of imaged brain tissue ensures this before further imaging. This allows samples to be acquired at submicrometer resolution, close to the diffraction limits of LM.

Two-photon microscopy [44] has been combined with tissue sectioning to create three different instruments; each instrument differs in its sectioning methodology. All-optical histology (AOH) [36] uses ultrasonic laser pulses, while two-photon tissue cytometry (TPTC) [38] applies a milling machine to section the tissue. In both of these methods, the sectioning process damages the tissue, inhibiting subsequent analysis. The third technique based on two-photon microscopy is STPT [41], which employs a vibrating microtome to section the tissue and is shown in Figure 2.1.a. This method allows the sectioned tissue to be collected for follow-up studies after the imaging process (it does not damage the tissue strips) [1].

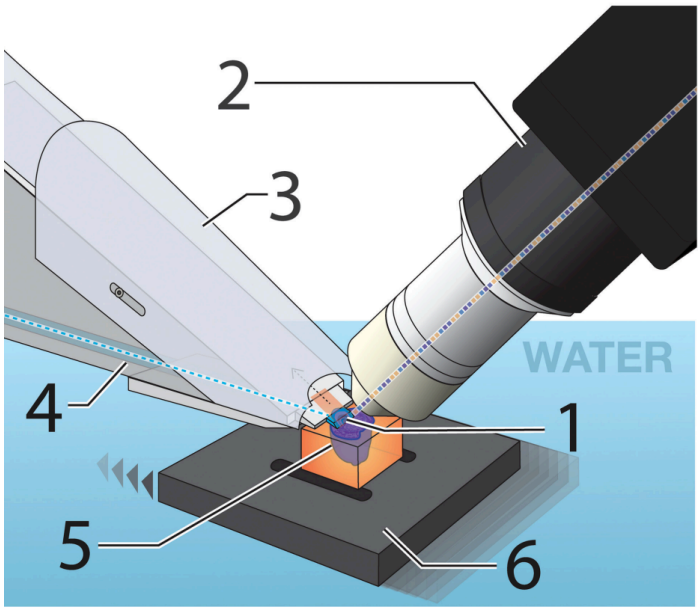


Figure 2.4: KESM instrumentation. The mechanical components include: (1) diamond knife; (2) objective lens; (3) suction channel for tissue removal; (4) fiber optic cable for illumination; (5) embedded tissue sample; (6) motorized stage. Reprinted, with permission, from [2].

Another and different means of whole-brain imaging is knife-edge scanning microscopy (KESM) [39] developed at Texas A&M University. KESM instrumentation, shown in Figure 2.4, combines bright-field line scan imaging and physical sectioning of the tissue by a diamond knife for whole-brain imaging. The KESM features a knife-collimator assembly for illumination and sec-

tioning, and a microscope objective positioned ninety-degrees to this unit and the motorized stage upon which the sample is mounted. Stair-step cutting is the tomographic sectioning mode used to simultaneously section and image across the sample. The resolution afforded by KESM is at approximately $1\,\mu\mathrm{m}$ voxel.

Due to its innovative design, KESM experiences unique noise due to the configuration of its instrumentation, which presents in addition to noise common to LM. Specifically, this unique noise includes knife chatter during the physical sectioning (Figure 2.6) and lighting defects (Figure 2.5) [3]. The lighting defects arise from two primary sources. The first, small misalignments between the knife and camera causing a steady change of in illumination across the sectioning axis of the image; (2), defects in the diamond knife (chips, uneven areas) causing reflection and refraction variations along the knife surface that in turn results in lighter or darker strips extending along the sectioning plane (see white arrows in Figure 2.5) [id.]. Fluctuations in illumination power can also cause differences in illumination over time, causing visible stripes in the image space that vary with sampling rate, but remain constant in the time domain [id.].

The lighting defects described are easily removed from KESM images because they formulate regular patterns; however, knife chatter (via knife vibrations) is more difficult to remove due to its irregularity [3]. Knife chatter typically presents across the sectioning plane as changes in illumination and arises due to tissue sectioning with its simultaneous imaging, making this type of noise rather unique to KESM-based modalities (see white arrows in Figure 2.6). While sampling very near to the knife edge is the primary reason for this kind of noise, sampling near the knife edge is necessary to ensure alignment between physical sections and to preserve the integrity of each section (from being warped or torn due to knife friction or water suction).

Micro-optical sectioning tomography (MOST) [40] is based off KESM technology. It is different from KESM [39] as it decouples the knife-collimator assembly, which has been shown to reduce chatter [43]. The image is also acquired using column-by-column tomographic sectioning in the same layer, with each layer immersed in water for a short but constant period. This process is thought to reduce the non-uniform deformation of the microstructures that is said to occur

Figure 2.5: Light intensity inconsistencies presenting in KESM imaging. The visible intensity shift along the x-axis is due to knife misalignment; defects in the diamond knife are responsible for the streaks extending along the y-axis (white arrows). Reprinted, with permission, from [3].

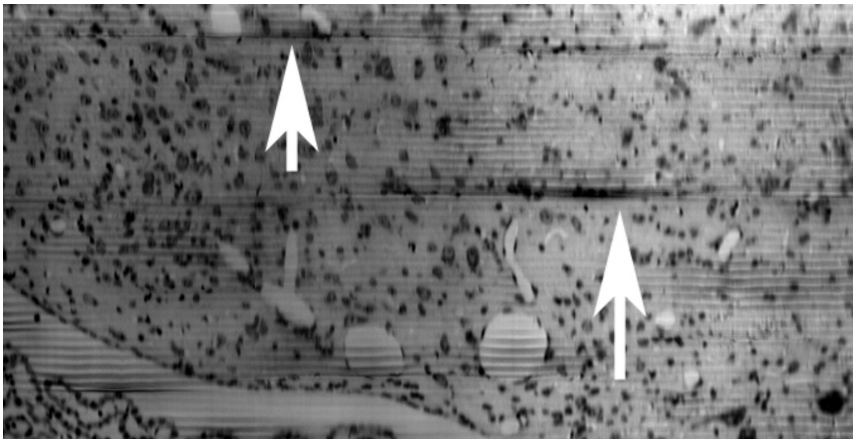


Figure 2.6: Knife chatter in KESM imaging, presenting along the x-axis of the image as darker bands; significant cases pointed out by the white arrows. Reprinted, with permission, from [3].

with KESM. MOST was recently configured for fluorescence imaging (fMOST) by using a tissue embedding protocol and employing confocal laser scanning microscopy [id.].

2.2.2 Light sheet fluorescence microscopy of chemically cleared brains

The combination of two century-old techniques, LSIM [45] and chemical clearing [46], by [47] provides another means to limit interference from out-of-focus background light. This technique, termed ultramicroscopy, illuminates a thin layer of the transparent brain that is imaged perpendicularly with wide-field LM. This works because only the tissue in the path of the light is illuminated. The whole brain is acquired by either shifting the optical imaging light path or moving the sample

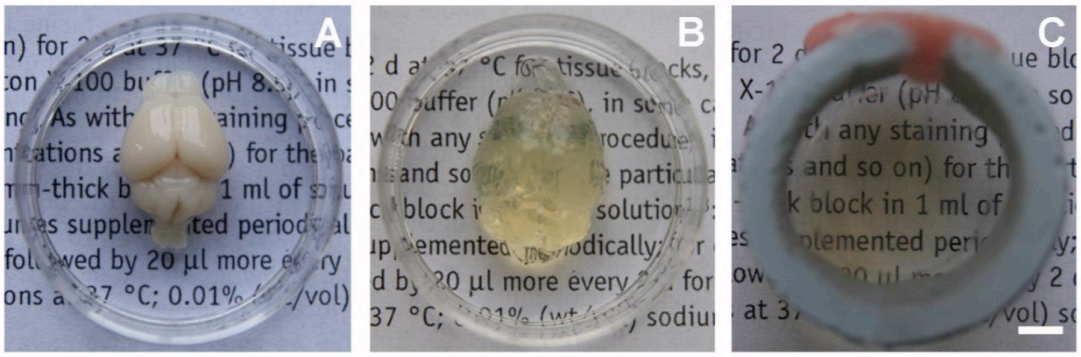


Figure 2.7: CLARITY method for rendering the whole-brain optically transparent: (a) whole-mouse brain before CLARITY application; (b) after removing lipid bilayers; (c) after refractive index matching. The scale bar is $1\,\mathrm{mm}$. Reprinted, with permission, from [4].

through it [43]. The whole mouse brain can be imaged using this methodology without the need for mechanical sectioning [1].

Critical to this technique is the optical clearing methodology applied to the brain, for the generation of transparent tissue. One of the well known and thus more popular techniques for achieving this is the CLARITY clearing method [48]. This method begins with a whole brain (Figure 2.7.a), employs electrophoresis to remove lipids from cerebral tissue (Figure 2.7.b), and then uses refractive index matching rendering the tissue transparent (Figure 2.7.c). Other techniques (e.x., [49, 50, 51, 52]) transform the tissue to clear form by immersing the brain in a clearing reagent. According to [43], this ensures that scatters in fixed tissues are resolved with optical clearing reagents, prior making the reactive index of the cerebral tissue uniform to reduce light scattering (thus rendering the tissue transparent).

While this technique is non-destructive (the tissue can be imaged multiple time; and quickly at that), it is currently not possible to render the brain completely transparent. As the imaging commences into the depths of the brain from its surface, a drop in imaging contrast and resolution can occur [43]. The axial resolution of LSIM is consequently limited to about $10\,\mu\mathrm{m}$ [id.].

2.2.3 Comparison of select whole-brain light microscopy techniques

In table 2.1, the attributes of fMOST, KESM, CLARITY LSIM, MOST, and STPT are compared. Specifically, the resolution at which a whole-brain is acquired using that modality, the imaging time necessary for that acquisition, and whether the technique supports fluorescence imaging.

Table 2.1: Comparison of select LM modalities on the acquisition of a whole-mouse brain.

Imaging modality	Lateral Resolution (x, y)	Axial Resolution (z)	Imaging Time	Fluorescence
fMOST [53]	$\sim$1 µm	$\sim$1 µm	$\sim$19-days	✓
KESM [54]	0.3$\sim$0.6 µm	$\sim$1 µm	$\sim$4-days	
CLARITY LSIM [55, 4]	$\sim$0.65 µm	$\sim$2 µm	$\sim$6-hours	✓
MOST [40]	$\sim$0.33 µm	$\sim$1 µm	$\sim$10-days	
STPT [41]	0.5$\sim$1 µm	$\sim$2.5 µm	$\sim$7-days	✓

STPT is currently the most broadly used approach to whole-brain imaging with LM. One reason why STPT is an attractive method is that the tissue removed during sectioning can be retrieved for subsequent analysis [1]. For instance, immunostaining techniques can be applied to the collected tissue. Such follow-up studies are not possible with AOH or TPTC [id.].

Another reason why STPT is currently popular is that tissue preparation for image acquisition has minimal effects on fluorescence and cerebral morphology [1]. KESM-based methods initially did not support fluorescence microscopy and did not do so at a time when STPT did. This made STPT a more attractive choice when whole-brain imaging was getting underway. The reason why KESM-based methods did not initially implement fluorescence imaging was due to how tissue was prepared for imaging: dehydration and plastic embedding was used in tissue preparation, which decreases GFP intensity [41]. Recently, fMOST has developed from MOST (which was based on KESM) imaging technology, which uses a resin-embedding method that maintains fluorescence.

LSFM coupled with chemical clearing for the imaging of whole-brain started slow. Initially, chemical clearing procedures quenched the fluorescence signals emitted by the tissue being im-

aged. The advent of CLARITY [48] has helped resolve this problem. As the interest in whole-brain imaging was growing, the spatial resolution of LSFM was limited by the objectives available at the time [1]. LSFM of the whole-mouse brain at submicrometer resolution has only been made possible recently with the development of objectives with large working distances and low numerical aperture. This coupled with other minor modifications to LSFM has dramatically increased its utility in research across the whole brain [id.].

STPT has been widely adopted as the imaging modality for many neuroanatomical investigations using fluorescent protein-based tracers [1]. This is more-or-less attributed to what STPT had to offer over its competitors as whole brain studies commenced. Large-Scale, whole-brain anatomical projects were pioneered by the Allen Mouse Brain Atlas of Gene Expression; the Mouse Connectivity project at the Allen Institute selected STPT to image a goal of $2,000$ brains. This has dramatically influenced the popularity of STPT within whole-brain studies.

2.3 Extracting vascular structures embedded in imaging volumes

2.3.1 Preprocessing

Every tissue sample and imaging technique has unique characteristics that affect optical penetration and light scattering during acquisition. During acquisition, signal heterogeneity frequently consequently presents with imaging distortions and illumination gradients that subsequently effect analysis [56]. Preprocessing is typical applied to acquired images in order to resolve common heterogeneities.

2.3.1.1 Downsampling

Due to various processing constraints, it is not uncommon for volumetric datasets to be downsampled, especially when processing whole-brain datasets [57, 58, 55, 41]. This process is trivial along the longitudinal and lateral axis: if the downsampling ratio r is an integer i, each new downsampled value is the average of i values from the original image. For the axial direction, image averaging can also be employed [6, 57, 59]. To do so, i images are averaged together to achieve a downsampling ratio of r. When r is not an integer value, the procedure is slightly more com-

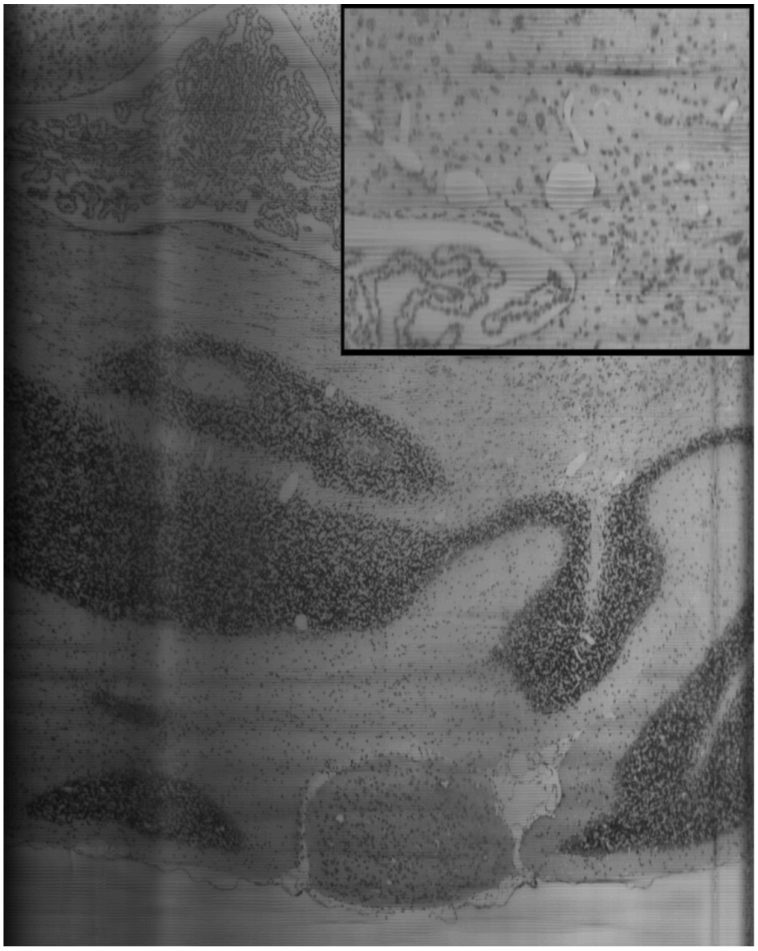


Figure 2.8: KESM acquired Nissl stained tissue images (raw) stitched together for viewing to illustrate lighting and knife irregularities. Application of light normalization and noise reduction are shown for the black-boxed region in Figure 2.9. Reprinted, with permission, from [3].

plicated as linear interpolation is necessary [60]. To return to the original resolution, interpolation during the upsampling process is required; downsampling a dataset and then upsampling it is *not* a lossless operation.

2.3.1.2 Light normalization

For 3D imaging stacks acquired by LM modalities, preprocessing typically begins with XY normalization and Z normalization [56]. The goal of this process is to help resolve uneven sample illumination and axial intensity dropout. For KESM-based imaging modalities (KESM, MOST, fMOST), lighting artifacts due to simultaneous imaging and physical sectioning are typically ad-

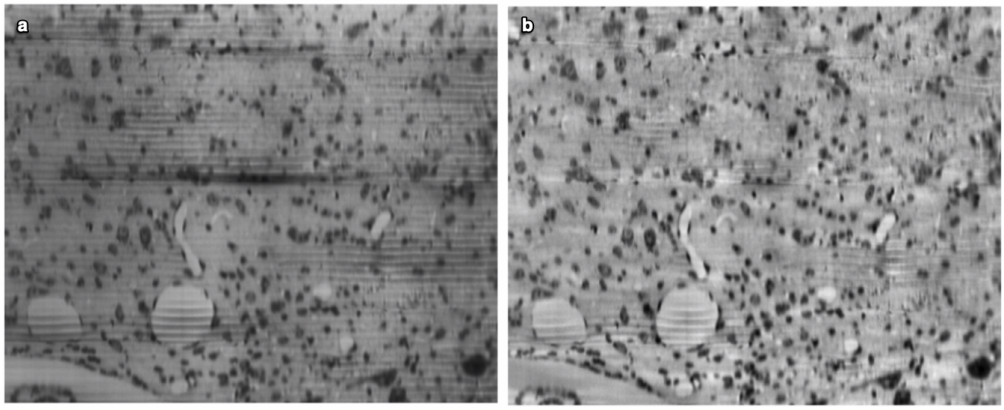


Figure 2.9: KESM Preprocessing of Figure 2.8. (a) Removal of lighting and knife irregularities by light normalization procedures; (b) removal of knife chatter. Reprinted, with permission, from [3].

dressed here [5, 3]. An example of the KESM process is shown applied to Figure 2.8 in Figure 2.9.a.

2.3.1.3 Noise reduction

In LM datasets, the number of pixels comprising the background far greater in number than that of the filaments. In order to reduce noise, the background of each image in the stack can be modeled using the Gaussian smoothing technique described in [61]. Once the background has been modeled, it can be subtracted from the original image using a procedure from [62]. While this technique can work well in practice, it can also be problematic when processing images with filament structures embedded within them. The reason for this is that local smoothing operations remove high frequency noise; the filaments are high frequency and low contrast [3]. To address the issues with local smoothing procedures, [3] scaled each value in the sample by a sliding window mean value of a small neighborhood about it. Using the averaging filter, high-frequency details were preserved while artifacts caused by knife chatter were removed, including those shown in Figure 2.6 as detailed in Figure 2.9.b.

Median filters provide considerably less blurring than a linear smoothing filter (such as the averaging filter) of the same size [60]. They are thus more commonly applied in LM studies. Median filters operate by moving a sliding window across the image, assigning the median value

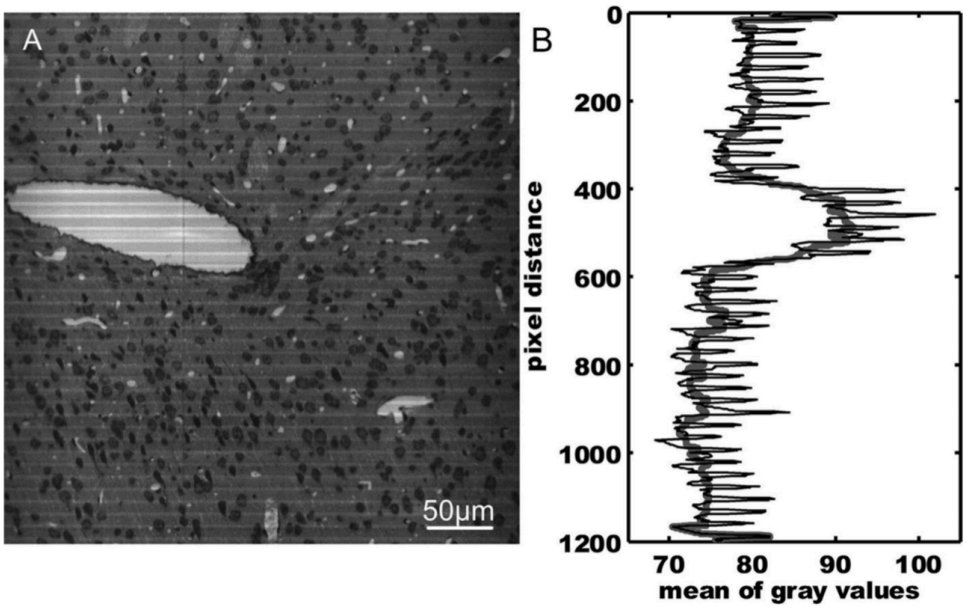
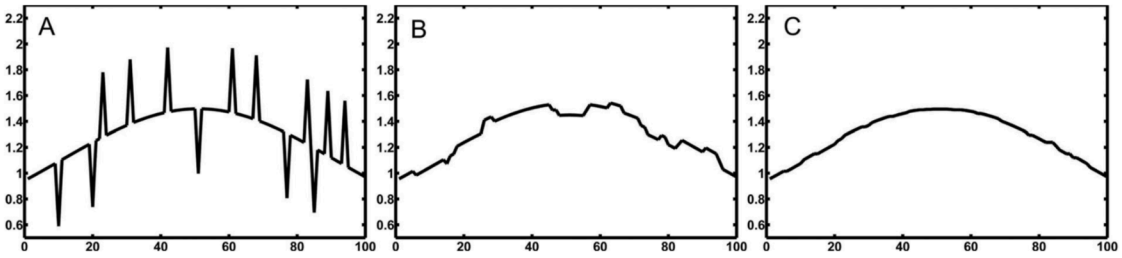


Figure 2.10: Streak removal procedure for MOST acquired images. (a) Picture of coronal section with visible streak noise; (b) mean projection along the horizontal plane of (a) shown by the thinner curve; thicker curve shows the result of the application of smoothing to the thinner curve. Reprinted, with permission, from [5].



captionoffigureSmoothing experiment to determine better method for streak removal in MOST images. (a) Original curve for a section of Figure 2.10; (b) result of using moving average filter with a neighborhood of 10; (c) result using a moving median filter with a neighborhood of 10. Reprinted, with permission, from [5].

of that window to the pixel located at its center [id.]. [5] illustrated that the median filter produces better results than the averaging filter in removing noise. First, [5] calculated a mean projection for an image $I(x, y)$ with streak artifacts presenting across its x-axis (Figure 2.10). A curve $m(y)$ for the y-axis was then calculated and was composed of the mean values of each row. Local extrema of the resultant mean-value curve were considered the streaks and the moving average filter and

median filter were applied to see which smoothed the curve best (eliminated local extrema). The median filter produces a smoother curve (Figure **??**); for MOST acquired images, the median filter removes streak artifacts better than the averaging filter.

When small filaments present, however, the median filter can be problematic. The reason for this is that the median filter forces points with distinct values to be more like their neighbors. Substantial data loss can result if the size of the filter size is selected incorrectly. Recall that the number of pixels in the background far dominates the number of pixels in the filament. Given a 5×5 median filter, isolated clusters of filament pixels whose area is less than $5^2/2$ will be eliminated from the image (will receive value of background).

To illustrate the potential implications of using the median filter, assume we have a dataset at a resolution of $1.0\,\mu\text{m} \times 1.0\,\mu\text{m} \times 1.0\,\mu\text{m}$. Furthermore, that we have applied a median filter of 5×5 to every image in that dataset. The median radius of the mouse capillary is near $2.0\,\mu\text{m}$ [29]. This means that capillaries diving axially into the volume will have an area of $12.75\,\mu\text{m}^2$ (on average). The median filter will thus 'eliminate' approximately half of all the capillaries in the dataset.

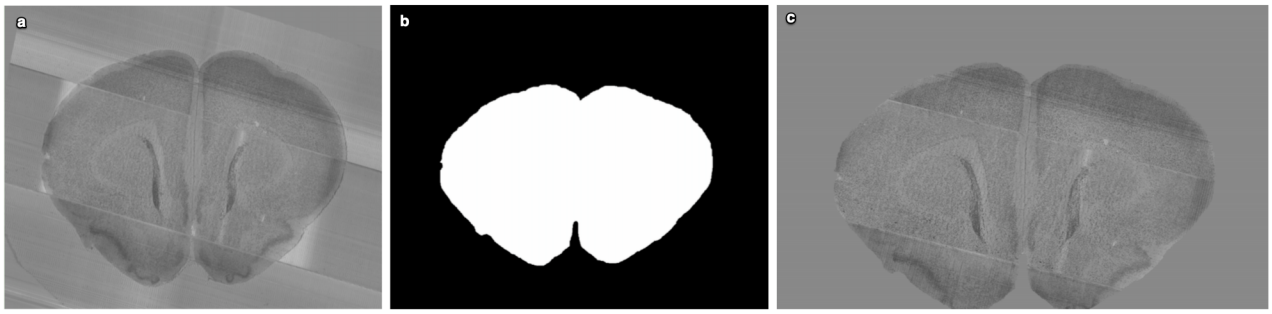


Figure 2.11: Application of freehand brain contour mask to KESM acquired image. (a) Whole-brain image slice with background noise present; (b) freehand brain contour mask generated for (a); (c) application of (b) to (a) to remove noise outside of the brain. Reprinted, with permission, from [6].

In many whole-brain studies, brain contour masks are first applied to imaging data to remove the noise presenting outside the periphery of the brain. These masks can be manually constructed

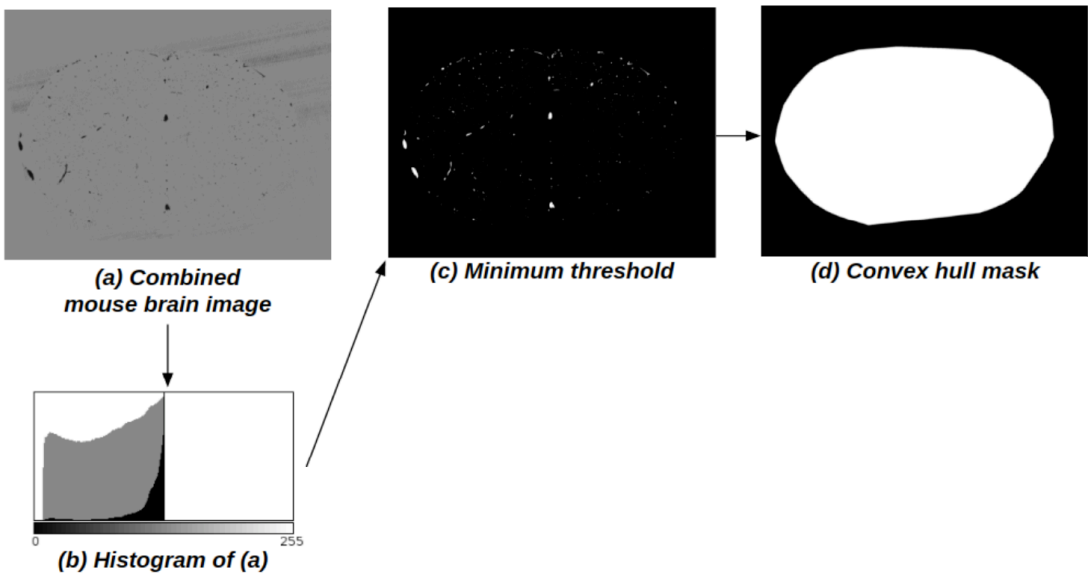


Figure 2.12: Fully automated brain contour mask generation by [6] to remove peripheral noise in KESM-acquired images. (a) Adjacent images are combined; (b) the histogram of the combined images is calculated; (c) minimum threshold is applied to the combined image with a value based on the histogram; (c) convex hull is calculated for the points remaining after threshold application. Reprinted, with permission, from [6].

by freehand using an illustrator program as detailed in [6, 63]. An example of a brain contour mask is shown in Figure 2.11 along with its application to eliminate noise outside of the brain. As described by [6], the freehand generation of brain contour masks is a time consuming. A fully automated method based on thresholding and the convex hull algorithm was designed by [6] and is detailed in Figure 2.12. Another fully-automated method of brain contour mask generation has been developed by and described in [5].

After removing noise outside the brain contour, there are many different kinds of filters that can be applied to imaging data [60] to help reduce the noise that presents within it. Popular choices include both the median filter [5, 6, 58, 4] and averaging filtering [3] as defined earlier in this section. As pointed out, it is imperative that the effect of the filter on filament structure is studied before its application as data loss is likely to otherwise occur. Simple investigations with filament models can help determine the extent to which the filaments are degraded the application of the

filter and thus can also help quantify how much data loss can be expected.

2.3.2 Segmentation algorithms

Segmentation algorithms play a crucial role in the processing of raw LM imaging data of the filament networks. LM acquired datasets, whether of the whole-brain or a local tissue sample, often contain imaging artifacts and noise. Furthermore, some filaments in the network typically drop below the sampling resolution of the objective used during acquisition. This means using a threshold value to simply differentiate between filament network and surrounding tissue is not adequate [64]. Instead, a more involved method is necessary to extract the filaments embedded in the imaging volume.

The algorithmic approaches used to extract vasculature vary depending on acquisition modality, domain, and method. [65] provides an extensive overview of the various methods commonly used to extract vasculature structures. After a detailed, yet widespread, survey of the literature, [65] proposed six categories of approaches: (1) pattern recognition techniques, (2) model-based approaches, (3) tracking-based approaches, (4) artificial intelligence-based approaches, (5) synthetic neural network-based approaches, and (6) tube-like object detection approaches. Many of the approaches reviewed were developed for medical imaging applications such as CT and MRI image segmentation and therefore focus on the extraction of large blood vessels [64].

[64] briefly reviewed the application of different classes algorithms to LM imaging. [64] details how multi-scale techniques (such as [66, 67]) are helpful for feature detection; however, this technique is inadequate for thin filaments that require a high-level of detail to resolve. As for centerline detection [68, 69] and thinning [70], [64] reports that the requirement to find an optimal threshold value in high-throughput LM data is difficult due to noise unique to LM. As a consequence, topological errors in filament centerlines are prevalent as is over segmentation. Region growing techniques [71, 72] are described by [64] as being difficult to apply due to their reliance on an initial surface, which is difficult to find given the complex structure of filament networks. Template matching approaches [73], which are robust to noise, are difficult due to their requirement of a large number of templates, oriented in different directions, and across multiple scales

[64]. Meanwhile, vector tracking algorithms [74] are described as operating well on well-defined surfaces of continuous structures, but quickly degrade when imaging contrast is low or when the border is not crisp [64].

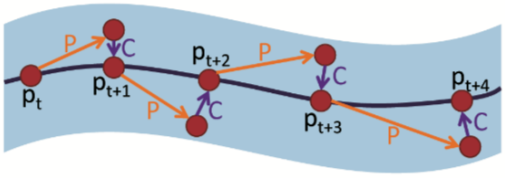


Figure 2.13: Mayerich et al.'s predictor-corrector filament tracking method. Initial position p_t is predicted along the particle path P. That prediction is then corrected to lie on the filament axis. Reprinted, with permission, from [7].

To address these concerns, [7, 75, 64] developed a filament tracking method that incorporated template matching. The algorithm operates from a seed point, where an estimation of vessel trajectory is made, and tracker is moved forward to that point. A Gaussian template is then used to model the mostly circular filament cross section. The estimation of the tracker position is then corrected based on the fit of the model. This is the foundation for Mayerich et al.'s 'predictor-corrector algorithm' (Figure 2.13).

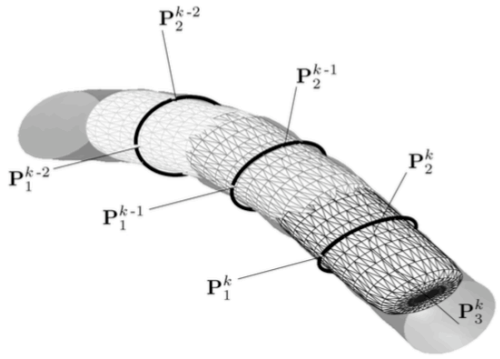


Figure 2.14: Tyrrell's approach to filament tracking. Sequence of k cylindroidal superellipsoid fits denoted as $k-2$, $k-1$, and k superimposed on filament. Reprinted, with permission, from [8].

Another method that uses local modeling is [8, 76], with the model used a superellipsoid primitive. Segmentation commences with a seed point, upon which a superellipsoid is fit. This model is then shifted across the vessel, with the parameters of the superellipsoid re-estimated each step along the way. This method can be summarized in short as an 'estimation and detection approach' (Figure 2.14).

Relatively new methods involving deep learning (convolutional neural networks) have been developed to extract filaments from 3D images [77, 78]. Like the aforementioned methods, these have neither been applied to whole-brain datasets in full; however, they have shown good correspondence with human-labeled ground truths on small samples of data. At the time of writing, a deep learning filament segmentation method has yet to be applied across a complete whole-brain dataset.

The techniques described to this point can all be applied directly to raw imaging data. Mayerich et al.'s predictor-corrector filament tracking method and Tyrrell's superellipsoid estimation and detection approach both operate well in noisy environments. This is in part because a filament-like template or shape is fit on an actual filament, a process which then proceeds out from the original seed point and along the filament axis. These methods are more sophisticated than what tends to be used in practice.

In our research, we have found that it is common practice to threshold images acquired from an LM sample, perform some morphological operations, and then apply an algorithm to the resultant binary raster images to construct a vectorized representation. We have already reported some of the problems with a thresholding approach, but nonetheless thresholding has been applied across whole brain datasets and small regional samples to segment the embedded filaments. Otsu's algorithm [60] appears to be the most commonly applied thresholding technique [57, 58, 4, 79, 63, 80, 81]. However, other thresholding algorithms including bilevel thresholding using local threshold calculated via max-entropy thresholding method [59], minimum thresholding [6], and many others are too used in practice.

The binarization process via thresholding is usually followed-up with morphological opera-

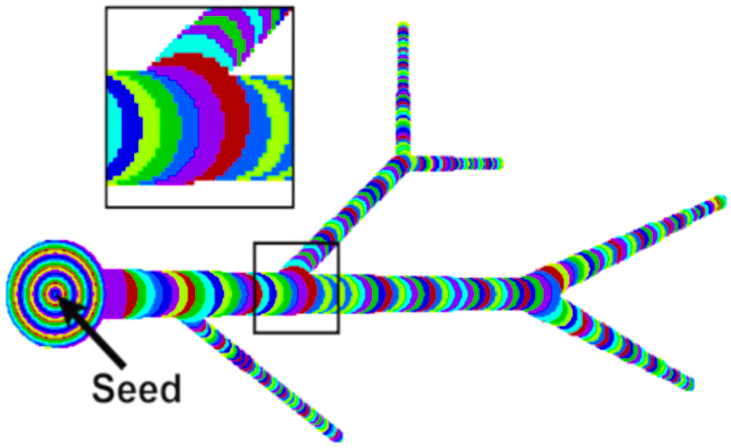


Figure 2.15: Clusters of voxels (shown in different colors) 'scooped' by the voxel scooping method as it extended outward from the seed point. Reprinted, with permission, from [9].

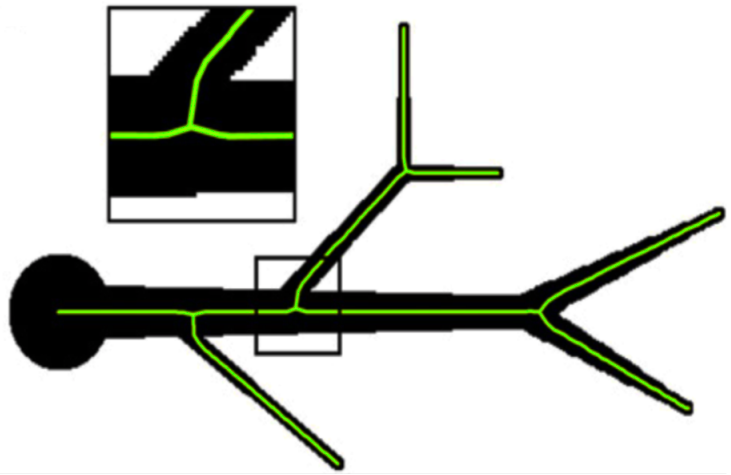


Figure 2.16: Centerline calculated through clusters of voxels 'scooped' by the voxel scooping method in Figure 2.15. Reprinted, with permission, from [9].

tions, such as closing and hole filling, to address structural integrity issues from the thresholding process. Thereafter, a vectorized representation of the filament network usually follows, upon which quantitative analysis will be performed. In [6]'s whole-brain study, the voxel scooping algorithm was used [9], although there are many different ways to perform vectorization. The voxel scooping method begins from a seed point and grows outward from that seed point by scooping

voxels into clusters, and continues this process until there is no unvisited region within the filament. This process results in a number of voxel clusters (Figure 2.15). The centroids of the scooped clusters are then connected to obtain the centerline of the vessel (Figure 2.16). The filament network isolated using the voxel scooping method is then written to an *swc* file: each line segment of the centerline is written along with an estimate of the radius about that line.

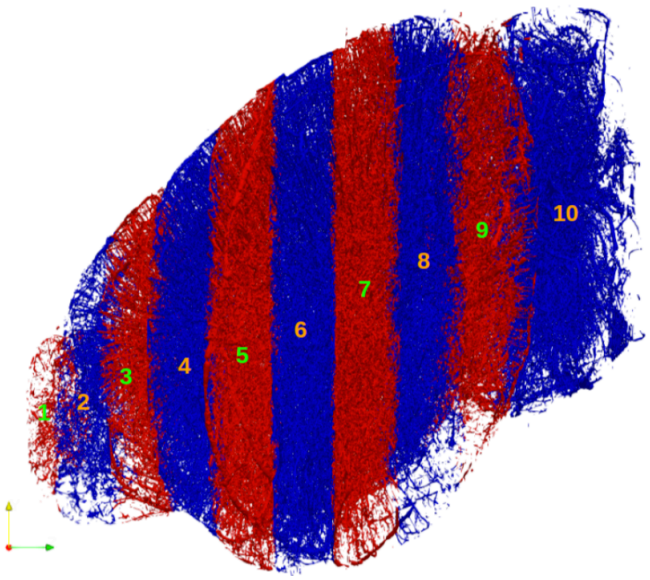


Figure 2.17: The whole-brain coronal images of the KESM India Ink dataset split into ten vertical sections for independent processing as completed by [6]. Reprinted, with permission, from [6].

When segmenting large datasets, such as that embedding the whole-brain, it can become necessary to split the imaging data into sections that can be processed independently by a segmentation algorithm. [6] adopted this approach when segmenting the vasculature from the KESM India Ink whole-brain dataset [39]: whole-brain coronal images were split into ten vertical sections (Figure 2.17); the filaments were extracted independently from these vertical sections using the voxel scooping method [9]. Independent processing leaves filament discontinuities between adjacent vertical sections (Figure 2.18) that must be stitched together before a quantitative analysis can commence.

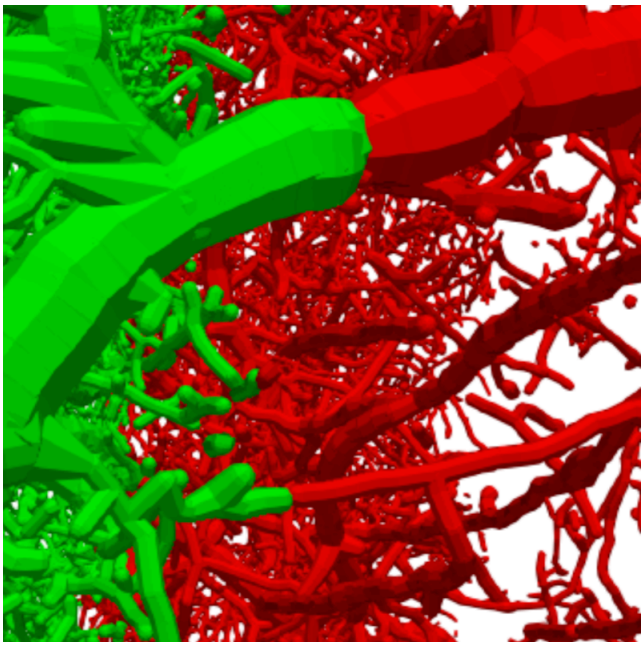


Figure 2.18: Example of the discontinuities that present between adjacent vertical sections (Figure 2.17) after being processed independently. Reprinted, with permission, from [6].

2.3.3 Segmentation algorithm validation

The validation of filament segmentation and tracing procedures, at the time of writing, is predominantly carried out using traditional validation volumes. This primarily includes digital phantoms and expert-labeled volumes [82, 12, 79, 77, 78, 55].

2.3.3.1 Digital phantoms

Digital phantoms are synthetic data volumes, each composed of a synthetic filament that is representative of a particular structural complexity potentially presenting in a 'real' biological dataset (Figure 2.19). The quality of these imaging volumes typically ranges from noise-free to noise levels representative of the target imaging modality to degradations that represent worst-case scenarios. The known characteristics and location of the 'phantom' embedded in the volume serve as its ground truth. Algorithmic performance is then evaluated on particular structural complexities perceived in 'real' cerebrovasculature filaments under controlled conditions [82]; accuracy is frequently assessed as the correspondence between the resultant segmentation and the 'phantom'

Figure 2.19: Characteristic example of the digital phantoms used in the validation of a filament segmentation algorithm. These three were used in the validation of [8]'s superellipsoid estimation and detection tracing algorithm. Reprinted, with permission, from [8].

ground truth [11].

2.3.3.2 Expert-labeled volumes

Subsets from the biological dataset are frequently sampled at random and subsequently segmented manually (labeled) by a human expert to devise validation sets that better reflect the structural complexities inherent to the structure of the cerebrovasculature system. The manual segmentation constitutes the volume's ground truth; it is composed of hand-drawn annotations by the expert, across every two-dimensional image-slice of the volume.

When expert-labeled volumes are constructed for filaments, the three-dimensional volume embedding those filaments is viewed through two-dimension slices. Due to this, manual ground-truth generated tends to be jagged in appearance, causing some small, thin filaments to become disconnected from their parent branches [79]. By its very nature, the process of generating labeled datasets is inherently time-consuming. It has been reported that an expert required 20-hours to label a $256 \times 256 \times 200\,\mathrm{pixels}$ dataset acquired by multiphoton microscopy that was resampled to a voxel size of $1\,\mu\mathrm{m}$ cube [78]. In addition to taking a significant amount of time to generate, the physical labeling process is inherently subjective and frequently inaccurate [79]. Furthermore, it is important to acknowledge that human-generated ground truths reflect only a small fraction of the structural complexity and variability that would be encountered in the complete volume [11].

2.4 Quantitative analysis of vascular structure

Given the significance of vessel morphology in pathophysiological conditions, many attributes of filament structure relating the large blood vessels have been characterized in medical imaging studies. In this section, we will summarize the common vascular element attributes, volume characteristics, and network characteristics reported in the literature. Metrics detailing these aspects of the vascular filaments presenting in the brain can be applied across the whole-brain; the limitations of preceeding imaging technology often focused quantitative analysis to small volumes of data at high resolution or large volumes of data at low resolution.

2.4.1 Vascular element attributes

There are several attributes frequently reported in the literature to describe the individual elements of the cerebrovasculature system. Diameter and length are perhaps the most regularly reported, with the reason being that, according to Poiseuilles formula, the flow rate (volume/time) through a respective vessel element is directly proportional to the product of the pressure difference across it by the fourth power of its diameter and is inversely proportional to its length. The fact that length is directly proportional to resistance implies that the longer a vessel is, the higher its resistance to flow will be; moreover, long vessels under lumen pressure can become unstable when the pressure exceeds some critical value, resulting in 'vessel bulking thus increased vessel tortuosity [83].

Vessel abnormalities characterized by irregular tortuosity affect blood flow by increasing resistance and decreasing perfusion [84]. As our derivation of Poiseuilles law in equation 2.4 applies to cylindrical shaped vessels, changes in blood flow due to tortuosity will not be reflected in the presented formulation; however, an adaptation of Poiseuilles law taking in to account the structural tortuosity of vessels appears in [85], and numerical models that have confirmed that tortuous vessels exhibit different hemodynamics from non-tortuous vessels [i.d.]. The disturbances that tortuosity has on physiological hemodynamics has made it a candidate biomarker for disease pathology

[84]. For instance, arteriolar tortuosity in human cerebral white matter confirmed morphological changes associated with vascular dementia [84, 86]. It is for these reasons, and amongst others (e.x., [87, 88, 89, 90]), that vessel tortuosity is also an attribute commonly reported in studies of cerebrovasculature morphometry [91].

The local geometrical properties of bifurcations are another aspect of the angioarchitecture that can affect blood flow. Recall that each vessel element has a specific resistance to flow that is inversely proportionate to its diameter and positively correlates to its length; that the resistance of the total network is dependent on the connectivity of these elements to one another [15]. Furthermore, that blood flow in the capillaries is dominated by viscous dissipative forces and that the apparent viscosity of blood is a function of hematocrit; that the distribution of hematocrit across vessels sourcing the capillaries is heterogeneous due to plasma skimming, red cell screening, and phase separation that occurs at bifurcations [91]. Given the influence of branching patterns on resistance and peripheral flow distribution, their angles of bifurcation and area ratios are frequently described as a function of their nature (arterioles versus venules), taxonomy (based on parent position in centripetal ordering scheme), or type [17].

There are three bifurcation types typically described: intra-element bifurcations (IEB) in which the three branches have the same order; homogeneous-symmetrical bifurcations (HSB) in which the order of the two child branches equals that of the parent minus one; and asymmetric-lateral bifurcations (ALB) in which a lower order vessel branches laterally between two segments of the same order [17]. The bifurcation angles are calculated using the boundaries of the intersecting vessels at a node n: for each segment i a vertex is identified inside, and another outside, a sphere centered at the node of an intersection, with the maximum radius of the three vessels. The mean tangential unit vectors as calculated at each of the identified vertices. This results in three-unit vectors u_0, u_1, and u_3 for the mother, major child, and minor child respectively. The branching angles are calculated as $\alpha_{ij} = \cos^{-1}(u_i \cdot u_j)$ and the asymmetry angle was defined as $\beta = \alpha_{01} - \alpha_{02}$.

The three branches of a bifurcation define a solid angle $\Omega = \alpha_{01} + \alpha_{02} + \alpha_{12} - \pi$; when $\Omega < \pi$, the three branches are co-planar; $\pi - \Omega$ is the out-of-plane angle, which can be used to define bifurcation planarity [i.d]. The total area ratio of the bifurcation is calculated as $((A_1 + A_2/A_0))$, the major area ratio as A_1/A_0, the minor area ratio as A_2/A_0, asymmetry ratio $\gamma = A_2/A_1$, where A_0, A_1, and A_2 are the cross-sectional areas of the parent, major child, and minor child branches respectively [17].

2.4.2 Vascular volume characteristics

A few additional measurements are commonly reported across cerebrovasculature datasets to quantify the properties of the filaments embedded within them together. The total vascular volume mm^3 can be estimated by iterating over each of the vascular elements, calculating its capacity, and adding the result to a running total. Similarly, the total vessel surface area mm^2 can be approximated by iterating over each of the vascular elements, calculating its surface area from cylindrical approximation, and adding the result to a running total. Subsequently, the surface area to volume ratio can be calculated by dividing the total vessel surface area by total vessel volume; the amount of surface the microvasculature has relative to its size directly affects its ability to participate in the transcapillary exchange.

Vascular volume density values are regularly conveyed in the literature as well but done so in some different ways [24, 91, 2, 92, 10]. In two-dimensional morphometric studies of histological sections, the universally utilized parameter is total vascular length mm by mm^3 [24]. Although this method is rarely used to report vascular density in three-dimensional imaging studies, it is employed to compare results to those published in histological studies. Mean vascular length mm by mm^3 is a more generally used quantification of vascular density, albeit with a high degree of variability due to the inherent inhomogeneity of cerebrovasculature within the brain [id.]. Vascular surface area mm^2 by mm^3 provides insight into the exchange surface, while vascular volume mm^3 by mm^3 gives an estimate of blood volume. Such measures of density can be used to quantify the regional differences in microvascular profile across the brain [2, 24].

2.4.3 Vascular network architecture

The vascular system of the brain is a highly interconnected network that rapidly redistributes oxygenated and nutrient-rich blood to regions in need by matching local blood flow to metabolic demand. Furthermore, the cost of angiogenesis, remodeling [20], and structural adaptation [21] is offset by an increased power that it affords in routing blood towards the neurons and cellular constituents of the brain in need [28]. The ability to redistribute flow and capability to add additional plumbing helps ensure that if the system fails, it does so gracefully [id.].

The appropriate mathematical representation of a highly interconnected network – from the Manhattan-like roadway systems to the social networks of the Internet – is the graph [93]. The brain's vasculature system is a highly interconnected system, and can thus be represented by a set of nodes (i.e., bifurcations) and the connectivity between those nodes (i.e., vessel segments) [91].

Many studies of the topology of the cerebrovasculature, as it presents in the territory of the middle cerebral artery (MCA), have leveraged the graph representation of vessel segments in their analyses. Investigations of the MCA backbone commenced with a graph representation of the entire vascular territory sourced from it; the backbone was isolated through the removal of all vertices with a coordination number of 1, followed by the iterative removal of all newly formed vertices with coordination numbers of 1 or 2 until only the backbone remained. The ratio of vertices to edges (used as a measure of network redundancy) was calculated as 3 to 2, implying that $3N/2$ edges within the backbone with N vertices, where $N/2$ of the edges represented anastomoses [28]. The collection that comprised the smallest independent loops (analogous to Kirchhoff's loops, circuit theory) was then determined; based on the distribution of edges in the circles comprising the backbone, a mode of four sides was observed. The observation of four edges per loop, together with a coordination number of three for each vertex, suggests that no regular lattice describes the backbone [id.].

The interconnected nature of the microvasculature was determined by assigning a fluid resistance to each edge and viewing the microvasculature as a three-dimensional network [29]. This resistance was calculated using the Hagen-Poiseuille law, and the total network resistance across pairs of vertices subsequently determined. It was found that the resistance asymptoted as a function of Euclidean separation distance, implying that the microvasculature formed a highly interconnected irregular lattice [id.].

The formulation of communities of vertices whose inter-connections were relatively more frequent than by chance and intra-connections at a level less than chance were derived [29]. The strength of the communities was evaluated by plotting the number of connections between communities versus associations in the community: a strong community would have interconnections that scale as the surface-to-volume ratio with a power law exponent of $2/3$. Based on an observed power-law relationship with an exponent of 0.83, it was concluded that vertices in a given spatial region make extensive connections with neighboring communities [id.].

The flow domain of penetrating vessels was calculated using Kirchhoff's law for current conservation at each vertex with the values of the resistance for each edge, and a constant pressure difference between open ends of the vessels. This flow calculation yield the vessels predominantly sourced from a given penetrating vessel [29].

2.5 Whole-brain dataset dissemination schemes

Many initiatives aim to make these whole-brain datasets publicly available; frequently, they are disseminated through web-based atlases. A web-based atlas is essentially a light-weight data browser that serves-up the raw data in raster graphics format [94]. Such an atlas provides for a straightforward qualitative assessment of microstructures, which can be conducted in a three-dimensional viewing environment that is produced by image overlays with distance attenuation and shearing to generate stereo pairs of the encapsulated data [95, 94]. While such atlases allow whole-brain datasets to be distributed across the Internet, their raster-nature inhibits fast quanti-

tative evaluation of the embedded microstructures due to the lack of readily available geometric data from which morphometric statistics can be derived. To quantitatively analyze the data the microstructures must first be extracted (i.e., segmented) and then geometrically reconstructed. The boundary towards such an analysis can be lowered by storing the data in vector graphics format; however, the data must nonetheless be converted into graph representation or stored within an organized database, to study filament connectivity and the implications thereof [96]. We will highlight some notable atlases before detailing how graph databases can be leveraged to store and disseminate connectivity data acquired from whole-brain imaging studies.

2.5.1 Web-based atlases

2.5.1.1 Knife-Edge Scanning Microscopy Brain Atlas (KESMBA)

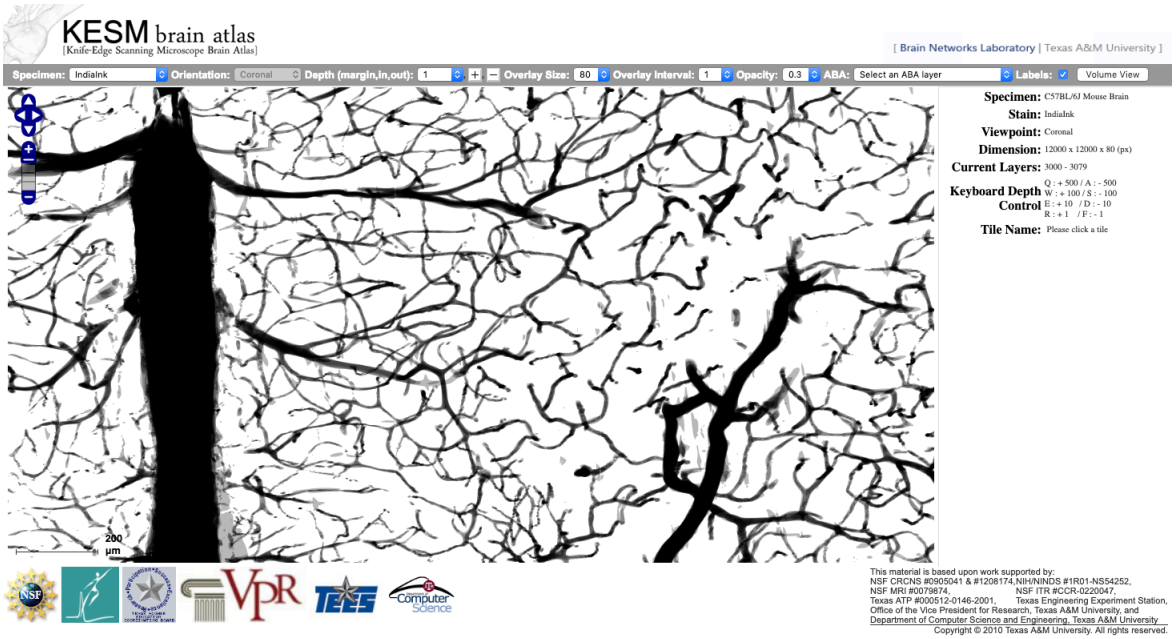


Figure 2.20: An example of the web-based presentation of the KESM Mouse cerebrovascular dataset in the KESMBA. Shown is an overview of 80 contiguous slices of coronal vasculature data at the highest available resolution. kesmba.cs.tamu.edu

The KESMBA, a collection of web-based atlases for the C57BL/6J mouse model, includes

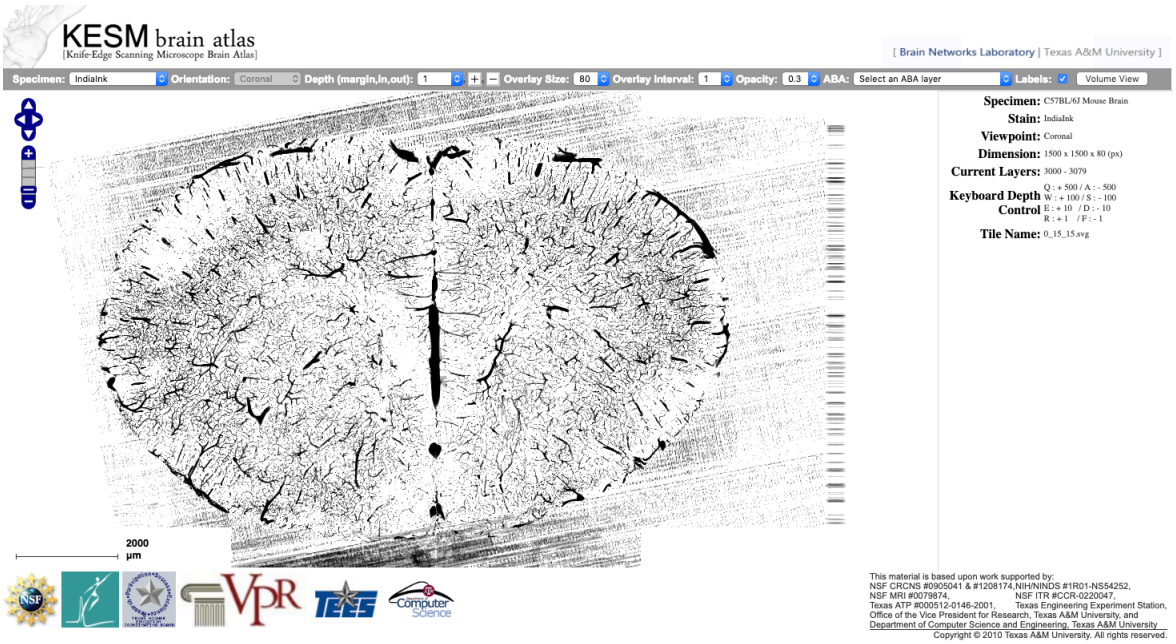


Figure 2.21: An example of the web-based presentation of the KESM Mouse cerebrovascular dataset in the KESMBA. Shown is an overview of 80 contiguous slices of coronal vasculature data at the lowest available resolution. kesmba.cs.tamu.edu

whole-brain Golgi (neuronal circuits), Nissl (soma distribution), and India ink (vascular networks) datasets [94, 39]. All datasets were acquired using Knife-Edge Scanning Microscopy, and are presented in the web-based atlas across multiple scales, at a maximum resolution of $0.6 \times 0.7 \times 1.0$ μm^3. These atlases were initially implemented using tiled images in raster graphics format in conjunction with the Google Maps JavaScript API. This version [94] has since been updated to use the OpenLayers API along with tiled images in vector graphics format. This change from raster to vector graphics format has decreased atlas load times. It has also introduced point-and-click reconstruction of vascular geometries and provided a means to calculate the corresponding morphometry data (such as diameter, segment/element length, surface area, volume, and the number of branches). These reconstructed geometries and relevant morphometry data are interacted with through a WebGL-based volume viewer. An example of the web-based atlas of the India Ink dataset is shown in Figures 2.20 and 2.21.

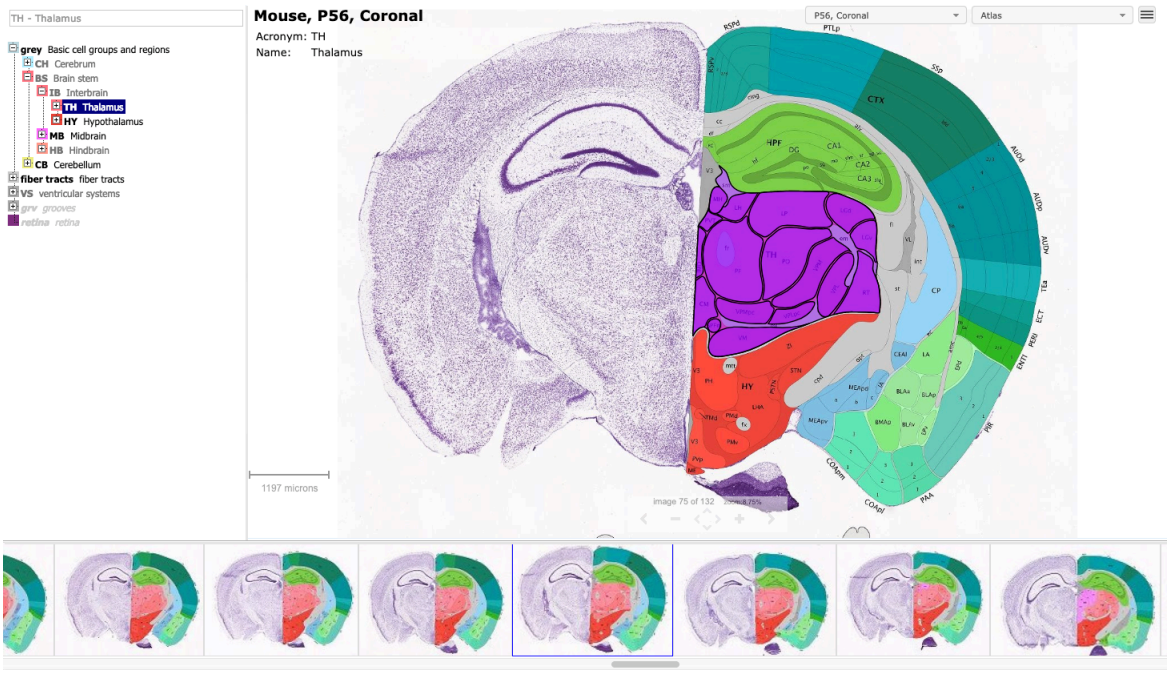


Figure 2.22: An example of the web-based presentation of the ALlen Brain Reference Atlas for the mouse. Shown is a coronal section of Thalamus brain tissue on the left, with a reference for cerebral structure overlaid on the right. atlas.brain-map.org

2.5.1.2 The Allen Brain Atlas

The Allen Brain Atlas also provides a collection of atlases that are publicly available to researchers through the Internet. The Allen Mouse Brain Connectivity Atlas [97] is a mesoscale whole-brain axonal projection atlas of the C57BL/6J mouse brain acquired using serial two-photon microscopy with automated vibratome sectioning. Through the online interface, users can interactively browse through all specimens contributing to this dataset (1772 at time of publication of [97]), or search for specific datasets by injection site, Cre-driver lines and/or target locations; axonal projection can then be investigated in the primary image data and through a three-dimensional summary. Multiple specimens can also be selected for co-viewing in numerous image-series viewers or in Composite View mode [id.]. Overall, this atlas provides the foundation for large-scale analysis of global neural networks, as well as between different local networks, and between different neural systems [id.].

The Allen Brain Atlas data portal also contains some additional public resources. This includes the Allen Mouse Brain Atlas (Figure 2.22, which provides anatomic coverage across the entire brain in conjunction with spatially-mapped, genome-wide coverage [98]. Also, the Allen Developing Mouse Brain Atlas, which contains projection mapping image data detailing axonal projections along with the imaging data derived from serial two-photon tomography [id.].

2.5.2 Atlas-based analysis

The whole-brain datasets currently published in the web-based atlas format provide a wealth of imaging data that can be qualitatively analyzed but are limited in the quantitative informatics that can be applied to them. The second generation of the KESMBA allows simple reconstruction along with necessary morphometry data for vasculature data. Meanwhile, the Allen Brain Atlas presents a wealth of data for neuronal connectivity, along with a precise means of querying that data for different neuronal motifs. Given that the nature of an angiome, i.e., distribution of vascular elements, can be considered a connectome of sorts, a more specific means analysis of the brain vasculature, outside of primary morphometry measures, is warranted.

It would make sense that different topologies could be analyzed with respect to resistance to flow, in addition to the examination and identification of the robust and fragile aspects of flow in different regions of the brain (for instance, in the area of MCA, subsurface and surface arterioles are robust, but the penetrating vessels connecting the two networks are fragile). Furthermore, that a similar querying system that can search and reconstruct vessel elements meeting some criterion (morphometry, taxonomy, type, etc.) within a specific scale (as organizational principles and morphometry may vary depending on scale) using a heuristic algorithm based on vessel characteristics is also reasonable.

2.5.3 Graph databases

The standard form of a graph model is the *labeled property graph*, which has the following characteristics: contains nodes and relationships; nodes contain properties; nodes include at least

one label; connections are named and directed; and, relationships contain properties [93]. Whole-brain imaging datasets embedding vascular connectivity can be converted into a labeled property graph storing filament topology and geometry [96, 99]. This representation can be entered into a graph database management system (i.e., graph database); such a system provides create, read, update, and delete (CRUD) methods that expose an underlying graph data model [id.]. In the graph data model, each node and edge is uniquely identified. A node can have characteristics, which could include radius, bifurcation type, branch angle, and pressure; meanwhile, an edge can also have properties, which could consist of length, resistance, and taxonomy.

The Neo4j analytics platform runs on top of the Neo4j graph database. In addition to implementing a platform to store graph data models, it provides a suite of dynamic graph analytic tools through an open library of high-performance algorithms. Included are algorithms for traversal, pathfinding, centrality, and community detection [100]. The traversal algorithms, such as breadth-first search (BFS) and parallel depth-first search (DFS), could be used to reconstruct particular aspects of the vascular network.

Pathfinding algorithms, such as single-source shortest path and all-pairs shortest path, would be useful in determining direct and alternating routing to different regions or in the analysis of network capacity [id.]. As for measures of centrality, the betweenness centrality algorithm could show utility in ascertaining the bottlenecks of blood flow into different regions of the brain [100]. Meanwhile, community detection algorithms such as the local clustering coefficient may prove helpful in determining a region's robustness to occlusion [id.].

3. BIOLOGICALLY GROUNDED SYNTHETIC CEREBROVASCULATURE MODELS

3.1 Motivation

The development of neuroimaging protocols for the acquisition of entire small animal organs at sub-micrometer resolution has helped integrate our knowledge about the cellular- and system-level aspects of the cerebrovasculature system. While the development of whole-brain imaging techniques has opened the door to data-driven, quantitative investigations into the cerebral angioarchitecture down to the level of brain capillaries, the accuracy of the segmentation algorithms used to extract the embedded filaments is customarily assessed using validation sets that are significantly smaller in sample size and dimensionality per grain [11]. Therefore, there is a critical need to scale the validation process to larger volumes of data, to better assess the scalability of vascular segmentation algorithms across larger-scale, higher-resolution whole-brain datasets.

To address this need, we must obtain a large number of models reflecting the morphological features and multiscale nature angioarchitecture of the cerebrovasculature system. The procurement of ground truths of such datasets has been a barrier to their incorporation to the validation process [id.]. Consequently, validation is frequently carried out using traditional validation volumes, such as digital phantoms and expert-labeled volumes [82, 12, 79, 77, 78], that are smaller-scale. The limitations of using digital phantoms and expert-labeled volumes have been detailed in Sections 2.3.3.1 and 2.3.3.2 respectively.

The limitations of digital phantoms and expert-labeled volumes have motivated the introduction of biologically-grounded, synthetic cerebrovasculature models. The primary objective is synthetic morphologies generated stochastically, according to the natural distributions of vessel diameters, lengths, and densities found in 'real' cerebrovasculature trees. In this chapter of my dissertation, I will illustrate how biological data describing the filament arborescence between the post-capillary

venules and the piemerian origin of the human collateral sulcus (Figure 3.1) [10] can be used to construct stochastically-generated, synthetic cerebrovasculature models.

3.2 Methodology

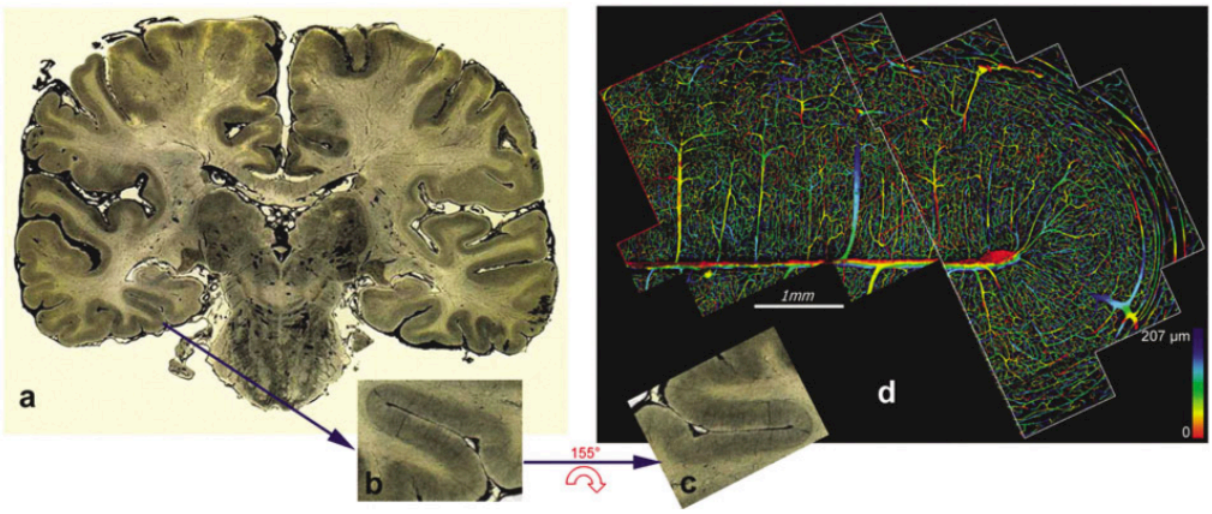


Figure 3.1: Slice of India ink-injected human brain. (a) Coronal slab of tissue containing the collateral sulcus in the temporal lobe; (b,c) close-up of the collateral sulcus; (d) reconstruction of confocal microscopy acquisition of the collateral sulcus. Reprinted, with permission, from [10].

3.2.1 Biological grounding

We produce synthetic models representative of the well-characterized cerebrovasculature of the collateral sulcus in the human temporal lobe (Figure 3.1). In this cerebral region, [10] carefully charted the branching pattern of cerebral vessels. They recognized that the filament arborescence in this sulcus could be modeled as a hierarchy, using the diameter-defined Strahler system (Figure 3.2) [23]. In this system, the Strahler number (i.e., the 'order') of each vessel segment is determined using a bottom-up process, beginning from the terminal vascular roots of order-0. As bifurcations are encountered, the confluent segment is given the order number of the larger 'child' element (assumed to be n), which is incremented by 1 (to order-$[n+1]$) if and only if its diameter is greater

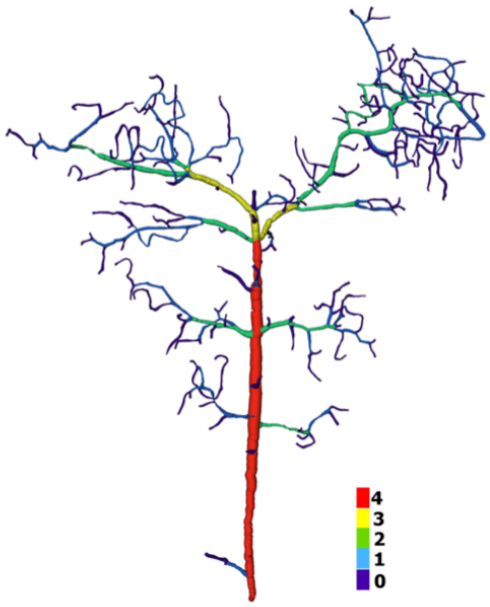


Figure 3.2: Example of the diameter-defined Strahler system used to label the branching pattern of a cortical vein in the collateral sulcus of the human brain. Reprinted, with permission, from [10].

than

$$[(D_{\mu_n} + D_{\sigma_n}) + (D_{\mu_n+1} + D_{\sigma_n+1})]/2, \tag{3.1}$$

where D_{μ_n} and D_{σ_n} denote the mean and standard deviation of the diameters of the order-n vessels respectively. [10] used this process to study characteristics (diameter, length, distribution) of the cerebrovascular filaments in this venule tree by hierarchical order. In our method [11, 12], these characteristics serve as the ground truth from which our synthetic models are derived.

3.2.2 Model-based generator

3.2.2.1 The 'backbone'

The 'backbone' of our synthetic models is an order-4 vessel (the 'root'), bifurcating into two order-3 vessels (the 'offshoots') (Figure3.3). The diameter and length of the resultant segment is assigned following a Gaussian distribution ($\mu = 0.0$; $\sigma = 0.5$) around D_{μ_n} and L_{μ_n} respectively, where n is the order of the branch created.

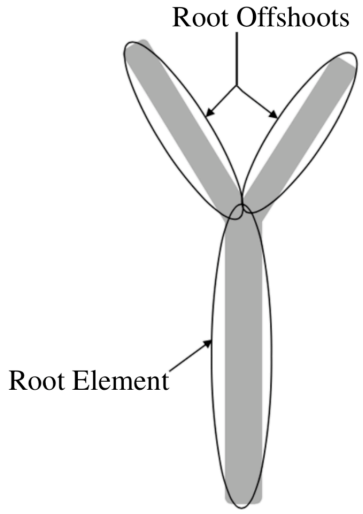


Figure 3.3: Illustrates the structure of the 'backbone' of the synthetic model. Reprinted, with permission, from [11, 12].

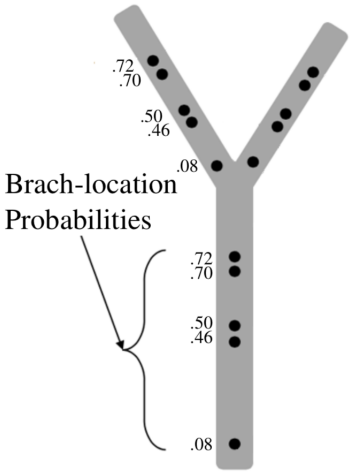


Figure 3.4: Branch location-probabilities on the 'backbone'. Reprinted, with permission, from [11, 12].

3.2.2.2 Branching from the backbone.

Branch points are subsequently assigned to the backbone's root and offshoots. The probability that a specific location on the backbone will receive a branch is dependent on branch-location probabilities (Figure 3.4) derived from the ground truth. If a branch point is assigned to a location on the root, an order-2 child results; from the offshoots, an order-2 or order-1 child is chosen by a random process. When a branch does arise, it diverges from the parent vessel at an angle ψ, such that $\frac{\pi}{4} \leq \psi \leq \frac{\pi}{3}$ is satisfied; ψ is then negated with probability $p = 0.5$. The diameter and length of the resultant segment are then assigned stochastically, as described in section 3.2.2.1.

3.2.2.3 Branching from non-backbone segments.

Each vessel segment can have branches, except order-0 elements. The spatial distribution of branch-location probabilities across the root vessel (Figure 3.4) is also used for the lower-order elements. If a bifurcation occurs, the child's vessel separates from its vessel in accordance the procedure detailed in section 3.2.2.2, with diameter and length assigned following a Gaussian distribution as described in section 3.2.2.1.

3.2.2.4 Vessel element generation.

The formation of each vessel element is carried out over a sequence of generation steps. For each step, we calculate a step-size (ϑ_i), lateral direction (α_{i_z}), and axial translation (ϕ_i). These values are subsequently used to grow the segment outward and towards its specified length during the i^{th} generation. The vessel element generation is summarized as a sequence of rotations, translations, and drawing processes. During the i^{th} generation, we rotate about the current position α_{i_z} radians, where α_{i_z} is randomly selected such that ($\frac{-\pi}{18} \leq \alpha_{i_z} \leq \frac{\pi}{18}$) is satisfied. Subsequently, we translate axially ϕ_i units, where ϕ_i is randomly selected such that $-\vartheta_i \leq \phi_i \leq \vartheta_i$ is satisfied, and then translate laterally ϑ_i units. At this time, a vascular segment is drawn between the new location and our previous one. We then continue this process until the desired branch length is obtained. Upon completion, if the branch has an order number greater than zero, we create a bifurcation at its terminal end: one child diverges at an angle ρ, such that ($\frac{\pi}{6} \leq \rho \leq \frac{\pi}{4}$) is satisfied; the other at an angle $-\rho$.

3.2.2.5 Synthetic volume construction.

Each synthetic model will be synthesized individually, in accordance with the preceding criteria. We will implement our model-based generator in the Python programming language, using the Visualization Toolkit (VTK) library. Specifically, we will construct the ground-truth of a respective model first, by constructing centerlines throughout each vessel element generation step (section 3.2.2.4). This data will be written in the serial vtkPolyData (VTP) format. We will then fit cylinders about these lines to create the synthetic volume of appropriate dimension, in accordance to sections 3.2.2.1, 3.2.2.2, and 3.2.2.3. This surface model will be exported in STereoLithography (STL) format and as a volumetric image stack. We will also export a comma-separated values (CSV) file for each model. For each vessel segment within the synthetic volume, we write its order number, length, and diameter.

3.2.3 Synthetic model analysis and validation

We analyze the properties of the filaments embedded in our synthetic volumes concerning their specific means (e.g., number of vascular elements, vessel length, and filament diameter) by Strahler number. We then compared each of these relationships against the ground truth data [12, 10]. To determine the extent of these relationships, we first used linear regression to fit curves of the following form:

$$\log_{10} \beta_n = a + bn, \tag{3.2}$$

where n is the Strahler number, β_n is an order-n mean characteristic of interest, and regression coefficients a and b. For the ground truth, the corresponding values of η_{μ_n}, D_{μ_n}, or L_{μ_n} were obtained from [10]. η_{μ_n} is the order-n mean number of elements; D_{μ_n} and L_{μ_n} are defined in section 3.2.2.1. We then considered the percent error amongst the regression coefficients for each relationship between groups as a measure of inconsistency between our models and the ground truth.

3.3 Results and analysis

Table 3.1: Mean values calculated across our synthetic models for the the number of vascular elements, vessel length, and filament diameter by order.

Order	No. Elements	Length (μm)	Diameter (μm)
0	214.78	109.97	6.05
1	49.81	212.38	8.76
2	8.05	411.05	12.59
3	2.92	802.22	18.29
4	1.00	1605.36	25.57

We present two of our synthetic cerebrovasculature models in Figure 3.5. The ground truth volume and centerline for Figure 3.5 right are shown in Figure 3.6. The mean number of vascular elements, vessel length, and filament diameter calculated across our models are presented by order

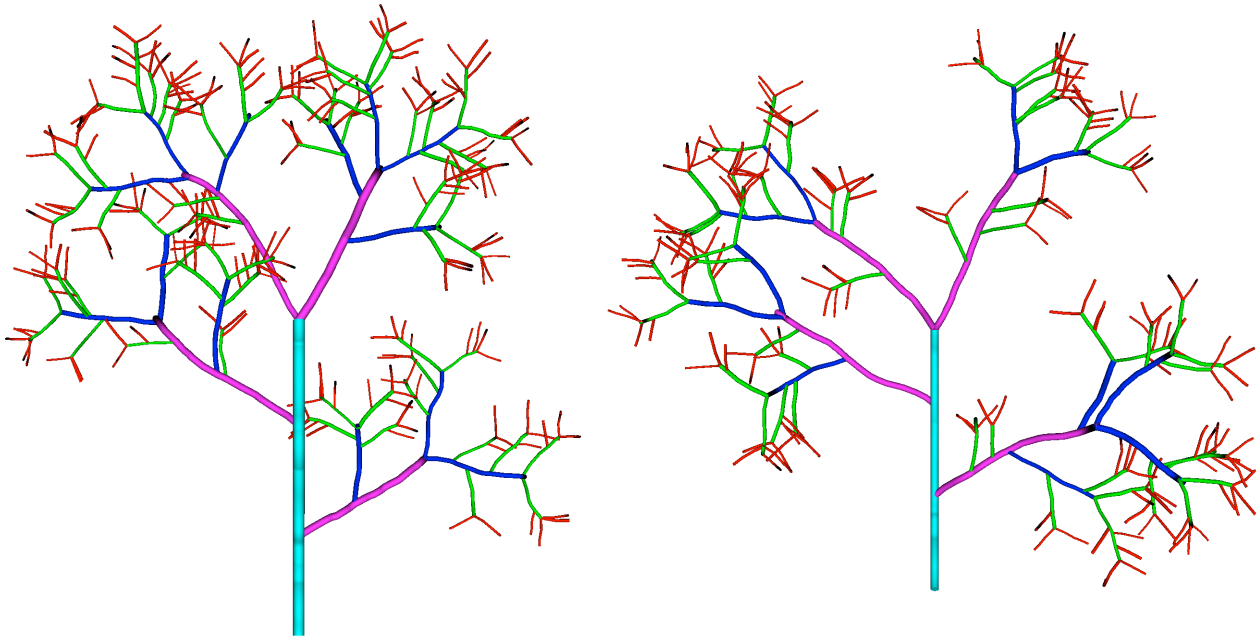


Figure 3.5: We present two of our three-dimensional, synthetic models in this figure. We have labeled the vessel segments of order no. 0, 1, 2, 3, and 4 in color by red, green, blue, magenta, and cyan respectively.

number in Table 3.1.

Table 3.2: The coefficients of regression and determination (R^2) of the fitted lines approximating the effect of order on the log-mean characteristics.

	Synthetic Data			**Ground Truth**		
	a	b	R^2	a	b	R^2
Elements	2.30	-0.59	0.988	2.19	-0.58	0.988
Length	2.03	0.30	0.999	2.11	0.29	0.971
Diameter	0.74	0.16	0.990	0.86	0.16	0.956

The relationships amongst log-mean characteristics and Strahler number between groups are reported in Table 3.2 by the regression coefficients of the fitted lines approximating them. The only percent error found in excess of 10% between groups was coefficient a for the order number by log-mean diameter relationship (at 12.98%).

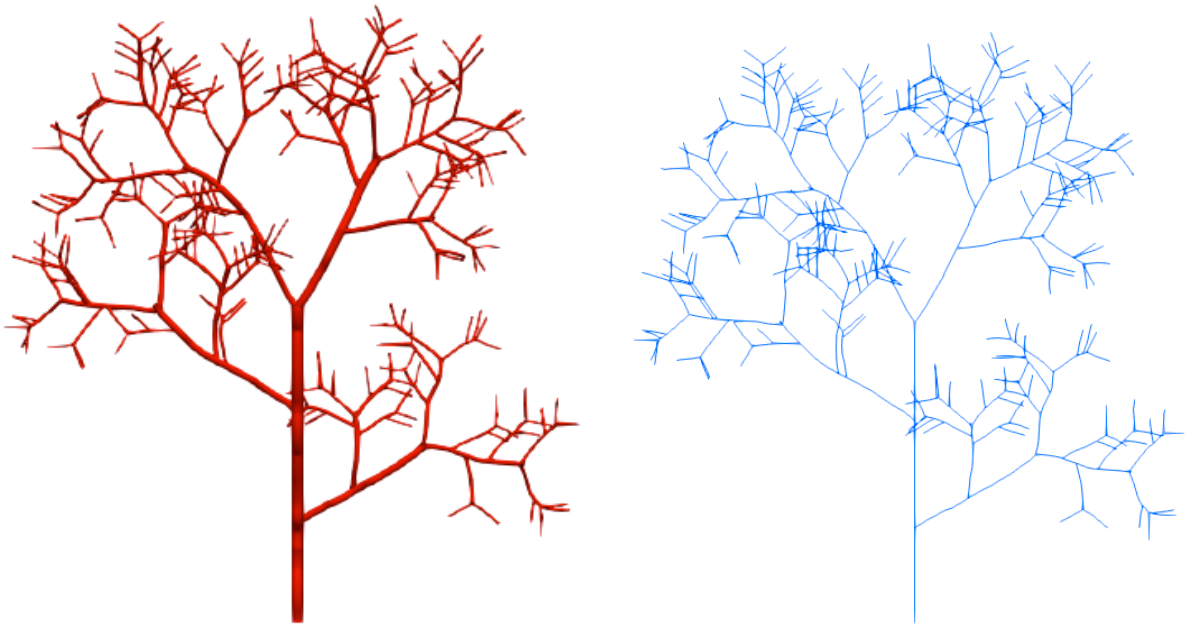


Figure 3.6: An example of one of our synthetic volumes is shown on the left; its corresponding ground truth centerline data are presented on the right.

3.4 Summary

In this chapter, I detailed our approach towards the construction of more complete cerebrovasculature models by addressing some of the limitations of digital phantoms and expert-labeled volumes. The primary contribution of this work is the introduction of the means to assemble synthetic datasets of vascular trees that are compared to biological datasets in vessel attributes and volume characteristics. It would be straight forward to extend the attributes considered in our models from diameter, length, and the number of elements to include additional features such as branch angle and vessel tortuosity. To model such characteristics, it would be necessary for a researcher to describe them as they present in a biological dataset. To model strong deviations from the tree-like structural topology considered here, for instance, the mesh-like capillary network, the method detailed in this chapter is limited. In the next chapter, we will present a technique capable of generating models with significantly different topologies within a single synthetic dataset.

Overall, our results show that our synthetic models are in good correspondence to the natural

distributions of vessel diameters and lengths found in the human collateral sulcus and its vascular density. Furthermore, by generating synthetic models, we know the composition of the vasculature embedded within and can thus make entirely accurate ground truths. Using the synthetic volumes and ground truth data, the accuracy of vascular segmentation algorithm can be assessed. Moreover, because the synthetic volumes are reconstructed directly from brain vasculature data, an assessment is made on embedded filaments that are representative of the topological and geometrical characteristics that would be encountered in practice.

4. DATA-DRIVEN SYNTHETIC CEREBROVASCULATURE MODELS

4.1 Motivation

In Chapter 3, we constructed synthetic morphologies according to the natural distributions of vessel diameters, lengths, and densities found in 'real' cerebrovasculature trees. We want to develop models that are more representative of the cerebrovasculature system by taking in to account additional characteristics, including those of vessel branch angle, circulatory anastomoses, and the mesh-like capillary network. One of the limitations of our previous method is that we must rely on biological ground truth data. This requires a researcher to go through and describe the different attributes of vascular elements residing in a specific cortical region. However, it is unlikely that descriptions derived from one area will apply to others [28]. To construct models that are more representative of the cerebrovasculature across the brain, the vasculature presenting in many brain regions would need to be described. If we would like to make our models more representative of filament structure, more attributes (such as branch angle, etc.) would need to be characterized. This would make the process of gathering ground truth data more difficult for the researcher (more attributes to record and more regions to describe).

In the current chapter, we detail a data-driven methodology to acquire vascular models that together characterize vascular structure across many different brain regions. We accomplish this by first obtaining the centerlines of vascular filaments embedded in a sub-sample of a vascular dataset using a segmentation algorithm. Reconstruction about the centerlines (i.e., the model's ground truth) begins after that. Subsequently, a three-dimensional stack of tiff images (i.e., 'the stack') embedding the vasculature of the model is generated using a voxelization algorithm. The initial segmentation serves as the ground truth from which the stack was built. This means that we know the location and characteristics of the vasculature embedded within the stack along with

the properties of the volume as a whole. Since the synthetic volume is biologically grounded with real cerebrovasculature, it is representative of the topological and geometrical characteristics of filaments from the biological dataset.

4.2 Methodology

4.2.1 Obtain a cerebrovasculature sample

To derive a synthetic cerebrovasculature model in a data-driven fashion, we first obtain representative vascular filaments for the cerebral region of interest. Such a sample can be acquired in many ways. Images slices of a randomly selected region can be downloaded directly from an online brain atlas [101] to construct a raster image stack or the geometries and topological information encoding a cerebrovasculature sample can be retrieved from a graph or database [99]. In practice, the former approach requires vascular filaments to be segmented from the raster image stack and converted to vector format. Once this is done, both approaches align at the same processing stage. Before proceeding, we require that the biological sample selected has its topology and geometry stored as a series of nodes and edges. We selected the *SWC* format (SWC is not an abbreviation, but the initials of the developers) for this encoding [102].

```
## id, type, x, y, z, radius, parent
25830 3 4915.886 4689.92 1370.629 8.18 -1
25831 3 4890.53 4722.949 1418.14 10.58 25829
25829 3 4896.63 4704.84 1398.8 11.559 25830
25832 3 4890.624 4734.837 1425.227 7.69 25831
25833 3 4890.051 4745.961 1431.775 6.245 25832
25835 3 4886.8 4757.2 1441.0 5.522 25833
```

Figure 4.1: Example of the *SWC* file format. Each line of the file represents a line end point. For each 'child' end point, its x, y, and z coordinates are stored, along with its radius and its 'parent' end point. Line segments are constructed between 'child' and 'parent' end points; these line segments encode the different vascular filaments embedded in the imaging volume.

The idea behind the *SWC* format is that each vascular segment (vessel between two bifurca-

tions) can be embedded as a series of line segments and line endpoints. The entire geometry and topology of a vascular sample are captured by ensuring that every time a vessel segment deviates from a straight line, that deviation is captured through the construction of another line endpoint that together with its parents, constructs a new line segment to represent a vascular element. This line represents the centerline of the vascular filament from which it is derived. The *SWC* format establishes a hierarchy of line endpoints connected into line segments encoding every vascular component in the dataset. Every line endpoint is written with a unique identifier (i.e., ID), its x, y, and z position, its radius, and parent endpoint's ID (of that to which it forms a segment). Based on this information, a sequence of straight lines can be drawn between parent and child endpoints, where each endpoint is centered at its x, y, and z position. A cone can be fit between the endpoints, with the radius of each end matching that of the endpoints. An example of this file format is shown in Figure 4.1.

4.2.2 Ground truth centerline and volume construction

Using the topological and geometrical information encoded about the cerebrovasculature sample in the *SWC* file, we reconstruct the volume in vector graphics form using geometrical primitives. Many different 3D graphics frameworks can be used to achieve this end. We selected the Visualization Toolkit [103] primarily due to our familiarity with it. Our reconstruction process can be summarized as the drawing of the lines (vascular filaments) as defined in the *SWC* file; the result is like that illustrated in Figure 4.2. We save this representation of the vessel centerlines in *VTK* format; this is the ground truth that describes precisely where the center of each vessel is located. To construct the volume, we fit a cone about each line, with the radius at each end matching that of the line's endpoint from the *SWC* file. The result of this process for Figure 4.2 is shown in Figure 4.3. The polydata for this volume is written out in *VTK* format; this is the ground truth volume of our cerebrovasculature model.

Figure 4.2: An example of the centerline reconstructed for a data-driven synthetic cerebrovasculature models.

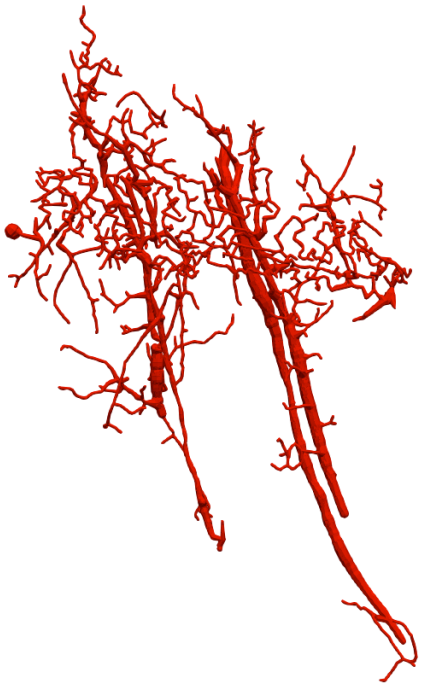


Figure 4.3: An example of the volume of one of our synthetic cerebrovasculature models. The volume was reconstructed by fitting cylinders about the centerline in Figure 4.2.

4.2.3 Ground truth volume voxelization

We must voxelize our vectorized ground truth volume to construct a raster imaging volume. Many different tools perform voxelization, and we selected *BINVOX* [104]. The way that *BINVOX* performs this operation is no different than most other voxelization methods. First, the mesh representation of the volume is normalized to fit inside a $1.0 \times 1.0 \times 1.0$ cube with its origin at (0.0,0.0,0.0). This is achieved by completing a translation operation with a uniform scale. The resultant unit cube is then voxelized. For a voxel model of resolution d and coordinates beginning at (0,0,0), the coordinates inside the unit cube of a given voxel located at (i, j, k) can be calculated by the following equation:

$$(x_n, y_n, z_n) = ((i + 0.5)/d, (j + 0.5)/d, (k + 0.5)/d), \tag{4.1}$$

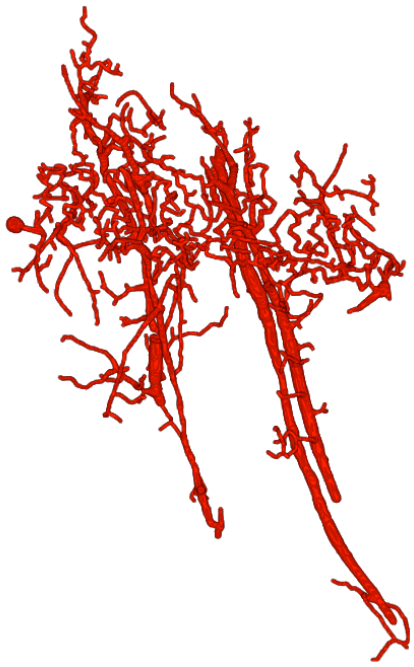


Figure 4.4: Voxelization of cerebrovasculature model presented in Figure 4.3.

with 0.5 added to each location point to obtain the center of the voxel cell.

In our work, we selected $d = 1024$, the highest resolution afforded by *BINVOX* for the voxelization process. This resulted in a $1024 \times 1024 \times 1024$ image volume of the voxelized model. We perform morphological hole filling on the resultant binary image volume. To make this raster image volume align with the vectorized ground truth volume and centerline, we scale it based on the scaling factor used during the normalization procedure of the mesh. Specifically, given the *BINVOX* normalization scaling factor sf_{norm} and model resolution $d = 1024$ (as denoted above), we calculate the scaling factor sf_{img} for the raster image volume by the following equation:

$$sf_{img} = (1.0/sf_{norm})/d \tag{4.2}$$

The voxelization of the cerebrovasculature model presented in Figure 4.3 according to this procedure is shown in Fig 4.4. At this point, the synthetic dataset is similar to what would be acquired from the biological specimen, albeit without noise.

4.2.4 Rough mesh construction of voxelized volume

We construct a rough mesh representation of the filament structure embedded in the voxelized volume derived in Section 4.2.3. To do this, we must first extract the surface of the filament structures. There are many different algorithmic approaches to this problem. We decided to solve it by contouring the image using the discrete marching cubes algorithm [105]. Once the surface is obtained as polydata by the marching cubes algorithm, we store the surface in *VTK* format.

4.2.5 Ground truth centerline and volume translation

To align the centerline and volume ground truth to the raster imaging volume, we first set the origin of the centerline and volume ground truth to $(0.0,0.0,0.0)$. This is achieved using a basic translation operation applied to the polydata of the centerline and volume ground truths. For this process, we use the rough mesh constructed in Section 4.2.4 as an orientation aid: it shares the same placement as the imaging data in Euclidean space. We compare the alignment of the centerline and volume to the rough mesh. If needed, we rotate the centerline and volume polydata as to align it with the rough mesh.

4.3 Results and analysis

Five data-driven synthetic cerebrovasculature models acquired using the methodology detailed in this chapter are presented in Figures 4.5, 4.6, 4.7, 4.8, and 4.9.

The extent to which the filaments embedded in these synthetic models correspond to the vasculature found in the animal model from which they are derived is dependent on many factors that make quantification difficult. When processing the biological imaging dataset acquired from an animal model, the application of different image processing techniques can damage the surface of the vascular filaments, causing measures of radius, branch angles, and other attributes to be close but not entirely correct. Furthermore, the segmentation algorithms used to extract the embedded vascular filaments can degrade in the presence of noise. In the generation of our models, we view these issues as irrelevant. Once we receive the segmented filaments, we assume the data ground truth and reconstruct an imaging volume from it. Our models are a *best approximation* of the un-

derlying cerebrovasculature attributes that one can expect to be found in the biological specimen. Thus, our models are well suited for the evaluation of image processing techniques and validation of segmentation algorithms across different experimental conditions.

4.4 Summary

In this chapter, I detailed our approach to constructing data-driven models that are more representative of the cerebrovasculature structure than those generated stochastically in Chapter 3. In addition to creating models that have filament diameter, length, and number similar to those found in biological datasets, our data-driven models also incorporate branch angles, circulatory anastomoses, and tortuosity measures. This is because our models have been derived from 'real' cerebrovasculature data; they can, therefore, capture the tree-like topology considered in Chapter 3 down to the mesh-like capillary networks all within a single dataset. Furthermore, using an array of biological samples, we can construct synthetic models that exhibit the variation in vasculature structure across the brain. This has been made feasible by the whole-brain cerebrovasculature datasets: samples of cerebrovasculature can be acquired from different brain regions.

Our synthetic models address not only the limitations of digital phantoms and expert-labeled datasets but also those developed in Chapter 3. However, our artificial models do not replace digital phantoms and expert-labeled datasets, but can be used in conjunction with them. We believe that initial rounds of evaluation should utilize smaller models (e.g., digital phantoms). Once the processing procedure or tracing algorithm appears stable, testing should then commence on a volume of similar sample size and dimensionality of cerebrovasculature as the dataset, such as our synthetic models. In the past, representative datasets were difficult to obtain. Our method detailed in this chapter provides a simple means to achieve such datasets while maintaining an underlying ground truth.

Our original contribution is the means to assemble large-scale, high-resolution datasets of vessel attributes and volume characteristics comparable to those in the biological specimen. Our method addresses the critical need for large-scale models for the validation process of algorithms applied to cerebrovasculature imaging datasets. Our models are readily constructed and reflect

Figure 4.5: Cerebrovasculature model acquired from KESM India Ink dataset. This voxel-based model is $1158 \times 1158 \times 1158$ pixels derived from imaging data acquired at a voxel resolution of $0.6\,\mu\text{m} \times 0.7\,\mu\text{m} \times 1.0\,\mu\text{m}$. The total size of this model is $1.4\,\text{GB}$.

Figure 4.6: Cerebrovasculature model acquired from KESM India Ink dataset. This voxel-based model is $1280 \times 1280 \times 1280$ pixels derived from imaging data acquired at a voxel resolution of $0.6\,\mu\mathrm{m} \times 0.7\,\mu\mathrm{m} \times 1.0\,\mu\mathrm{m}$. The total size of this model is $2.1\,\mathrm{GB}$.

Figure 4.7: Cerebrovasculature model acquired from KESM India Ink dataset. This voxel-based model is $1617 \times 1617 \times 1617$ pixels derived from imaging data acquired at a voxel resolution of $0.6\,\mu\text{m} \times 0.7\,\mu\text{m} \times 1.0\,\mu\text{m}$. The total size of this model is $4.2\,\text{GB}$.

Figure 4.8: Cerebrovasculature model acquired from KESM India Ink dataset. This voxel-based model generated from imaging data acquired at a voxel resolution of $0.6\,\mu\mathrm{m} \times 0.7\,\mu\mathrm{m} \times 1.0\,\mu\mathrm{m}$; model is in excess of $2400 \times 2400 \times 2400$ pixels. The total size of this model is greater than of $10\,\mathrm{GB}$.

Figure 4.9: Cerebrovasculature model acquired from KESM India Ink dataset. This voxel-based model generated from imaging data acquired at a voxel resolution of $0.6\,\mu\mathrm{m} \times 0.7\,\mu\mathrm{m} \times 1.0\,\mu\mathrm{m}$; model is in excess of $2400 \times 2400 \times 2400$ pixels. The total size of this model is greater than of $10\,\mathrm{GB}$.

the morphological features and multi-scale nature of the brain's angioarchitecture. Through their application, we can quantitatively access the accuracy of filament processing and extraction techniques and their scalability across larger-scale, higher-resolution datasets.

5. MODEL-BASED VALIDATION SYSTEM

5.1 Motivation

It is challenging to quantitatively evaluate the effect of image processing algorithms on filament structure and determine the correctness of a segmentation algorithm (SA) in filament extraction. This is true irrespective of the methodology used to image the biological specimen, and especially so when the sample size and dimensionality per grain of cerebrovasculature in acquired image volume is significantly large [106]. When it comes to vascular datasets, there can be many different image processing procedures applied to an imaging volume before an SA extracts the embedded filaments; such methods comprise the 'preprocessing' stage. When it comes time to segment the vasculature from the image volume, there are many different SAs that one can choose. Section 2.3 provides an overview of some of the commonly applied preprocessing and segmentation techniques for cerebrovasculature datasets acquired by LM. The primary goal in this chapter is to describe our model-based validation system and analyze results from its application. Our system is designed to quantify the extent to which a preprocessing procedure degrades filament structure and the degree that an SA successfully extracts the embedded vasculature (albeit not at the same time).

In the literature, validation is typically limited to the SA and is frequently conducted upon digital phantoms and manual-labeled volumes [96]. This is achieved by measuring the correspondence of the medial axis of a model in some validation set against the centerline obtained by an SA or by determining the overlap of the extracted filaments to those in a ground truth volume. The effect on filament integrity of image processing techniques is rarely quantified. This practice occurs despite the availability of digital phantoms and manual-labeled volumes to which image processing techniques could be applied. The result of this application could then be compared against the digital phantom or manual-labeled ground truth.

In cerebrovasculature research, it is imperative that vessel integrity is maintained. This is because the relationship of the microvasculature to surrounding tissue is critical to perfusion, as is the geometry and topology of the cerebrovasculature that influences the resistance to blood flow. The implications of incorrect measurements of vasculature attributes are significant, yet are overlooked in the literature. To emphasize this, take the Hagen-Poiseuill solution (Eq. 2.4) to volumetric flow rate (volume/time). This equation describes flow rate as directly proportional to the product of the radius to the fourth power and the pressure drop across the vessel; inversely proportional to filament length. This means that the blood vessel radius does not need to change significantly to increase (or decrease) flow, limiting the extent to which dilation (or constriction) is required under hypoxic (or hyperoxic) conditions [15]. What this also means is that processing or segmentation procedures that induce *small* damages to filament structure can have a *large* impact on our conclusions drawn about blood flow and tissue perfusion in studies using imaging data.

Furthermore, inaccurate measurements of the radius can impact flow separation at bifurcations, and this may have implications in studies of blood flow. Murray's law (Eq. 2.3) describes the optimal angle between parent and child vessels and is calculated for the radius of the vessels comprising the bifurcation. As the angle changes, the velocity profile, wall shear stress, and other attributes of the flow change too. Moreover, the geometry of the bifurcation determines how blood is distributed to the child's vessels and also influences the resistance to flow. Degradation at points of bifurcation in the imaging data from processing or segmentation can also impact analyses about blood flow and tissue perfusion.

In chapters 3 and 4, we have introduced two ways to generate synthetic models that address some of the limitations of digital phantoms and manually-labeled volumes. In this chapter, we will detail a model-based validation system that can be used in conjunction with those models. Our system does not only validate segmentation algorithm performance but also can determine whether a specific (or combination of) image processing algorithm(s) damages filament structure.

5.2 Methodology

The generation of our cerebrovascular models in Chapters 3 and 4 results in ground truth vectorized skeletons, vectorized surfaces, and rasterized image volumes. Using our models, we can assess the accuracy of different segmentation algorithms or the effect of image processing regiments on the structures embedded within them. Our method presented here provides the means to study these aspects quantitatively. We provide the ability to conduct such studies by reporting quantitative measurements of the centerline and volumetric correspondence of experimental data to a model's ground truth.

5.2.1 Volumetric correspondence

5.2.1.1 Rasterize the segmented volume

If the experimental result (ER) is not already in raster graphics form, we will first generate a binary image volume from the vectorized data using *BINVOX* [104]. We first align the ER to the GT using the procedure detailed in Section 4.2.5. We then follow the same procedure as in Section 4.2.3 to voxelize the ER, albeit without rescaling. This process results in a binary image volume of resolution $d = 1024$. We subsequently compare this image volume for the ER against that generated by *BINVOX* for the ground truth (GT).

5.2.1.2 Calculate sensitivity, specificity, precision, Jaccard index, and dice coefficient

We calculated sensitivity, specificity, precision, Jaccard index, and dice coefficient as measures of correspondence of the ER to the GT. For the ER and GT binary image volumes, foreground (FB) is denoted as the positive and the background (BG) is denoted as the negative class. True positive (TP) is defined as the total number of voxels labeled as foreground by both the ER and GT. Meanwhile, true negative (TN) is defined as the total number of voxels labeled as BG by both the ER and GT. False positive (FP) and false negative (FN) are defined accordingly. Following these definitions, we derive sensitivity, specificity, precision, Jaccard index, and dice coefficient as:

$$\text{Sensitivity} = P(y = FG \mid GT = FG) = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5.1}$$

$$\text{Specificity} = P(y = BG \mid GT = BG) = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{5.2}$$

$$\text{Precision} = P(GT = FG \mid y = FG) = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5.3}$$

$$\text{JI} = P(y = SA \cap GT = FG \mid y = SA \cup GT = FG) = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{5.4}$$

$$\text{DC} = \frac{2 \times \text{JI}}{1 + \text{JI}} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{5.5}$$

Both the DC and JI approach a value of 1.0 for very similar results and 0.0 when they share no voxels. They are appropriate when the number of FG voxels is much less than BG or when the detection accuracy of FG detection is more critical than BG [78].

We calculate these measures of correspondence by iterating over the image slices of the ER and GT volumes, determining the TP, FP, TN, and FN values for corresponding slices, and summing these metrics as we proceed through the volumes. The overall measures of TP, FP, TN, and FN are used to calculate sensitivity, specificity, precision, Jaccard index, and dice coefficient.

5.2.1.3 Report of results

Our model-based validation system reports sensitivity, specificity, false positive rate (FTR), false negative rate (FNR), precision, Jaccard index, and dice coefficient to the user. We leave the interpretation of these measures of correspondence up to the user. We do this as what constitutes an acceptable level is often dependent on the application.

5.2.2 Centerline correspondence

When an SA is applied to an imaging model, and the output is filament centerline information, quantitative analysis of SA accuracy is more complicated given the sparse nature of centerline data [106]. If the radius is encoded, in conjunction with the centerline data, the volume can be reconstructed and volumetric correspondence assessed according to Section 5.2.1. Depending on the segmentation protocol, the centerline may be extracted first, and the radius calculated using that data in conjunction with the imaging volume [7]. In these scenarios, the radius is calculated through the evolution of a level-set surface outward from the skeleton [107], or by extracting the network isosurface. The information acquired about the surface structure and diameter of vessel filaments can be used to determine each element's radius [7]; the volume can then be readily reconstructed.

If the volume cannot be reconstructed, we can still quantify the agreement of the SA's centerline output to our model's GT. Due to the sparse nature of the centerline representation, there will not necessarily be a one-to-one correspondence between the centerline ground truth and the centerline calculated by an SA. Therefore, we propose a novel graph-based approach to compare the ground truth of a model to that output by an SA.

5.2.2.1 Graph construction

Before proceeding, we require that the ER and GT centerlines are stored as a series of nodes and edges. We selected the *SWC* format (SWC is not an abbreviation, but the initials of the developers) for this encoding [102] and elaborate on this scheme in Section 4.2.1. The centerlines for the ER and GT can be represented as a sequence of nodes with a straight line-edge between them. Viewing

the ER and GT centerlines in this fashion allows us to construct independent graph representations of each easily.

There are many different libraries that can be used to construct a graph given the nodes and edges. We decided to use the networkx package [108] based on our familiarity with it. Given that the *SWC* format establishes a hierarchy of line endpoints connected into line segments encoding every vascular component in the dataset, we can easily convert this encoding into graph form. For every line endpoint in the file, we construct a new node n into our graph encoding with its unique identifier (i.e., ID) specified in the *SWC* file, along with with its x, y, and z position, and its radius (if applicable). Given that the node's parent's endpoint ID is also specified on its line, we insert an edge into our graph connecting n with its parent. Based on this information, the sequence of straight lines representing the centerlines is encoded in graph form [96].

5.2.2.2 Initialization of k-d tree with ground truth centerline nodes

We subsequently iterate over the nodes comprising the GT centerline and insert them into the k-d tree data structure in $\mathcal{O}(\log n)$ time.

5.2.2.3 Centerline correspondence

We iterate through the nodes comprising the ER centerline. We view each node x as an SAddle-point of sorts and find the node closest to n to node x in the GT using the k-d tree. This process returns node n from the GT. We then find two additional nodes, a and b, from the GT closest to x, with the constraint that those nodes must be connected to n by an edge. We formulate a line $l1$ between n and a and calculate the distance $d1$ of x to that line. Subsequently, we formulate another line $l2$ between n and b and calculate the distance $d2$ of x from that line. We report the minimum distance calculated as the deviation of x from the GT. An illustrative example is shown in Figure 5.1. This process is done for each node in the graph encoding the ER centerline.

5.2.2.4 Correspondence metric reported

We indicate the correspondence of the ER centerline to that of the ground truth as the average deviation of each node x comprising the ER from the ground truth centerline.

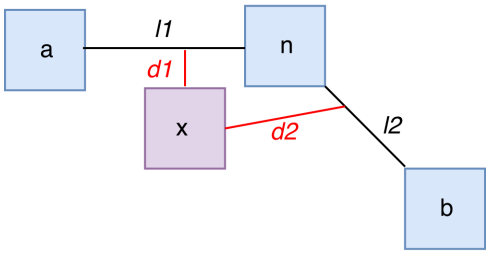


Figure 5.1: Diagram describing the calculation of deviation of a node x from the line described described by points a, n, and b with edges $l1$ and $l2$. In order to calculate the deviation distance of node x from the line, we take the minimum Euclidean distance that it is from $l1$ and $l2$.

5.2.2.5 Length difference reported

The correspondence metric described in Section 5.2.2.4 is necessary, but not sufficient on its own. For example, if the segmentation output consists of two nodes that both happen to lie on the ground truth centerline, then the correspondence metric reported would suggest perfect correspondence. However, if those two nodes only encoded a fraction of the ground truth centerline, then this is wrong. Therefore, we will also report the length difference (ϕ) as:

$$\phi = |1 - \frac{\mathrm{L_R}}{\mathrm{L}_A}|, \tag{5.6}$$

where $\mathrm{L_R}$ and $\mathrm{L_A}$ are the length of vascular data in ground truth and segmentation output respectively.

5.3 Results and analysis

5.3.1 Volumetric correspondence

To illustrate the utility of our model-based validation for volumetric correspondence, we acquired a model from the KESM India Ink dataset using our method detailed in Chapter 4. The model selected is shown in Figure 5.2. In this model, the background color is grayscale 144 and

Figure 5.2: Cerebrovasculature model acquired from KESM India Ink dataset. This voxel-based model is $633 \times 633 \times 633$ pixels derived from imaging data acquired at a voxel resolution of $0.6\,\mu\mathrm{m}$ $\times$ $0.7\,\mu\mathrm{m}$ $\times$ $1.0\,\mu\mathrm{m}$. The total size of the voxelized model is $253.7\,\mathrm{MB}$.

the vessel 44. Specifically, we wanted to quantify the accuracy of the voxel scooping segmentation algorithm [9] when applied to a dataset that had been degraded with Gaussian noise and subsequently processed using a two-dimensional 5×5 median filter ($n = 5$) and followed-up by global thresholding ($t = 100$). These are common preprocessing procedures and are detailed in Section 2.3. Thereafter, we explore how the median filter alone effects filament structure when no noise is present.

5.3.1.1 Quantification of algorithmic performance in the presence of noise

To illustrate the utility of our model-based validation system, we degraded the model presented in Figure 5.2 with Gaussian noise. Specifically, we constructed seven noisy models:

- Noisy Model I, a pixel's grayscale value was offset by -5–5
- Noisy Model II, offset by -10–10

- Noisy Model III, offset by -20–20
- Noisy Model IV, offset by -40–40
- Noisy Model V, offset by -60–60
- Noisy Model VI, offset by -80–80
- Noisy Model VII, offset by -100–100.

The probabilities for these offsets are shown in Figure 5.3 for each noisy model. The maximum grayscale value that a pixel could become in any noisy model was 255 and the minimum 0.

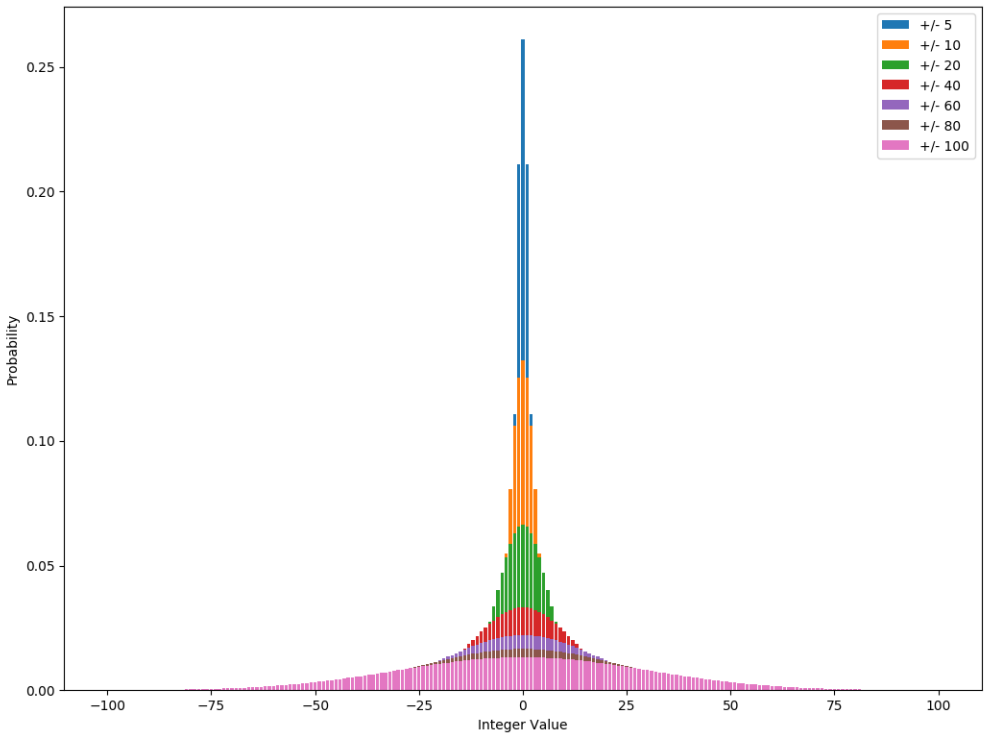


Figure 5.3: The probability that a grayscale pixel in a respective model will be offset by the value in specified range. The minimum value after the offset is 0; the maximum value is 255.

A slice of the binary ground truth for Figure 5.2 is shown in Figure 5.4. We illustrate the level of filament structure degradation in the noisy models in the same slice for Noisy Models I-VII in Figures 5.5, 5.7, 5.9, 5.11, 5.13, 5.15, and 5.17. The degradation in these figures is representative of that across each respective noisy imaging volume. We also present the histograms for each of the noisy image volumes in Figures 5.6, 5.8, 5.10, 5.12, 5.14, 5.16, and 5.18.

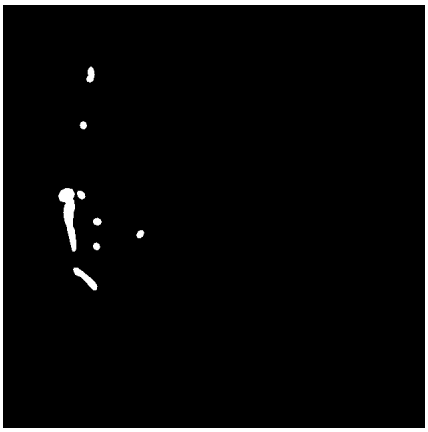


Figure 5.4: Slice from the binary ground truth volume.

For each noisy model, we applied a two-dimensional median filter of neighborhood $n = 5$ slice-by-slice, followed by global threshold at $t = 100$. We subsequently applied the voxel scooping method [9] to the volume to segment the filaments. We calculated volumetric correspondence of the segmented filaments (i.e., the ER) to the model's GT per the procedure outlined in Section 5.2.1. We report the algorithm's sensitivity (true positive rate), specificity (true negative rate), false positive rate, and false negative rate in Figures 5.19, 5.20, 5.21, and 5.22 respectively. Precision (positive predictive value), Jaccard index, and dice coefficient are reported in Figures 5.23, 5.24, and 5.25 respectively.

It is interesting that TPR (Figure 5.19) increases in the presence of more severe noise while the TNR decreases (Figure 5.21). Likewise, that FPR (Figure 5.21) increases while the FNR decreases (Figure 5.21). The observed trends for TPR, TNR, FPR, and FNR are not strongly linear in nature.

Figure 5.5: Noisy Model I. Image slice from the middle of the image volume detailing the level of noise representative throughout the model.

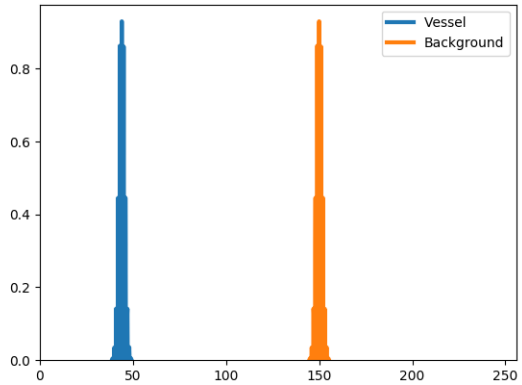


Figure 5.6: Histogram of the image volume of Noisy Model I. The x-axis shows the potential grayscale values and the y-axis the probability that they occur across the volume. The grayscale values of the filaments (vessels) are shown in blue; the values of the background in orange.

Figure 5.7: Noisy Model II. Image slice from the middle of the image volume detailing the level of noise representative throughout the model.

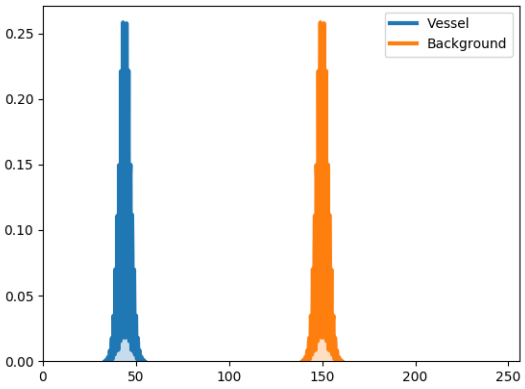


Figure 5.8: Histogram of the image volume of Noisy Model II. The x-axis shows the potential grayscale values and the y-axis the probability that they occur across the volume. The grayscale values of the filaments (vessels) are shown in blue; the values of the background in orange.

Figure 5.9: Noisy Model III. Image slice from the middle of the image volume detailing the level of noise representative throughout the model.

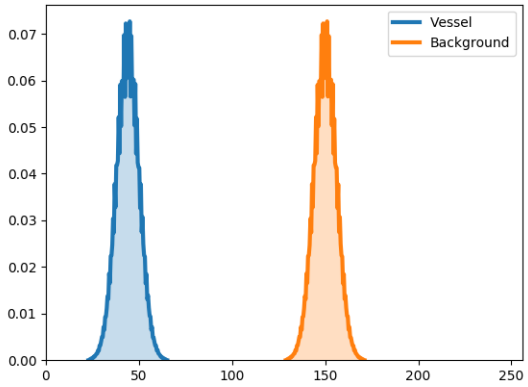


Figure 5.10: Histogram of the image volume of Noisy Model III. The x-axis shows the potential grayscale values and the y-axis the probability that they occur across the volume. The grayscale values of the filaments (vessels) are shown in blue; the values of the background in orange.

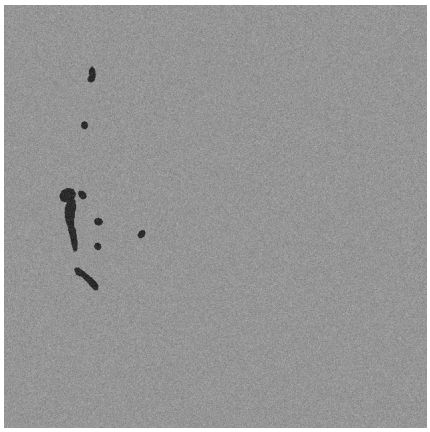


Figure 5.11: Noisy Model IV. Image slice from the middle of the image volume detailing the level of noise representative throughout the model.

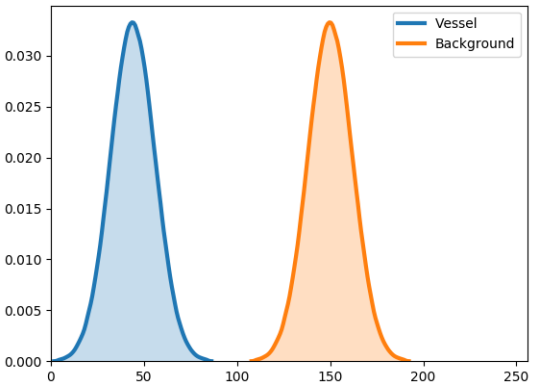


Figure 5.12: Histogram of the image volume of Noisy Model IV. The x-axis shows the potential grayscale values and the y-axis the probability that they occur across the volume. The grayscale values of the filaments (vessels) are shown in blue; the values of the background in orange.

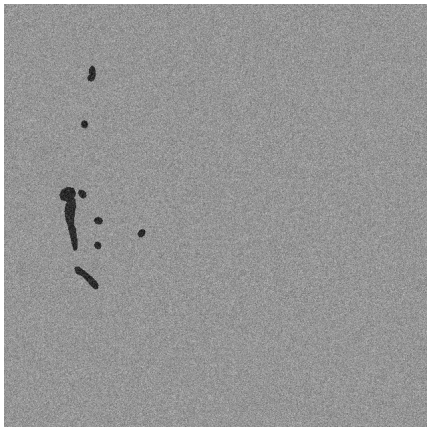


Figure 5.13: Noisy Model V. Image slice from the middle of the image volume detailing the level of noise representative throughout the model.

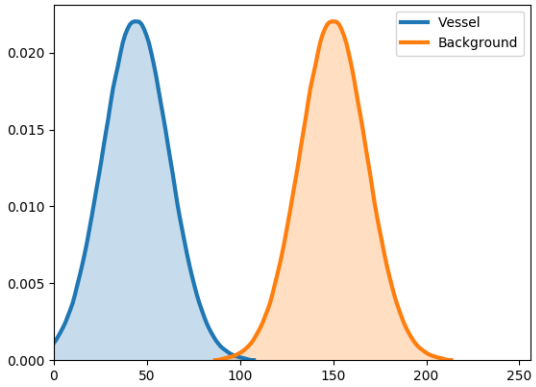


Figure 5.14: Histogram of the image volume of Noisy Model V. The x-axis shows the potential grayscale values and the y-axis the probability that they occur across the volume. The grayscale values of the filaments (vessels) are shown in blue; the values of the background in orange.

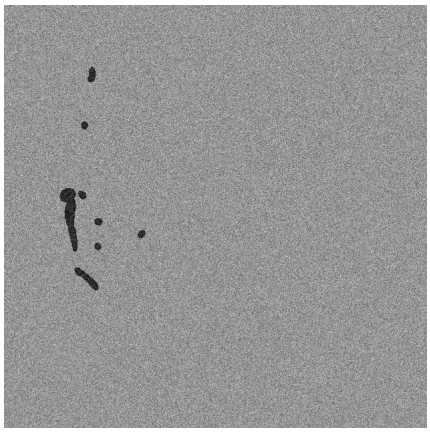


Figure 5.15: Noisy Model VI. Image slice from the middle of the image volume detailing the level of noise representative throughout the model.

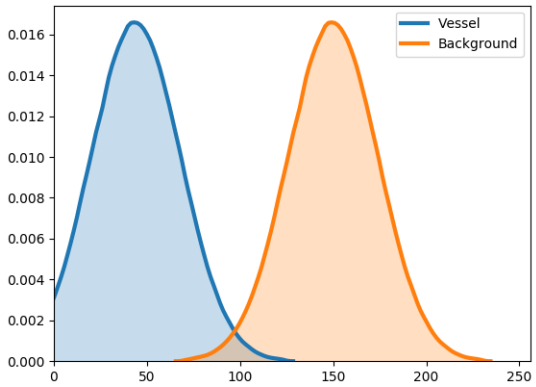


Figure 5.16: Histogram of the image volume of Noisy Model VI. The x-axis shows the potential grayscale values and the y-axis the probability that they occur across the volume. The grayscale values of the filaments (vessels) are shown in blue; the values of the background in orange.

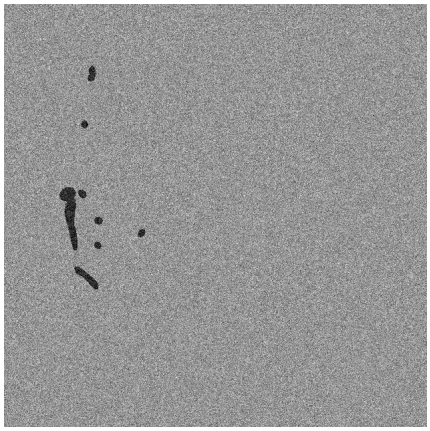


Figure 5.17: Noisy Model VII. Image slice from the middle of the image volume detailing the level of noise representative throughout the model.

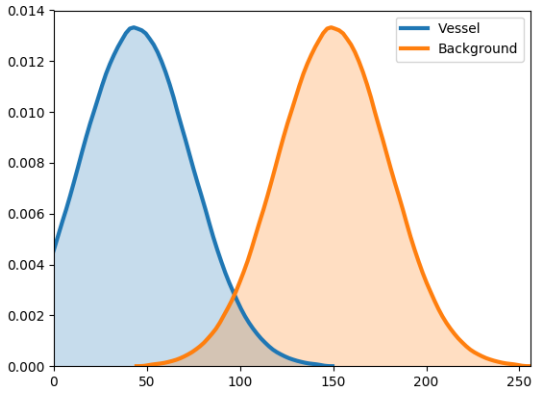


Figure 5.18: Histogram of the image volume of Noisy Model VII. The x-axis shows the potential grayscale values and the y-axis the probability that they occur across the volume. The grayscale values of the filaments (vessels) are shown in blue; the values of the background in orange.

Regardless of the shape of the trend, it is important to question the nature of it. That is, how could TPR and FPR both increase in the presence of noise? How could TNR and FNR both decrease?

As the image volumes are degraded with more and more Gaussian noise, the pixel values of the foreground and background *can* become closer together and eventually overlap. When the median filter is applied, it is likely that more pixels comprising the filament's structure will be characterized correctly as foreground. However, by the same token, that filament pixels will also be labeled incorrectly as part of the background. This increases both TPR and FPR, respectively. With little noise, a filament's edges are sharp as the dark foreground meets the lighter background pixels. When the median filter is applied to datasets where the ROI presents in high contrast to the background, the edges at the boundary of the filaments stay sharp; however, they are often degraded. Figure 5.19 and Figure Figure 5.20 show that even in low noise and high foreground to background contrast conditions, only 80% of the pixels are labeled correctly. Furthermore, this percentage does not change drastically as the level of Gaussian noise present in the image volume increases. These effects suggest that the median filter could be more responsible for the resultant

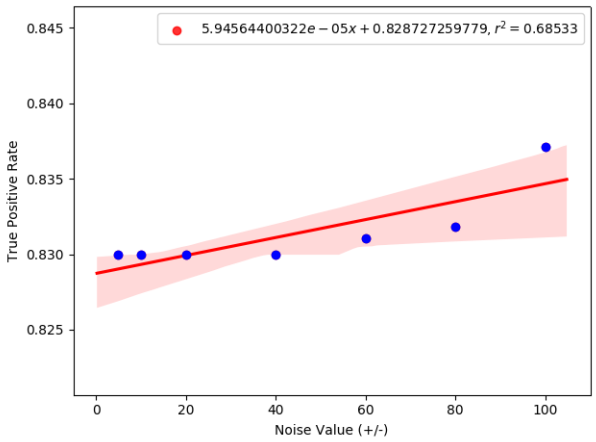


Figure 5.19: Noise value by the true positive rate for the correspondence of filaments segmented from volumes degraded with varying levels of Gaussian noise to the volume's ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

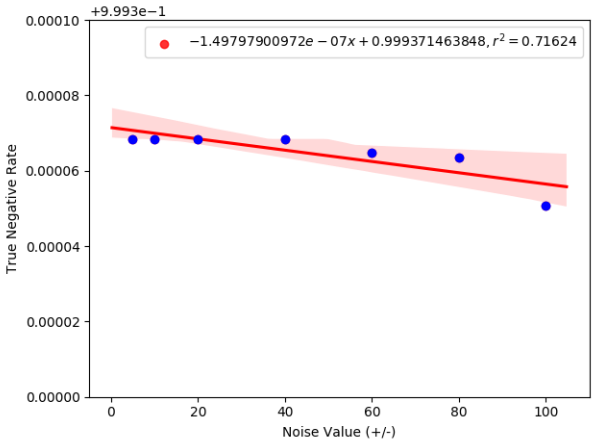


Figure 5.20: Noise value by the true negative rate for the correspondence of filaments segmented from volumes degraded with varying levels of Gaussian noise to the volume's ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

degradation of the filament structure than Gaussian noise in our investigation. In the next section, we explore the effect of the median filter on a noiseless image volume.

5.3.1.2 Quantification of median filter effect on filament structure

To quantify the effect of the median filter on filament structure, we constructed 12 filtered models. Each filtered model is derived from the model presented in Figure 5.2; it is unique in the two-dimensional $n \times n$ median filter applied to it slice-by-slice. The models were constructed for the different median filter neighborhoods of $n = 0, 3, 5, 7, 11, 13, 15, 17, 19, 21, 23, 31$. We then applied the voxel scooping method [9] to each filtered model segment the embedded filaments. We calculated volumetric correspondence of the segmented filaments (i.e., the ER) to the model's GT per the procedure outlined in Section 5.2.1. We report sensitivity (true positive rate), specificity (true negative rate), false positive rate, and false negative rate in Figures 5.26, 5.27, 5.28, and 5.29 respectively. Precision (positive predictive value), Jaccard index, and dice coefficient are reported in Figures 5.30, 5.31, and 5.32 respectively.

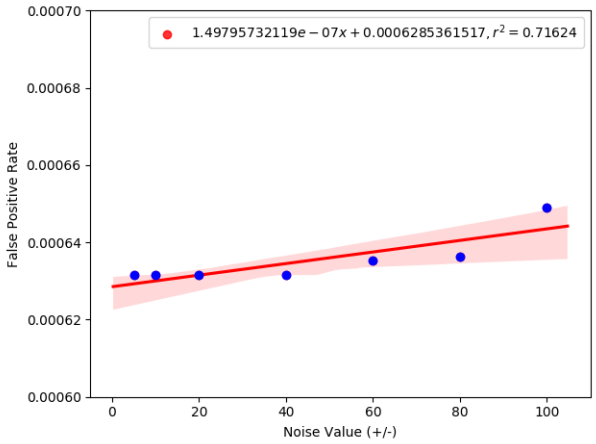


Figure 5.21: Noise value by the false positive rate for the correspondence of filaments segmented from volumes degraded with varying levels of Gaussian noise to the volume's ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

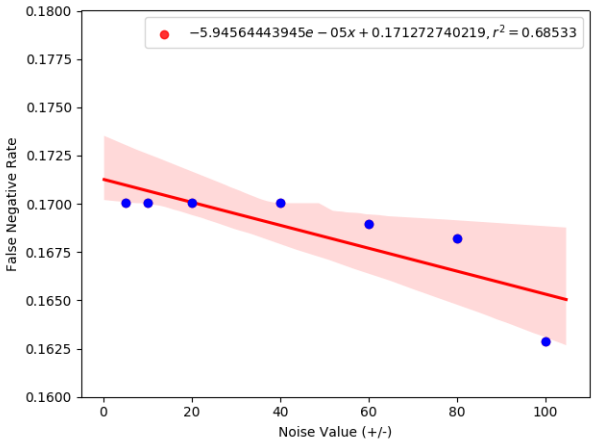


Figure 5.22: Noise value by the false negative rate for the correspondence of filaments segmented from volumes degraded with varying levels of Gaussian noise to the volume's ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

As expected, the 0×0 median filter (i.e., no filter) model ER showed perfect correspondence with the GT, as filament structure was not damaged at all in the ER. Filament structure began to degrade significantly with the application of a non-zero median filter. Simply applying a 3×3 median filter resulted in a reduction of approximately 20% (from 100% down to about 80%) of the pixels labeled correctly. Qualitatively, the degradation is hardly visible along the isosurface, as illustrated in Figure 5.33. However, these small defects accumulate into significant degenerations across the whole, as evident by the significant difference between the ER and GT.

With the larger 7×7 median filter ER, pixel label correctness doesn't change substantially, and is right around 80%. Qualitatively, the aberrations along the isosurface of this ER are only slightly more pronounced. These defects are captured in Figures 5.35 and 5.36. The correspondence between a single slice of the GT and ER is shown in Figure 5.37. As the size of the median filter is increased, the correspondence of the ER to GT continues to diminish. We continue to illustrate the degradation qualitatively (quantitative data already presented in Figures 5.26, 5.27, 5.28, 5.29,

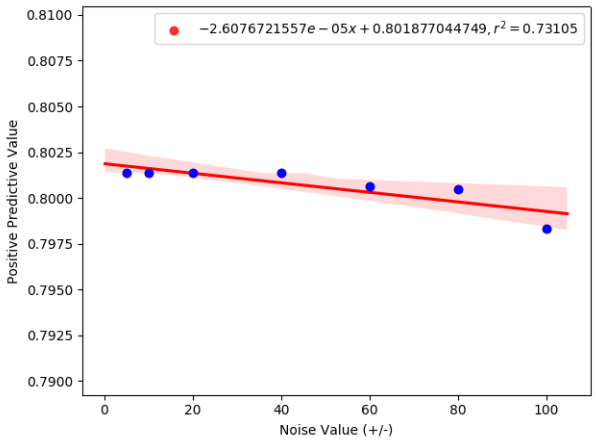


Figure 5.23: Noise value by the precision for the correspondence of filaments segmented from volumes degraded with varying levels of Gaussian noise to the volume's ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

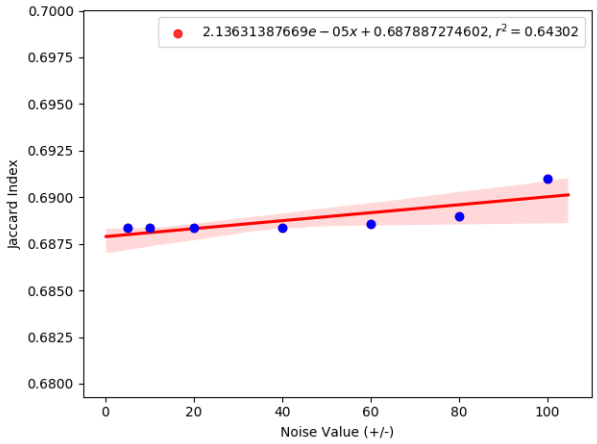


Figure 5.24: Noise value by the Jaccard index for the correspondence of filaments segmented from volumes degraded with varying levels of Gaussian noise to the volume's ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

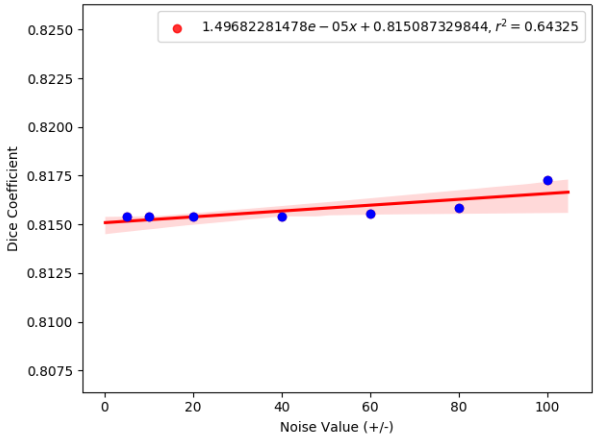


Figure 5.25: Noise value by the dice coefficient for the correspondence of filaments segmented from volumes degraded with varying levels of Gaussian noise to the volume's ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

5.30, 5.31, and 5.32) of the application of a:

- 15×15 median filter in Figures 5.38, 5.39, and 5.40;
- 17×17 median filter in Figures 5.41, 5.42, and 5.43; and,
- 23×23 median filter in Figures 5.44, 5.45, and 5.46; and,

The increase in degradation is observed as the pixels involved in the neighborhood of the median filter becomes more dominated by "background" grayscale values. This effect is most prevalent when the size of the median filter exceeds that of an embedded filament element: the filament element is absorbed into the background and thus "removed" from the image. An $n \times n$ median filter has n^2 pixels involved in the computation of the pixel under consideration's value. Therefore, isolated clusters of filament pixels whose area is less than $n^2/2$ are eliminated from the image (will receive the value of background). In our experiments, as the median filter increased in size, we saw degradation along the surface of filaments in the ER. Once $n^2/2$ exceeded the diameter of isolated filaments, the ER became discontinuous as filaments pieces are lost to the background. It is important to emphasize that even with the smallest median filter, there is a degeneration of the quality of the surface.

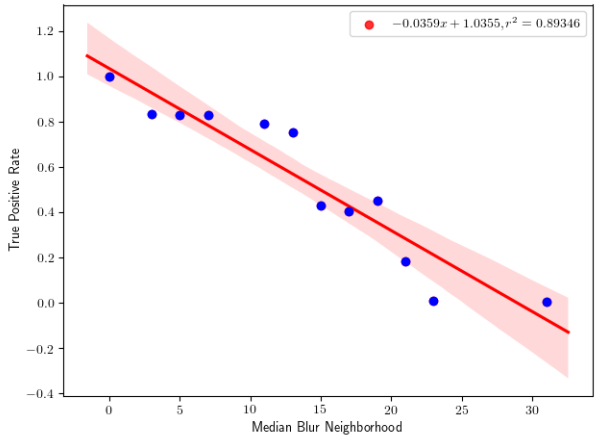


Figure 5.26: Median filter of neighborhood n by true positive rate for the correspondence of filaments segmented from volumes with applied $n \times n$ median filter to the volumes ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

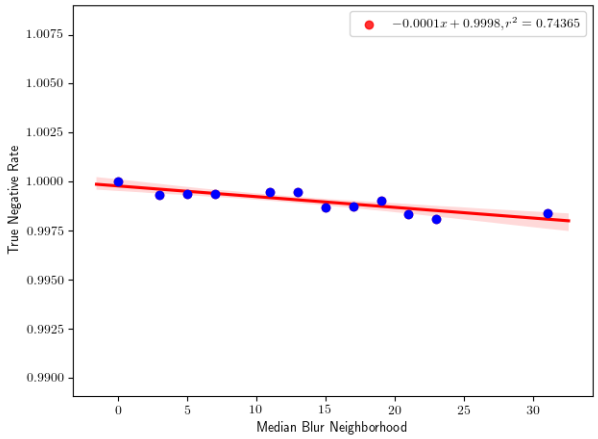


Figure 5.27: Median filter of neighborhood n by true negative rate for the correspondence of filaments segmented from volumes with applied $n \times n$ median filter to the volumes ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

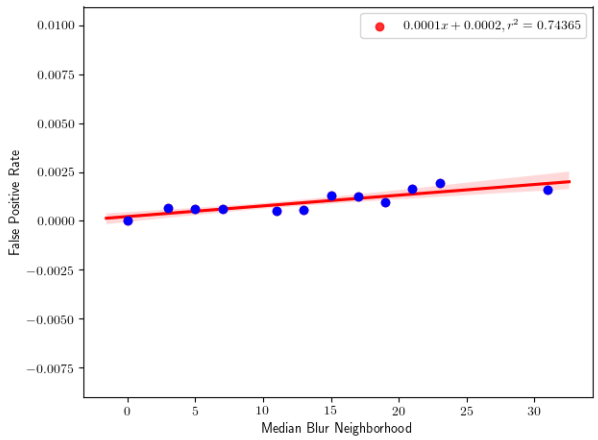


Figure 5.28: Median filter of neighborhood n by false positive rate for the correspondence of filaments segmented from volumes with applied $n \times n$ median filter to the volumes ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

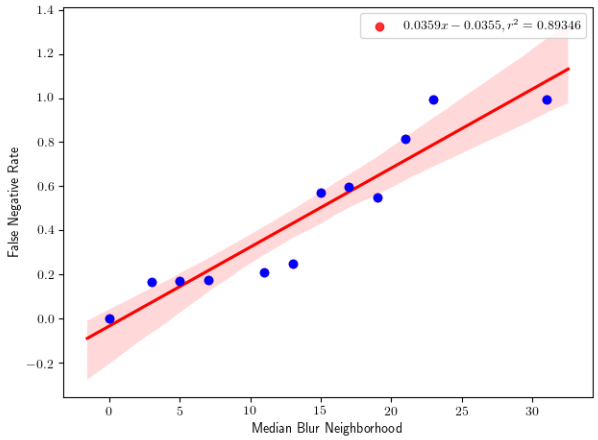


Figure 5.29: Median filter of neighborhood n by false negative rate for the correspondence of filaments segmented from volumes with applied $n \times n$ median filter to the volumes ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

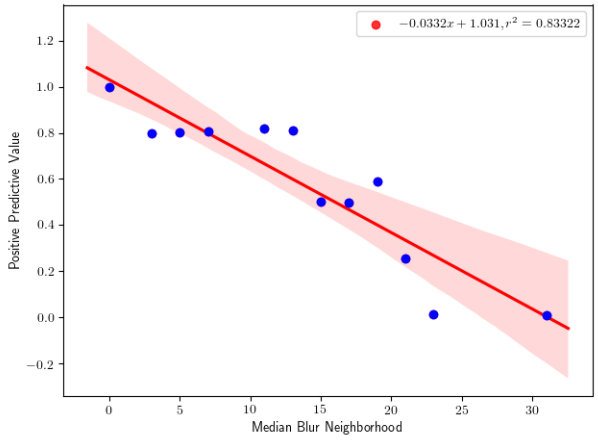


Figure 5.30: Median filter of neighborhood n by positive predictive value for the correspondence of filaments segmented from volumes with applied $n \times n$ median filter to the volumes ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

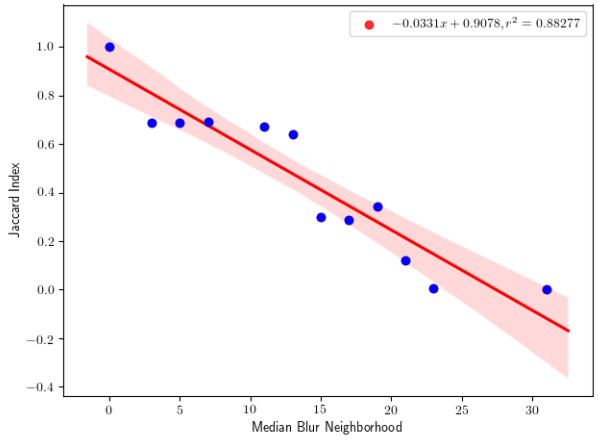


Figure 5.31: Median filter of neighborhood n by Jaccard index for the correspondence of filaments segmented from volumes with applied $n \times n$ median filter to the volumes ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

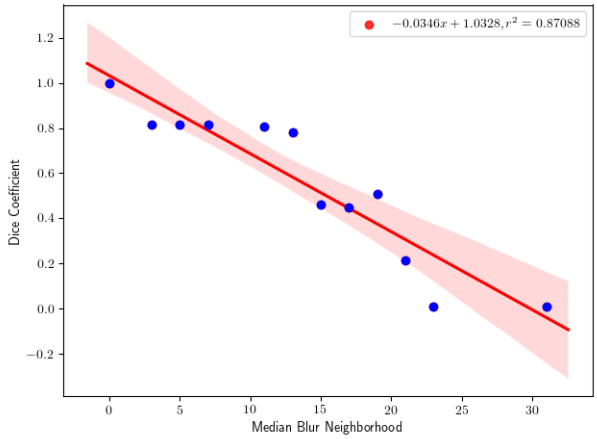


Figure 5.32: Median filter of neighborhood n by dice coefficient for the correspondence of filaments segmented from volumes with applied $n \times n$ median filter to the volumes ground truth. The linear regression line fit to the data presented in this plot is drawn in red; the equation for that fitted line, along with its R-squared value, is boxed within the graph.

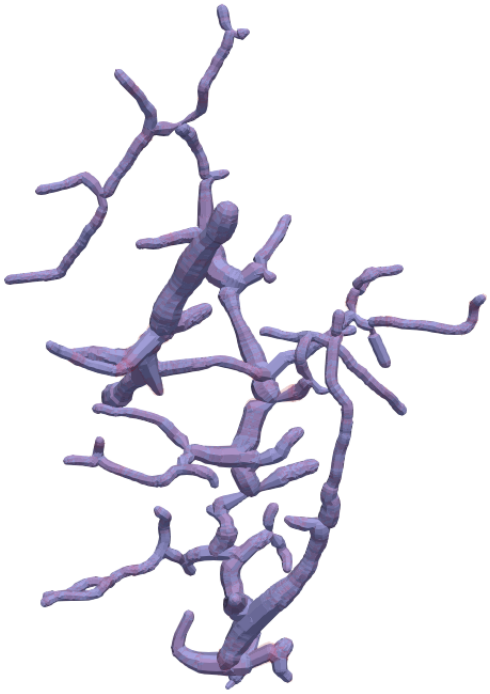


Figure 5.33: Surface of 3×3 median filtered ER (in purple), overlaid atop the GT, which presents in low opacitiy red as it "bleeds" through the ER in areas of lack of correspondence.

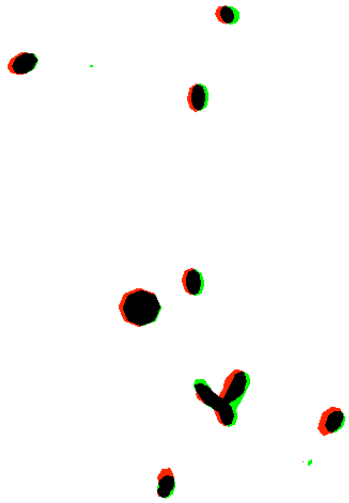


Figure 5.34: Diff image of the same slice acquired from the middle of the 3×3 median filtered ER and the GT. Black pixels are common to both the ER and GT slice; green pixels are in the GT but not the ER; red pixels are in the ER but not the GT.

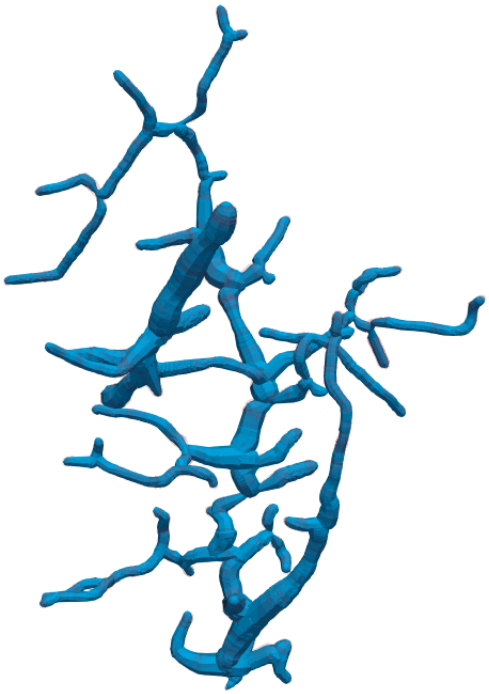


Figure 5.35: Surface of 7×7 median filtered ER (in blue), overlaid atop the GT, which presents in low opacity red as it "bleeds" through the ER in areas of lack of correspondence.

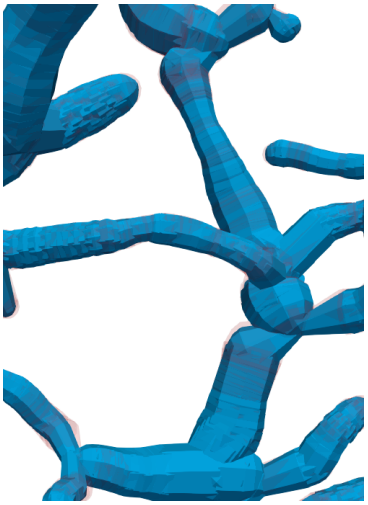


Figure 5.36: Close-up of Figure 5.35.

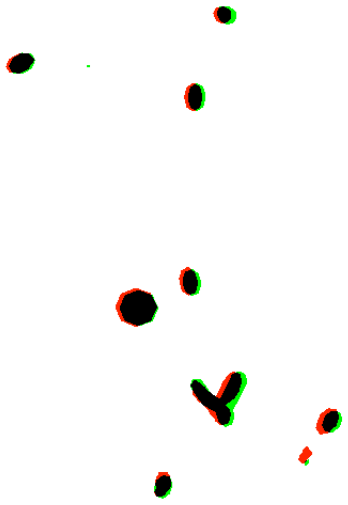


Figure 5.37: Diff image of the same slice acquired from the middle of the 7×7 median filtered ER and the GT. Black pixels are common to both the ER and GT slice; green pixels are in the GT but not the ER; red pixels are in the ER but not the GT.

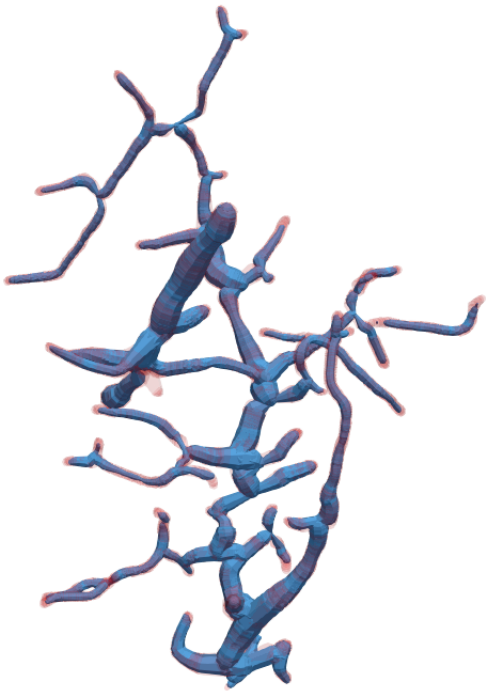


Figure 5.38: Surface of 15×15 median filtered ER (in blue), overlaid atop the GT, which presents in low opacitiy red as it "bleeds" through the ER in areas of lack of correspondence.

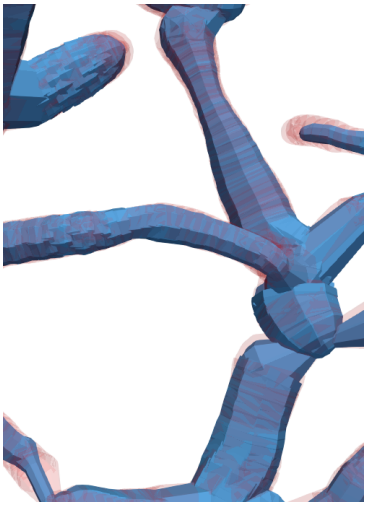


Figure 5.39: Close-up of Figure 5.38.

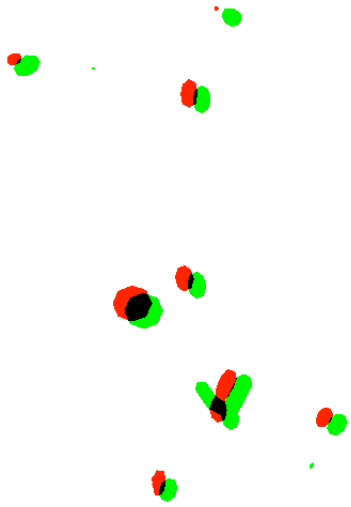


Figure 5.40: Diff image of the same slice acquired from the middle of the 15×15 median filtered ER and the GT. Black pixels are common to both the ER and GT slice; green pixels are in the GT but not the ER; red pixels are in the ER but not the GT.

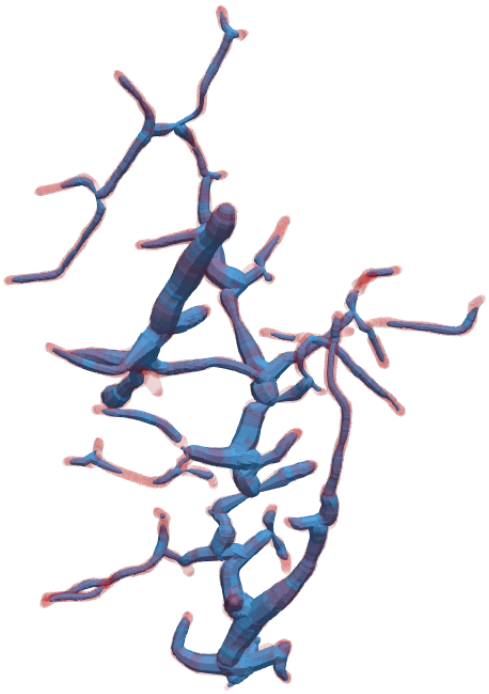


Figure 5.41: Surface of 17×17 median filtered ER (in blue), overlaid atop the GT, which presents in low opacitiy red as it "bleeds" through the ER in areas of lack of correspondence.

Figure 5.42: Close-up of Figure 5.41.

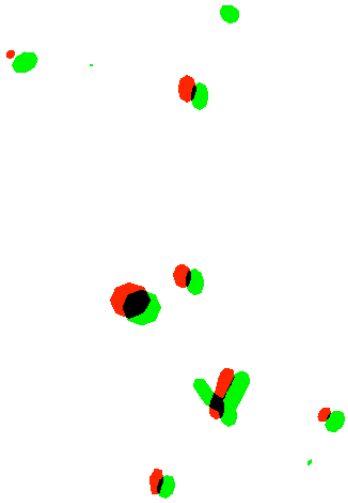


Figure 5.43: Diff image of the same slice acquired from the middle of the 17×17 median filtered ER and the GT. Black pixels are common to both the ER and GT slice; green pixels are in the GT but not the ER; red pixels are in the ER but not the GT.

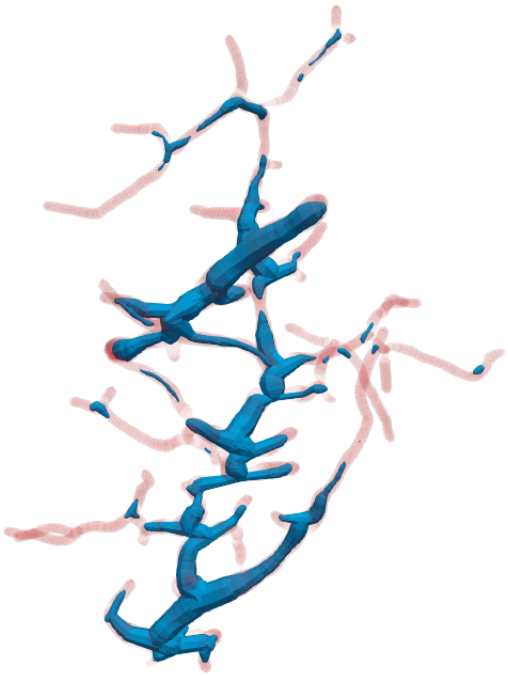


Figure 5.44: Surface of 23×23 median filtered ER (in blue), overlaid atop the GT, which presents in low opacitiy red as it "bleeds" through the ER in areas of lack of correspondence.

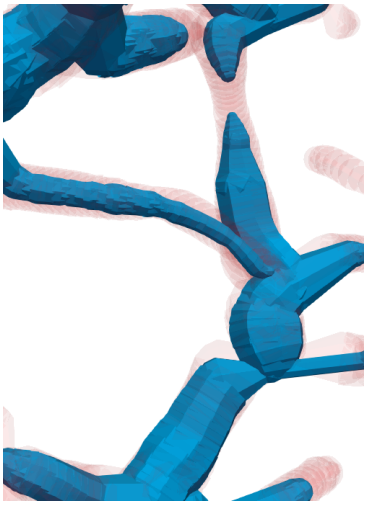


Figure 5.45: Close-up of Figure 5.44.

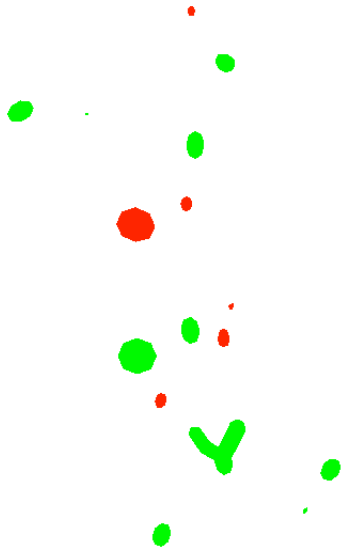


Figure 5.46: Diff image of the same slice acquired from the middle of the 23×23 median filtered ER and the GT. Black pixels are common to both the ER and GT slice; green pixels are in the GT but not the ER; red pixels are in the ER but not the GT.

5.3.2 Centerline correspondence

5.3.2.1 Digital phantom experiment

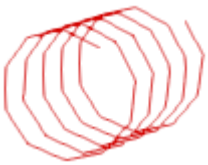


Figure 5.47: Centerlines of spiral pattern model used to illustrate our procedure.

Figure 5.48: Tubular surface fit to centerlines in Figure 5.47 with constant radius.

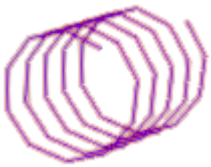


Figure 5.49: Centerlines of spiral pattern model described in Figure 5.47 overlaid with its surface as shown in Figure 5.48.

We elected to use a $40\,\text{pixel} \times 40\,\text{pixel} \times 70\,\text{pixel}$ spiral pattern to illustrate our method detailed in Section 5.2.2; the validation techniques illustrated here are directly applicable to more complicated, and more vascular-like, geometries without modification. To construct the spiral geometry from which surface, skeleton, and image volumes used in this experiment were derived, we used The Visualization Toolkit 8.0 (vtk) [103] in conjunction with the Python 2.7 programming language to draw a spiral using geometric primitives. The first step of our procedure was to calculate the three-dimensional points and connect those points through line segments. We used the following algorithm to calculate the point sequence necessary to construct a spiral geometry:

Algorithm 1 Generating spiral coordinate points

Input: Number of vertices, cycles, radius of spiral, height, and number of sides
Output: Coordinates for line describing the spiral provided as input

1: pts = { }
2: **for** $i = 1$ to number of verticies **do**
3: calculate vx, vy, vz
4: pts += $\{i, vx, vy, vz\}$
5: **end for**
6: **return** pts

where

$$vx = rSpiral * cos(2 * \pi * nCycles * i/(nVertices - 1)) \tag{5.7}$$

$$vy = rSpiral * sin(2 * \pi * nCycles * i/(nVertices - 1)) \tag{5.8}$$

$$vz = height * i/nVertices \tag{5.9}$$

for $nVerticies$ number of vertices, $rSpiral$ radius, $height$ height, and i vertex number composing the coordinated spiral.

This series of points, connected through line segments, comprised the skeleton of the spiral geometry, and this was written out in vector graphics format. We did this using a Serial vtkPolyData (vtp) [103] formatted file due to our familiarity with its encoding scheme. Specifically, this file served as the centerline GT and is shown in Figure 5.47. About the centerline GT, we fit a tube of fixed radius to represent a vascular-like cylindrical structure and wrote the geometrical data describing the resultant polydata surface to a vtp formatted file. The result of this process is shown in Figure 5.48. Finally, we then constructed a volumetric imaging file containing this vessel-like geometry encoded. We wrote the volumetric imaging file out in raster graphics format. Specifically, we chose the meta image format (MDH), due to our experience with it. The surface overlaid with the skeleton can be seen in Figure 5.49.

The ERs are presented in Figures 5.50 in blue alongside the ground truth centerlines shown in green. A close-up detailing the deviation more profoundly is presented in Figure 5.51. In Figure5.52 we present the representation of the ground truth centerline as series of nodes connected by edges. We show the same for the ERs Figure5.53.

Using our centerline correspondence procedure detailed in Section 5.2.2, we found that on average, each node comprising the ER centerline was $0.129{,}4\,\mathrm{pixels}$ from the ground truth centerline. The nodes in the ER exceeding that value are shown in Figure 5.54.

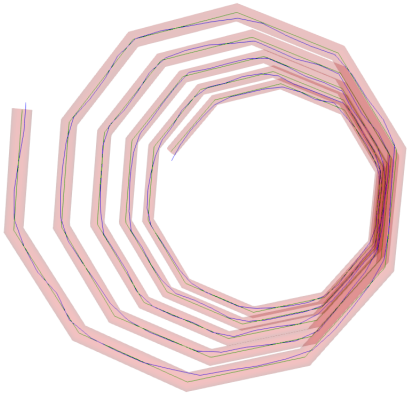


Figure 5.50: Centerline determined by superellipsoid tracing algorithm upon application to image volume containing the spiral geometry described by the surface in Figure5.48 shown in blue against the ground truth data shown in green.

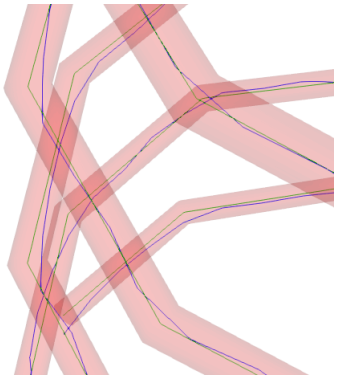


Figure 5.51: Close-up of the centerline determined by superellipsoid tracing algorithm upon application to image volume containing the spiral geometry described by the surface in Figure5.48 shown in blue against the ground truth data shown in green.

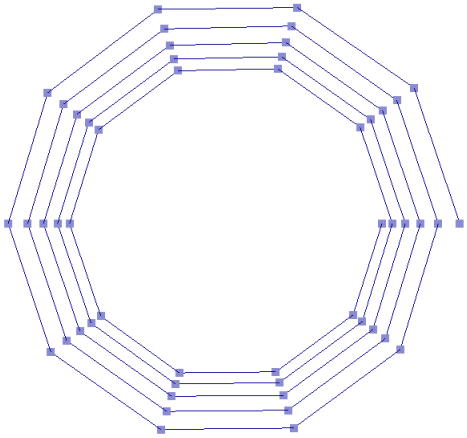


Figure 5.52: Graph representation of the spiral geometry skeleton described in Figure 5.47.

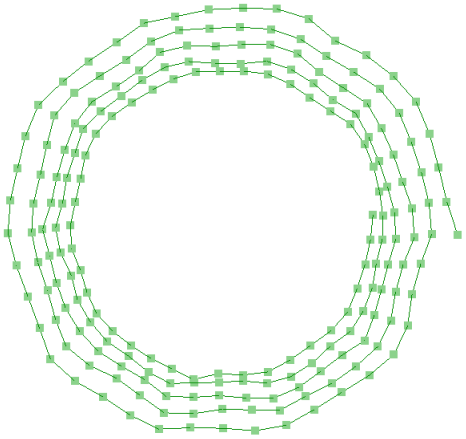


Figure 5.53: Graph representation of the centerline determined by superellipsoid tracing algorithm upon application to image volume containing the spiral geometry described in green in Figure 5.50.

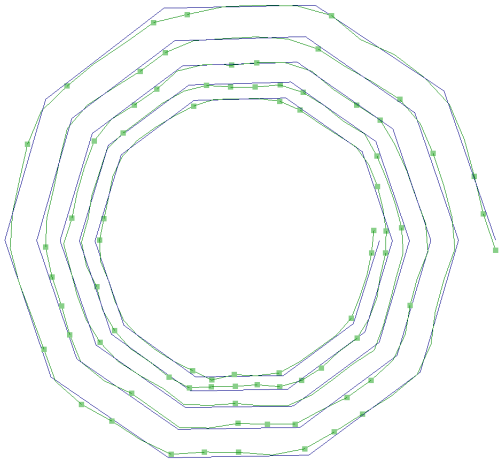


Figure 5.54: The nodes (as green squares) composing the ER (shown in green) that exceeded $0.129{,}4\,\text{pixels}$ from the ground truth centerline (shown in blue).

5.3.2.2 Data-driven model experiment

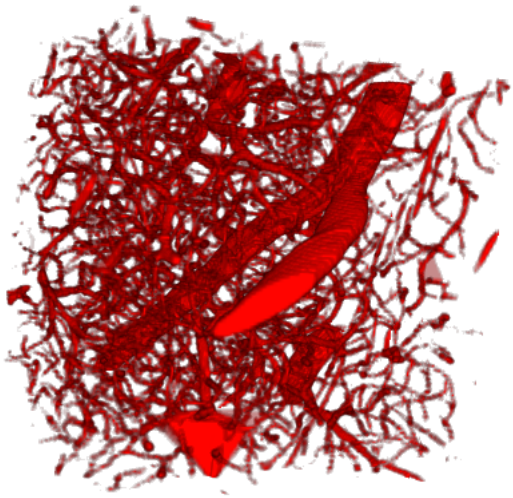


Figure 5.55: Sample of C57BL/6J mouse cerebrovasculature obtained from the KESMBA India Ink whole-brain dataset for use in this study.

We isolated an arbitrary region of cerebrovasculature – $256\,\text{pixel} \times 256\,\text{pixel} \times 256\,\text{pixel}$ at a voxel resolution $1.2\,\mu\text{m} \times 1.4\,\mu\text{m} \times 1.0\,\mu\text{m}$ – from the KESMBA C57BL/6J whole-brain India Ink dataset. We subsequently composed a volume from the obtained image slices (Figure 5.55) and

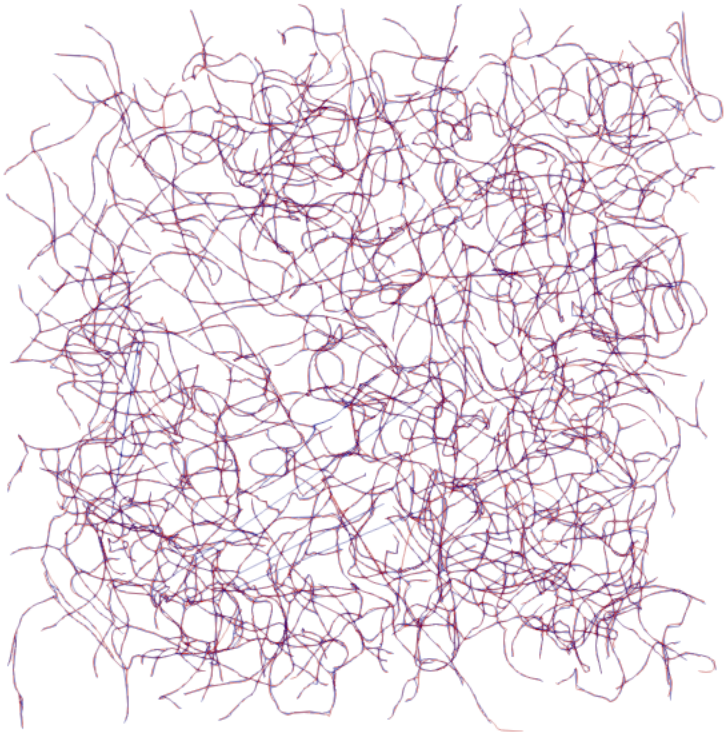


Figure 5.56: Overlay of the ground truth centerline (red) and the algorithm centerline (blue) for the reconstructed volume. These centerlines are bound within a of $256\,\text{pixel} \times 256\,\text{pixel} \times 256\,\text{pixel}$ volume.

wrote out this "imaging volume" in a raster graphics format. We reconstructed constructed a data-driven model from this imaging volume using our method described in [101], a precursor to that detailed in Chapter 4. Once we derived the imaging volume along with the ground truth centerline data, we applied the superellipsoid tracing method [76] from the Farsight Tool Kit v0.4.4 to extract the centerline of the embedded filaments. We did not fine-tune any of the tracing parameters; Farsight's default values were used.

We compare the acquired ER centerline to the GT centerline using our method detailed in Section 5.2.2. We present these centerlines overlaid in Figure 5.56. For each node in the ER graph, we calculated its deviation for a corresponding node in the GT accordingly. We took the average of this value over all nodes and found an average deviation to be approximately $2.45\,\mu\text{m}$.

5.3.3 Summary

The primary contribution of this work is our model-based validation system that consolidates metrics of correspondence for the quantitative evaluation of image processing techniques and segmentation algorithm accuracy in large-scale datasets composed of embedded filaments. Our validation system provides the capability to quantify the volumetric agreement between an ER and corresponding GT of the model used in that experiment. Furthermore, our model-based validation system provides the capability to quantitatively analyze centerline deviation as the distance between ER and GT.

The introduction of our system is a critical step towards easier quantitative analyses into the effect of different image processing regiments, or the accuracy of segmentation algorithms, applied to large volumes of imaging data. Moreover, our centerline graph-based approach will provide enhanced utility in predicting the accuracy on high-resolution, high-volume cerebrovasculature data. For example, a correspondence measure based on voxel deviation works well for digital phantoms composed of simple geometries or with small volumes of expert-labeled ground truths. It's easy to assess where the algorithm went astray qualitatively. We make the argument for the use of more complex geometries as validation sets that represent aspects of the brain vasculature in Chapters 3 and 4 and in [11, 101]. When dense synthetic datasets are applied, it is difficult to assess the discrepancies between the segmentation result and ground truth occur.

Our contribution is not only a method that reports centerline deviation but one that also collects the exact locations where the segmentation algorithm's performance deteriorates. The geometry and topology at such areas can then be examined and the algorithm revised accordingly. However, we do not believe that this approach is sufficient on its own. Understanding the true correspondence at a voxel-based level between an ER and GT is extremely important. We previously discussed in Section 2.1 how small changes in morphology can have a profound effect on cerebrovasculature studies. Therefore, we believe that a multi-faceted approach, using both volumetric and centerline correspondence, is important and that our model-based validation system provides the foundation for such an approach.

6. INFORMATICS PLATFORM

6.1 Motivation

There are many high-throughput LM modalities capable of complete imaging of organs of small animals at sub-micron resolution (review, Section 2.2). In cerebrovasculature research, the number of and diversity in whole-brain filament acquisition protocols is growing each day. Common to all of these techniques is that the acquisitions are large in volume, high in resolution (sub-micrometer), and multi-scale by nature (spanning from cellular-scale to whole-brain scale). For cerebrovasculature datasets, these characteristics make the quantitative analysis of the embedded filaments challenging.

Whole-brain datasets have become increasingly easier to acquire with the introduction of new LM techniques. However, lagging behind is the development of informatics platforms capable of analyzing the acquired image volumes. Recently, research has started to shift away from acquisition modalities and towards quantitative analysis. For cerebrovasculature data, many groups go about the extraction process in diverse ways (see Section 2.3) and report different attributes about the filaments and characteristics about the volume (see Section 2.4). Due to these various directions in processing and analysis, a standardized means of quantitative analysis for whole-brain cerebrovasculature datasets has yet to emerge. Given the variety in acquisition protocols for complete cerebrovasculature systems, an informatics platform that can be applied to data acquired across an array of LM modalities is required.

The development of such an informatics framework would help elucidate filament attributes across the whole-brain, whether these attributes vary depending on the cerebral region, and how they differ between physiological and pathophysiological states. The structural and functional integrity of the brain critically depends on the interdependence between neuronal activity and vascular dynamics, which ensures that adequate oxygen and nutrients are delivered to neurons by matching local blood flow to metabolic demand [14]. Disruptions to the integrity of the cere-

brovasculature architecture can have a profound impact on the cellular constituents of the brain, which can lead to cognitive decline and a diminished quality of life [14]. In fact, the pathogenesis of many acute conditions (such as ischemic stroke) and chronic illnesses (including Alzheimer's and Parkinson's) are associated with changes to the cerebral microcirculation.

To better understand the processes underlying healthy and diseased brains, there is a critical need to foster an accessible and standardized open-source informatics framework that facilitates the quantitative analysis of the cerebrovascular system. In the current chapter, we develop an open-source informatics framework that can be applied to data acquired across an array of LM modalities. Our method focuses on the utilization of a graph-based representation of vascular topology and geometry. The first part of this chapter covers the construction of such a graph. Thereafter, we detail procedures that calculate vascular volume characteristics across the whole brain and the attributes of the individual vascular segment comprising it. Finally, we present two case studies that illustrate the utility of our informatics platform.

6.2 Methodology

6.2.1 Graph construction

Before the graph can be constructed, we require that the filaments are extracted from the imaging volume and encoded in a format that stores the filaments as a series of nodes and edges. We selected the *SWC* format for this encoding [102] and elaborate on this scheme in Section 4.2.1. In short, the idea behind this format is that each vascular segment (vessel between two bifurcations) can be embedded as a series of line segments. The filaments are stored in the *SWC* format as a hierarchy of line segments for every connected vascular component in the dataset.

In studies of the brain's vascular system, graph representations of vessel connectivity are used to analyze the underlying structural properties of its networks [92, 28, 29, 91, 96, 101]. This description G is composed of a set of nodes (i.e., vertices) V and a set describing the relationships between the nodes (i.e., edges) E. In our representation, we define a vessel segment as a sequence of interconnected edges by nodes of degree two presenting between two nodes of degree at least

three (i.e., a bifurcation point) or exactly one. We extend this topological encoding by associating geometrical properties with the vertices and edges to create a more biologically meaningful description of the vessel segments comprising the vascular network.

To construct a graph representation of a filament network, we construct nodes and edges from the hierarchy of line segments stored in the *SWC* file. We do this by iterating over the nodes in the *SWC* file that comprise the vascular filaments. During this process, we store each respective endpoint's unique ID along with its coordinates and radius in a "container of nodes". While iterating over an endpoint, we also establish a relationship between its unique ID and that of its parents. We store this connectivity information in a "container of edges". Once the nodes and edges are realized, there are many different graph libraries that handle the encoding of nodes and edges into graph form. We elected to construct a *NetworkX* [108] graph primarily due to our familiarity with it and due to the breadth of graph algorithms that it provides.

To build the *NetworkX* graph, we begin by initializing a new *NetworkX* graph. Given the bidirectional nature of blood flow through the cerebral arterial [31] and venous circulation [16], we model undirected relationships between nodes in the graph model. We add into the graph our "container of nodes", along with the edges in our "container of edges". If desired, during this process, we can use segment radius [91] or resistance [17] to discriminate between capillaries and non-capillaries. Each node can be labeled before it is entered into the graph, thereby estimating the separation between the tree-like arterioles [17] and venules from the mesh-like capillary network [29]. Our graph construction process results in a *NetworkX* graph that encodes both the topology and geometry of the vascular data.

6.2.2 Metrics reported

Our informatics framework will report an array of attributes for each vessel segment and general characteristics of a complete network. The analysis commences on filament data stored in graph representations in full or upon a subgraph obtained from it for a region of interest. A subgraph can be obtained using traversal operations such as breadth-first and depth-first search or through the application of a more sophisticated graph algorithm that identifies a subgraph satisfy-

ing some constraints

6.2.2.1 Vascular volume characteristics

For the graph representation (or subgraph) under consideration, we report the following vascular volume characteristics:

- Total length
- Total surface area
- Total volume

To perform these calculations, we iterate over every edge `e` in our graph `G`. We look up the two nodes, `n1` and `n2`, that are connected through `e` in `G`. The Euclidean distance `length` between `n1` and `n2` is used to calculate length and is added to an accumulator for length across the edges. Subsequently, we imagine the edge as a truncated cone, with circles of radius `n1.radius` and `n2.radius` presenting at either end, separated by some distance `length`, and connected through a trapezoid funnel. We diagram this structure in Figure 6.1.

In order to calculate the surface area of the imagined truncated cone, we flattened it into three geometric primitives, two circles, and a trapezoid, as shown in Figure 6.2. We subsequently calculate the circumference of the base circles, and multiply their average by the distance `length` between them.

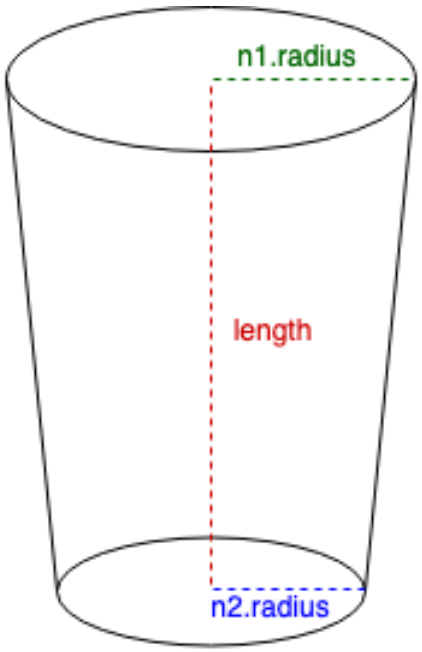


Figure 6.1: Truncated cone used to model an edge e in graph G during the calculation of vascular volume characteristics across e. As illustrated in this figure, we imagine the edge as a truncated cone, with circles of radius `n1.radius` and `n2.radius` presenting at either end, separated by some distance `length`, and connected through a trapezoid funnel.

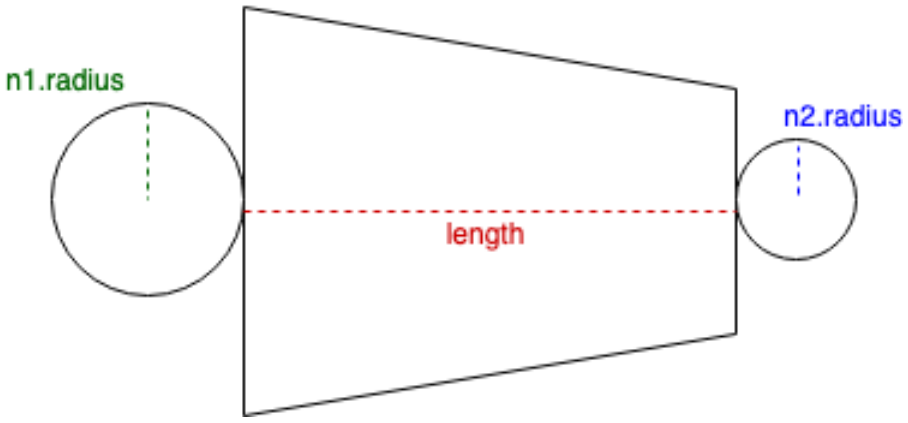


Figure 6.2: Truncated cone presented in Figure 6.1 flattened into three geometric primitives, two circles and a trapezoid.

Algorithm 2 Vascular volume characteristics

Input: Graph `graph`
Output: Calculation of vascular volume characteristics

Let `total_length` be 0
Let `total_surface_area` be 0
Let `total_volume` be 0
for each edge `e` in graph `G` **do**
    Let `n1` and `n2` be the nodes comprising `e`
    Let `length` be the Euclidean distance between `n1` and `n2`
    Increment `total_length` by `length`
    Let `circC1` be $2\pi \times$ `n1.radius`
    Let `circC2` be $2\pi \times$ `n2.radius`
    Let `surface_area` be $(\texttt{circC1} + \texttt{circC2}) \div 2 \times \texttt{length}$
    Increment `total_surface_area` by `surface_area`
    Let `volume` be $\pi \times ((\texttt{n1.radius} + \texttt{n2.radius}) \div 2)^2 \times \texttt{length}$
    Increment `total_volume` by `volume`
end for
return `total_length`, `total_surface_area`, `total_volume`, `total_surface_area` $\div$ `total_volume`

6.2.2.2 Vessel segments attributes

Our framework will report the following attributes for each vessel segment present in the graph representation (or subgraph) under analysis:

- Radius
- Length
- Tortuosity, using the distance metric defined in [87]
- Distance (between beginning and terminal end of segment)
- Surface area
- Volume
- Surface-area-to-volume ratio (i.e., SA:V)
- Length-to-diameter ratio (i.e., LEN:DIA)

In order to compute the attributes of vessel segments, we create a copy of the graph constructed in section 6.2.1 and reduce it such that only topological information for points of bifurcation and endpoints remain (we call this graph the "reduced graph"). This process is carried out per Algorithm 3. In short, this procedure isolates each vessel segment. To perform this operation, we pick an arbitrary node of degree 2 from the graph, bag it, and then visit its neighbors in either direction, bagging each until we arrive at a point of bifurcation or a node of degree 1 in both directions. Let these two "endpoint" nodes be called `end1` and `end2`. At this point, we have determined all of the nodes comprising one vessel segment.

We calculate the attributes of that respective vessel segment using `end1` and `end2` along with the nodes and edges presenting between them. The radius is taken simply the average radius across all of the nodes comprising the vessel segment. Length, surface area, and volume are calculated using the same procedure used for the volume (described in Section 6.2.2.1). The only difference

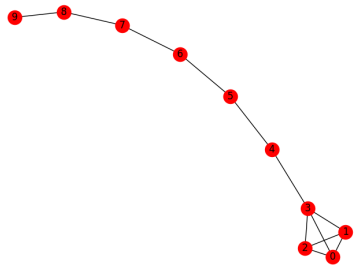


Figure 6.3: Lollipop graph used as input to Algorithm 3 to illustrate graph reduction.

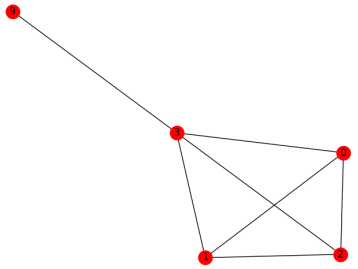


Figure 6.4: Output of Algorithm 3 after being applied to lollipop graph shown in Figure 6.3.

is that instead of using all of the edges in the graph, we're considering only those for a respective vessel segment. Surface to volume ratio and length to diameter ratio are calculated as ratios are. Tortuosity for the vessel segment is calculated as the length divided by the Euclidean distance between `end1` and `end2`. If the vessel segment forms a cycle, the denominator will be zero, and tortuosity will be reported as *inf* (infinity).

After calculating the attributes of a vessel segment, we connect the two-node endpoints, `end1` and `end2`, through an edge in the reduced graph and assign that edge a unique index for the vessel segment. We then store calculated attributes for that vessel segment in the new edge and subsequently eliminate the transition nodes between `end1` and `end2` to reduce the graph. In the original graph, we subsequently label all edges within that segment with this unique index. This creates a mapping between the reduced graph and the original graph. We continue this process for each vessel segment. At the end of this procedure, only edges at bifurcations and endpoints remain in the reduced graph [91]. We show the application of our graph reduction algorithm (Algorithm 3) applied to the lollipop graph shown in Figure 6.3 in Figure 6.4. The beauty of our procedure is that each edge in the reduced graph represents a uniquely identifiable vessel segment with that segment's attributes. Using a vessel segment's unique index label in the reduced graph, we can easily isolate the nodes and edges comprising that segment from the original graph, allowing that segment to be easily reconstructed, should that be desired.

Algorithm 3 Graph reduction and vessel segment attributes calculation algorithm

Input: Graph G
Output: Graph R reduced

Let R be a copy of G
for each node n in R **do**
  if degree of n is 2 or 1 **then**
    determined adjacent nodes $n1$ and $n2$ to n in R
    while degree of $n1$ is 2 **do**
      bag edge $e1$ between $n1$ and its predecessor
      assign $n1$ its adjacent node, moving away from n
    end while
    while degree of $n2$ is 2 **do**
      bag edge $e2$ between $n2$ and its predecessor
      assign $n2$ its adjacent node, moving away from n
    end while
    calculate attributes for each edge in the bag
    let *id* be a unique ID for isolated vessel segment
    connect $n1$ and $n2$ in R with an edge e; label e with *id*
    store accumulated vessel segment attributes in edge e
    label all corresponding edges of bag in G with *id*
    remove all nodes between $n1$ and $n2$ from R
  end if
end for
return G, R

6.2.3 Open-source dissemination

The source code for our informatics platform is published in a Github repository available via www.github.com/michaelrnowak.

6.3 Results and analysis

6.3.1 Case study #1, digital phantoms

We will first validate our informatics platform using six digital phantoms. These phantoms have been developed to capture different junctions that might present in the graphs derived from biological datasets. Therefore, some of our phantoms include aspects such as cycles and bridges.

The first digital phantom that we use is presented in Figure 6.5. This phantom exhibits two branches that are connected through a bridge. The result of the graph construction process on

this volume, per our procedure detailed in Section 6.2.1, is shown in Figure 6.6. We present the location of these nodes in the phantom volume in Table 6.1, along with the radius measurements at those node positions. The application of graph reduction to this "original" graph (i.e., that in Figure 6.6) is shown in Figure 6.7. Each edge in our "reduced" graph is labeled with a vessel segment number that maps to edges composing that vessel segment in the "updated original" graph (shown in Figure 6.8). This allows for the easy reconstruction of the different vessel segments. For instance, to reconstruct the vessel segment denoted by edge 1 in the "reduced" graph (i.e., 6.7), we would derive all edges labeled with 1 from the original graph (i.e., Figure 6.8).

The attributes (as defined in Section 6.2.2.2) for each vessel segment (i.e., vessel arising between branch points/endpoints) in the digital phantom are reported in Table 6.2. The overall characteristics (as defined in Section 6.2.2.1) for the vascular volume follow: total volume, $188.20\,\text{pixel}^3$; total surface area, $376.40\,\text{pixel}^2$; and total length, $59.91\,\text{pixel}$.

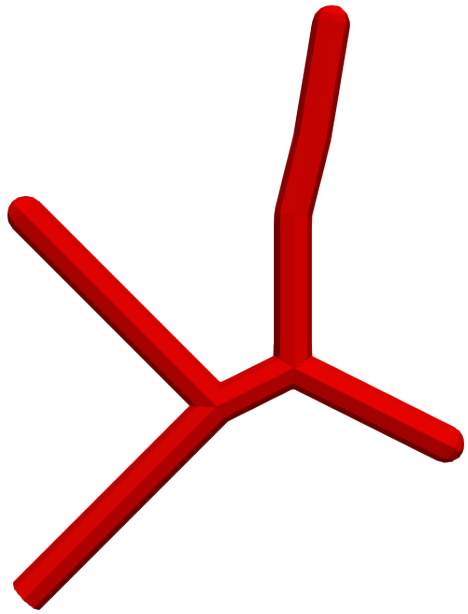


Figure 6.5: Phantom 1 volume.

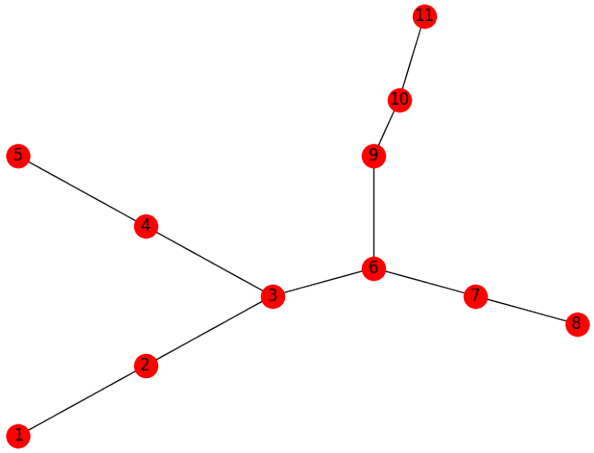


Figure 6.6: Graph representation of phantom 1 volume (Figure 6.5).

Table 6.1: Correspondence of nodes in the graph representation of phantom 1 (Figure 6.6) to their location in the volume (Figure 6.5), along with the radius measurement at that position.

Node No.	***x***	***y***	***z***	**Radius** (pixel)
1	-10	-10	0	1
2	-5	-5	0	1
3	0	0	0	1
4	-5	5	0	1
5	-10	10	0	1
6	4	2	0	1
7	8	0	0	1
8	12	-2	0	1
9	4	10	0	1
10	5	14	0	1
11	6	20	0	1

Table 6.2: Vessel segment attributes (defined in Section 6.2.2.2) calculated across the unique segments composing phantom 1.

Segment ID	**Length** (pixel)	**Surface Area** (pixel^2)	**Volume** (pixel^3)	**Distance** (pixel)	**SA:V**	**LEN:DIA**	**Tortuosity**
1	14.14	88.86	44.43	14.14	2.00	7.07	1.00
2	14.14	88.86	44.43	14.14	2.00	7.07	1.00
3	8.94	56.20	28.10	8.94	2.00	4.47	1.00
4	18.21	114.39	57.20	18.11	2.00	9.10	1.01
5	4.47	28.10	14.05	4.47	2.00	2.24	1.00

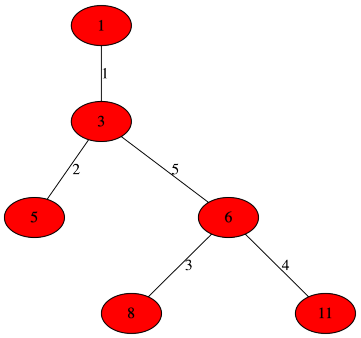


Figure 6.7: "Reduced" graph representation of phantom 1 (Figure 6.6) obtained by algorithm 3. The labels on each edge serve as a unique index for a respective vessel segment presenting in the volume (Figure 6.5).

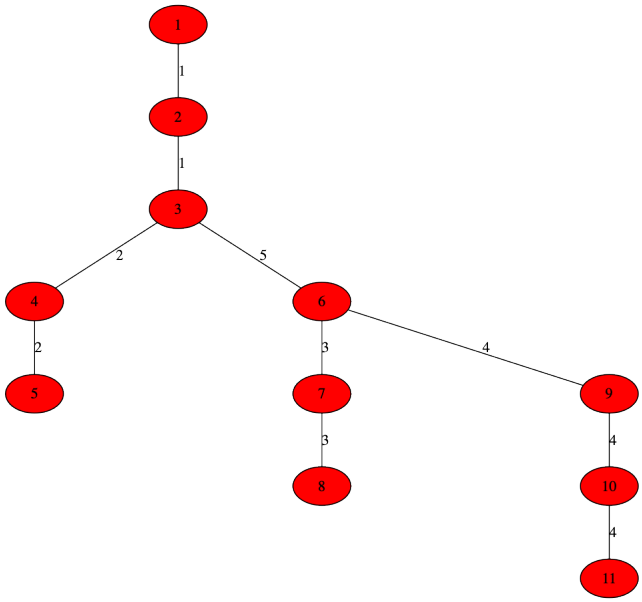


Figure 6.8: Graph representation of phantom 1 (Figure 6.6) updated to contain labeled edges denoting the unique index of the vessel segment that each serves as a filament in.

The second phantom used is also branch-like and is nearly identical to the first phantom (Figure 6.5). The second phantom exhibits two branches that are connected through a bridge; however, unlike phantom one, this model has variable radius measures (not all the same). We present this phantom in Figure 6.9. The result of the graph construction process is shown in Figure 6.10. We present the location of these nodes in the phantom volume in Table 6.3, along with the radius measurements at those node positions. The application of graph reduction to the "original" graph is shown in Figure 6.11. The "updated original" graph is shown in Figure 6.12.

The attributes (as defined in Section 6.2.2.2) for each vessel segment in the digital phantom are reported in Table 6.4. The overall characteristics (as defined in Section 6.2.2.1) for the vascular volume follow: total volume, $617.03\,\text{pixel}^3$; total surface area, $675.22\,\text{pixel}^2$; and total length, $59.91\,\text{pixel}$.

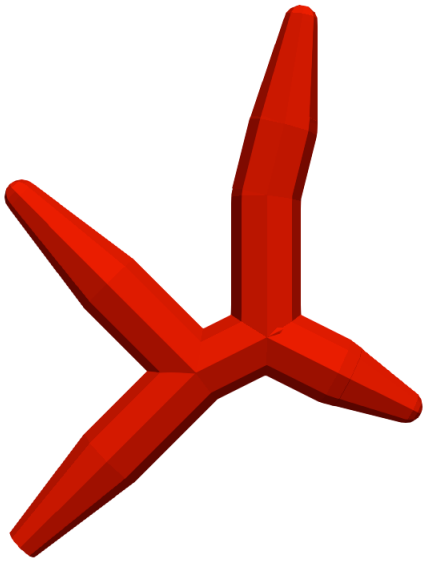


Figure 6.9: Phantom 2 volume.

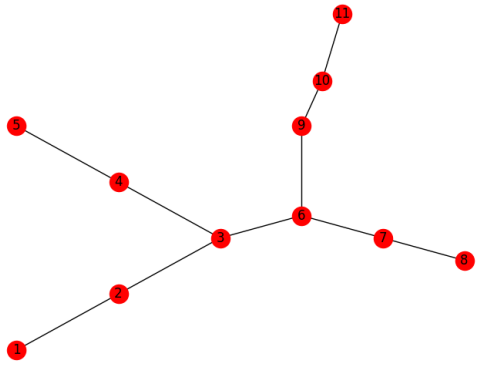


Figure 6.10: Graph representation of phantom 2 volume (Figure 6.9).

Table 6.3: Correspondence of nodes in the graph representation of phantom 2 (Figure 6.6) to their location in the volume (Figure 6.9), along with the radius measurement at that position.

Node No.	***x***	***y***	***z***	**Radius** (pixel)
1	-10	-10	0	1
2	-5	-5	0	2
3	0	0	0	2
4	-5	5	0	2
5	-10	10	0	1
6	4	2	0	2
7	8	0	0	2
8	12	-2	0	1
9	4	10	0	2
10	5	14	0	2
11	6	20	0	1

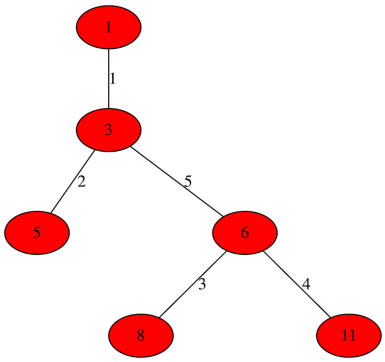


Figure 6.11: "Reduced" graph representation of phantom 2 (Figure 6.10) obtained by algorithm 3. The labels on each edge serve as a unique index for a respective vessel segment presenting in the volume (Figure 6.9).

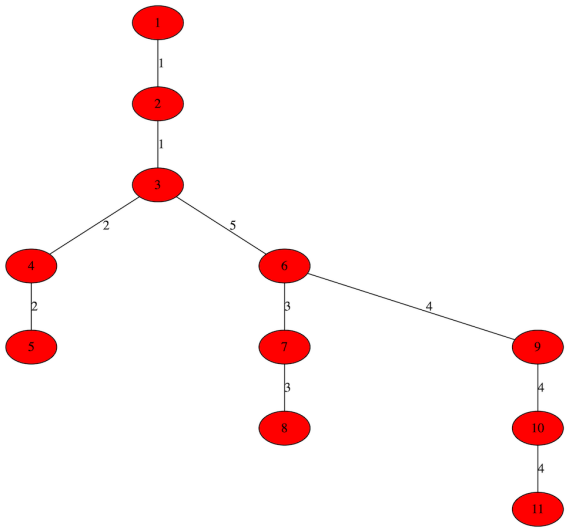


Figure 6.12: Graph representation of phantom 2 (Figure 6.10) updated to contain labeled edges denoting the unique index of the vessel segment that each serves as a filament in.

Table 6.4: Vessel segment attributes (defined in Section 6.2.2.2) calculated across the unique segments composing phantom 2.

Segment ID	**Length** (pixel)	**Surface Area** (pixel^2)	**Volume** (pixel^3)	**Distance** (pixel)	**Radius** (pixel)	**SA:V**	**LEN:DIA**	**Tortuosity**
1	14.14	155.50	138.84	14.14	1.5	1.12	4.71	1.00
2	14.14	155.50	138.84	14.14	1.67	1.12	4.71	1.00
3	8.94	98.35	87.81	8.94	1.67	1.12	2.98	1.00
4	18.21	209.67	195.34	18.11	1.75	1.07	6.07	1.01
5	4.47	56.20	56.20	4.47	2.00	1.00	1.49	1.00

The third phantom used is kite-like by nature: it includes a tail connected to a square. This geometry comprises a cycle connected to a line. We present this phantom in Figure 6.13. The result of the graph construction process is shown in Figure 6.14. We present the location of these nodes in the phantom volume in Table 6.5, along with the radius measurements at those node positions. The application of graph reduction to the "original" graph is shown in Figure 6.15. The "updated original" graph is shown in Figure 6.16.

The attributes (as defined in Section 6.2.2.2) for each vessel segment in the digital phantom are reported in Table 6.6. The overall characteristics (as defined in Section 6.2.2.1) for the vascular volume follow: total volume, $57{,}136.84\,\text{pixel}^3$; total surface area, $22{,}854.74\,\text{pixel}^2$; and total length, $727.49\,\text{pixel}$.

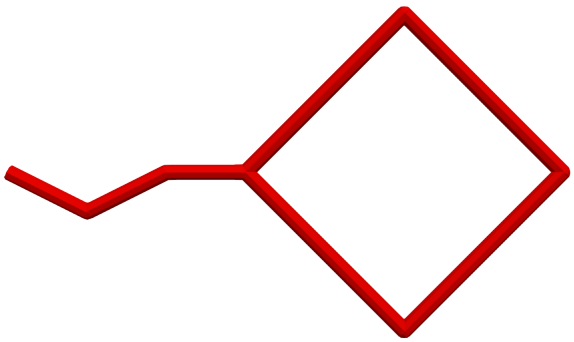


Figure 6.13: Phantom 3 volume.

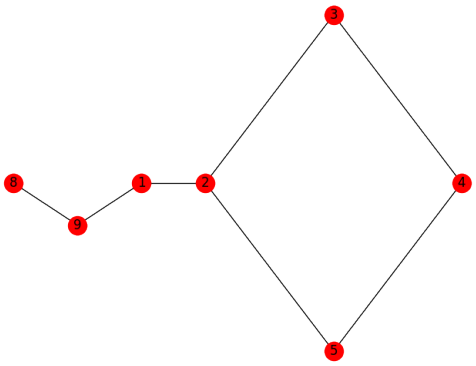


Figure 6.14: Graph representation of phantom 3 volume (Figure 6.13).

Table 6.5: Correspondence of nodes in the graph representation of phantom 3 (Figure 6.14) to their location in the volume (Figure 6.13), along with the radius measurement at that position.

Node No.	***x***	***y***	***z***	**Radius** (pixel)
1	-50	0	0	5
2	0	0	0	5
3	100	100	0	5
4	200	0	0	5
5	100	-100	0	5
8	-150	0	0	5
9	-100	-25	0	5

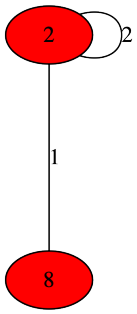


Figure 6.15: "Reduced" graph representation of phantom 3 (Figure 6.14) obtained by algorithm 3. The labels on each edge serve as a unique index for a respective vessel segment presenting in the volume (Figure 6.13).

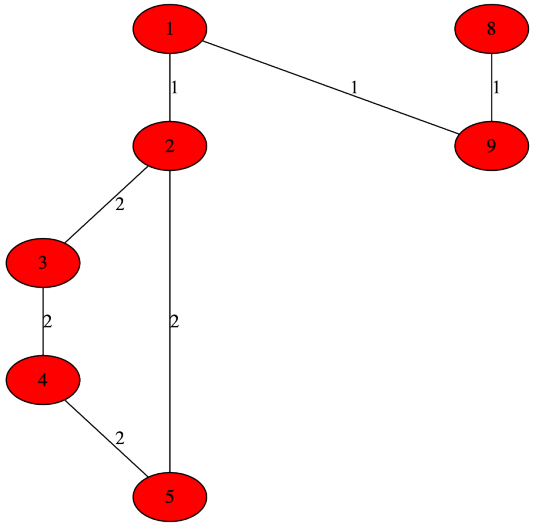


Figure 6.16: Graph representation of phantom 3 (Figure 6.14) updated to contain labeled edges denoting the unique index of the vessel segment that each serves as a filament in.

Table 6.6: Vessel segment attributes (defined in Section 6.2.2.2) calculated across the unique segments composing phantom 3.

Segment ID	**Length** (pixel)	**Surface Area** (pixel^2)	**Volume** (pixel^3)	**Distance** (pixel)	**SA:V**	**LEN:DIA**	**Tortuosity**
1	161.80	5083.20	12708.01	150.00	0.40	16.18	1.08
2	565.69	17771.53	44428.83	0.00	0.40	56.57	*inf*

The fourth phantom used is also kite-like by nature: it includes a tail along with a square partitioned in two by a line down the middle. This partition creates a bridge between one cycle and two paths to the tail. We present this phantom in Figure 6.17. The result of the graph construction process is shown in Figure 6.18. We present the location of these nodes in the phantom volume in Table 6.7, along with the radius measurements at those node positions. The application of graph reduction to the "original" graph is shown in Figure 6.19. The "updated original" graph is shown in Figure 6.20.

The attributes (as defined in Section 6.2.2.2) for each vessel segment in the digital phantom are reported in Table 6.6. The overall characteristics (as defined in Section 6.2.2.1) for the vascular volume follow: total volume, $72{,}844.80\,\text{pixel}^3$; total surface area, $29{,}137.92\,\text{pixel}^2$; and total length, $927.49\,\text{pixel}$.

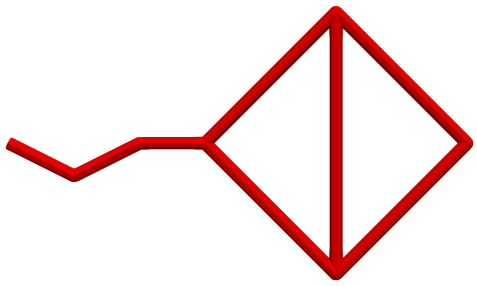


Figure 6.17: Phantom 4 volume.

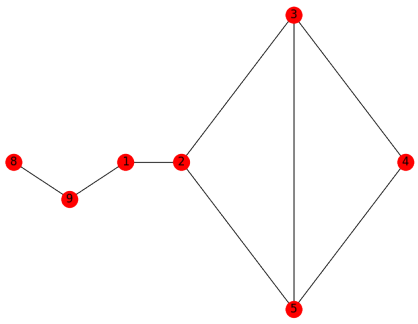


Figure 6.18: Graph representation of phantom 4 volume (Figure 6.17).

Table 6.7: Correspondence of nodes in the graph representation of phantom 4 (Figure 6.18) to their location in the volume (Figure 6.17), along with the radius measurement at that position.

Node No.	***x***	***y***	***z***	**Radius** (pixel)
1	-50	0	0	5
2	0	0	0	5
3	100	100	0	5
4	200	0	0	5
5	100	-100	0	5
8	-150	0	0	5
9	-100	-25	0	5

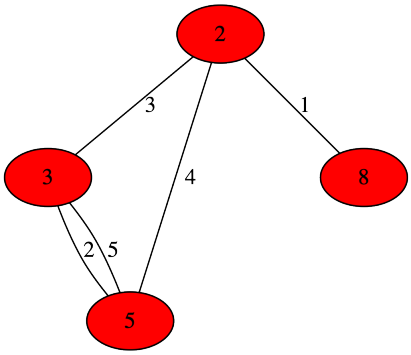


Figure 6.19: "Reduced" graph representation of phantom 4 (Figure 6.18) obtained by algorithm 3. The labels on each edge serve as a unique index for a respective vessel segment presenting in the volume (Figure 6.17).

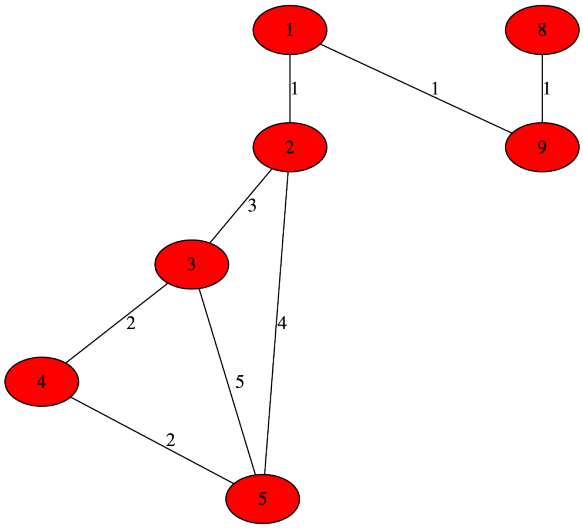


Figure 6.20: Graph representation of phantom 4 (Figure 6.18) updated to contain labeled edges denoting the unique index of the vessel segment that each serves as a filament in.

Table 6.8: Vessel segment attributes (defined in Section 6.2.2.2) calculated across the unique segments composing phantom 4.

Segment ID	**Length** (pixel)	**Surface Area** (pixel^2)	**Volume** (pixel^3)	**Distance** (pixel)	**SA:V**	**LEN:DIA**	**Tortuosity**
1	161.80	5083.20	12708.01	150.00	0.40	16.18	1.08
2	282.84	8885.77	22214.41	200.00	0.40	28.28	1.41
3	141.42	4442.88	11107.21	141.42	0.40	14.14	1.00
4	141.42	4442.88	11107.21	141.42	0.40	14.14	1.00
5	200.00	6283.19	15707.96	200.00	0.40	20.00	1.00

The fifth phantom used is figure-eight, like in its appearance. This structure includes two cycles spanning from the same node. We present this phantom in Figure 6.21. The result of the graph construction process is shown in Figure 6.22. We present the location of these nodes in the phantom volume in Table 6.9, along with the radius measurements at those node positions. The application of graph reduction to the "original" graph is shown in Figure 6.23. The "updated original" graph is shown in Figure 6.24.

The attributes (as defined in Section 6.2.2.2) for each vessel segment in the digital phantom are reported in Table 6.10. The overall characteristics (as defined in Section 6.2.2.1) for the vascular volume follow: total volume, $355{,}430.64\,\text{pixel}^3$; total surface area, $71{,}086.13\,\text{pixel}^2$; and total length, $1{,}131.37\,\text{pixel}$.

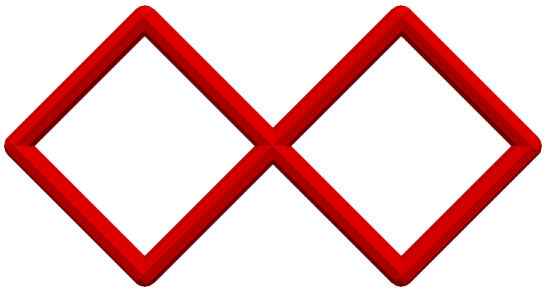


Figure 6.21: Phantom 5 volume.

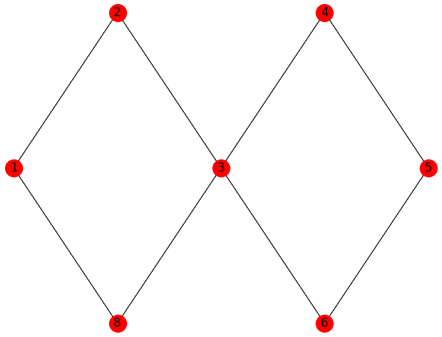


Figure 6.22: Graph representation of phantom 5 volume (Figure 6.21).

Table 6.9: Correspondence of nodes in the graph representation of phantom 5 (Figure 6.22) to their location in the volume (Figure 6.21), along with the radius measurement at that position.

Node No.	***x***	***y***	***z***	**Radius** (pixel)
1	0	0	0	10
2	100	100	0	10
3	200	0	0	10
4	300	100	0	10
5	400	0	0	10
6	300	-100	0	10
8	100	-100	0	10

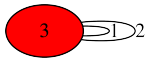


Figure 6.23: "Reduced" graph representation of phantom 5 (Figure 6.22) obtained by algorithm 3. The labels on each edge serve as a unique index for a respective vessel segment presenting in the volume (Figure 6.21).

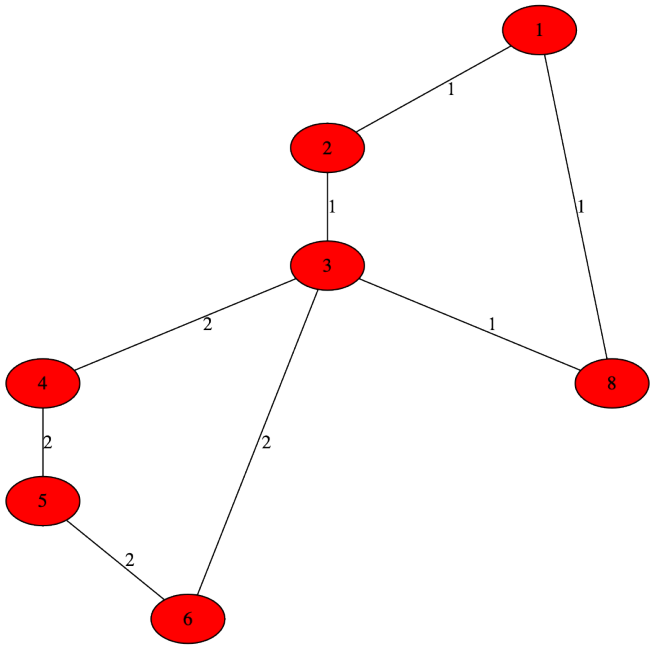


Figure 6.24: Graph representation of phantom 5 (Figure 6.22) updated to contain labeled edges denoting the unique index of the vessel segment that each serves as a filament in.

Table 6.10: Vessel segment attributes (defined in Section 6.2.2.2) calculated across the unique segments composing phantom 5.

Segment ID	**Length** (pixel)	**Surface Area** (pixel^2)	**Volume** (pixel^3)	**Distance** (pixel)	**SA:V**	**LEN:DIA**	**Tortuosity**
1	565.69	35543.06	177715.32	0.00	0.20	28.28	*inf*
2	565.69	35543.06	177715.32	0.00	0.20	28.28	*inf*

The sixth and final phantom used is composed of two figure-eight like structures overlapped such that the top 'O' of one is shared with the bottom 'O' of the other. This composition results in two cycles at either end of the structure, with two bridges emerging from the same points that connect these two cycles. We present this phantom in Figure 6.25. The result of the graph construction process is shown in Figure 6.26. We present the location of these nodes in the phantom volume in Table 6.11, along with the radius measurements at those node positions. The application of graph reduction to the "original" graph is shown in Figure 6.27. The "updated original" graph is shown in Figure 6.28.

The attributes (as defined in Section 6.2.2.2) for each vessel segment in the digital phantom are reported in Table 6.12. The overall characteristics (as defined in Section 6.2.2.1) for the vascular volume follow: total volume, $533{,}145.95\,\text{pixel}^3$; total surface area, $106{,}629.19\,\text{pixel}^2$; and total length, $1{,}697.06\,\text{pixel}$.

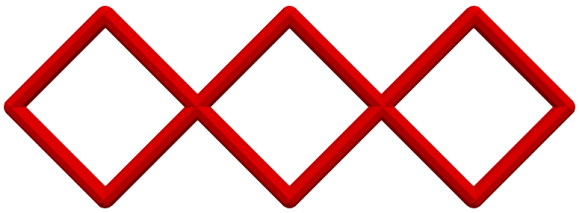


Figure 6.25: Phantom 6 volume.

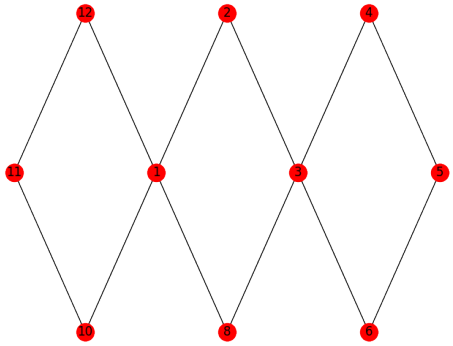


Figure 6.26: Graph representation of phantom 6 volume (Figure 6.25).

Table 6.11: Correspondence of nodes in the graph representation of phantom 6 (Figure 6.26) to their location in the volume (Figure 6.25), along with the radius measurement at that position.

Node No.	x	y	z	**Radius** (pixel)
1	0	0	0	10
2	100	100	0	10
3	200	0	0	10
4	300	100	0	10
5	400	0	0	10
6	300	-100	0	10
8	100	-100	0	10
10	-100	-100	0	10
11	-200	0	0	10
12	-100	100	0	10

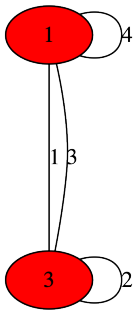


Figure 6.27: "Reduced" graph representation of phantom 6 (Figure 6.26) obtained by algorithm 3. The labels on each edge serve as a unique index for a respective vessel segment presenting in the volume (Figure 6.25).

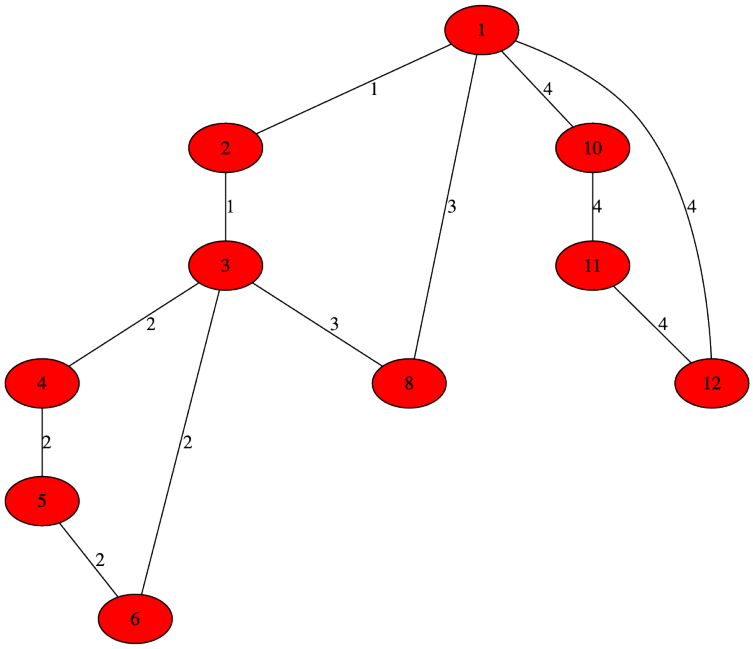


Figure 6.28: Graph representation of phantom 6 (Figure 6.26) updated to contain labeled edges denoting the unique index of the vessel segment that each serves as a filament in.

Table 6.12: Vessel segment attributes (defined in Section 6.2.2.2) calculated across the unique segments composing phantom 6.

Segment ID	**Length** (pixel)	**Surface Area** (pixel^2)	**Volume** (pixel^3)	**Distance** (pixel)	**SA:V**	**LEN:DIA**	**Tortuosity**
1	282.84	17771.53	88857.66	200.00	0.20	14.14	1.41
2	565.69	35543.06	177715.32	0.00	0.20	28.28	*inf*
3	282.84	17771.53	88857.66	200.0	0.20	14.14	1.41
	565.69	35543.06	177715.32	0.00	0.20	28.28	*inf*

6.3.2 Case study #2, data-driven synthetic model

To further illustrate the utility of our informatics platform, we provide an analysis of the filaments presenting in one of our data-driven models constructed in Chapter 4. The model we selected for this case study is presented in Figure 6.29 (for full-page image, see Figure 4.5) and was imagined as isotropic in resolution at $1.0\,\mu\text{m}$.

Figure 6.29: Cerebrovasculature model acquired from KESM India Ink dataset. This voxel-based model is $1158 \times 1158 \times 1158$ pixels derived from imaging data acquired at a voxel resolution of $0.6\,\mu\text{m} \times 0.7\,\mu\text{m} \times 1.0\,\mu\text{m}$. The total size of this model is $1.4\,\text{GB}$.

The graph for this model is shown in Figure 6.38 and the "reduced" graph in Figure 6.39. The overall characteristics for the vascular volume follow: total volume, $2{,}286{,}872.45\,\mu m^3$; total surface area, $636{,}994.41\,\mu m^2$; and total length, $16{,}293.23\,\mu m$. We found that this volume is comprised of $N = 286$ vessel segments. For the vessel segment attributes, we detail the mean, median, min, and max values in Table 6.13. We also present the distributions of these attributes across the vessel segments embedded in the model in Figures 6.30, 6.31, 6.32, 6.33, 6.34, 6.35, 6.36, and 6.37.

Table 6.13: Mean, median, minimum, and maximum values for vessel segment attributes calculated for the synthetic model presented in Figure 6.29.

Vessel Attribute	***Mean***	***Median***	***Min***	***Max***
Radius	$6.00\,\mu m$	$5.08\,\mu m$	$4.80\,\mu m$	$16.11\,\mu m$
Length	$56.97\,\mu m$	$40.67\,\mu m$	$4.20\,\mu m$	$474.25\,\mu m$
Surface Area	$2{,}227.25\,\mu m^2$	$1{,}499.82\,\mu m^2$	$126.59\,\mu m^2$	$38{,}630.19\,\mu m^2$
Volume	$7{,}996.06\,\mu m^3$	$4{,}434.60\,\mu m^3$	$303.83\,\mu m^3$	$269{,}709.58\,\mu m^3$
Distance	$52.77\,\mu m$	$38.29\,\mu m$	$4.20\,\mu m$	$450.14\,\mu m$
Tortuosity	1.06	1.02	1.00	1.93
LEN:DIA	4.93	3.68	0.42	27.25
SA:V	0.35	0.39	0.12	0.42

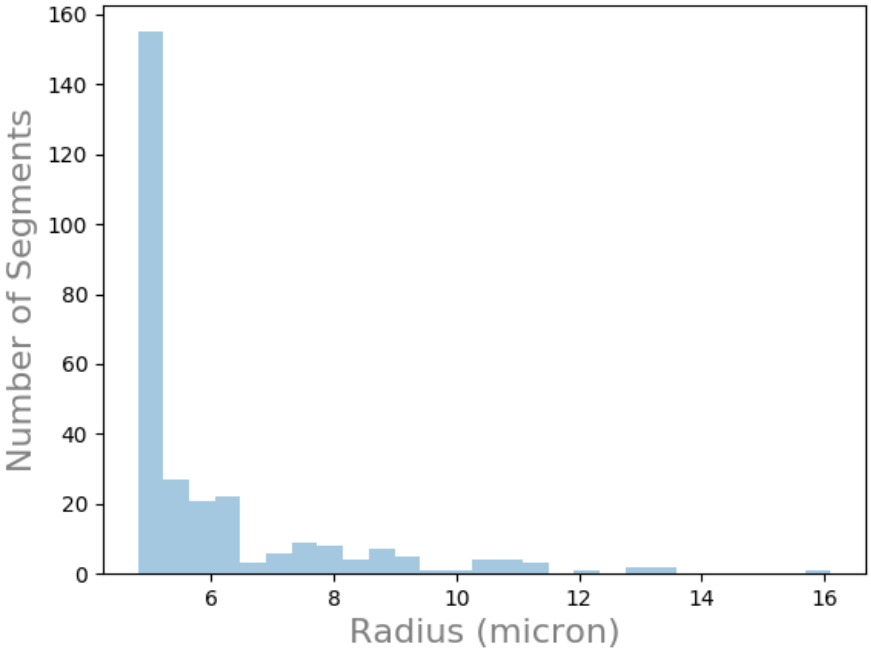


Figure 6.30: Number of segments by radius (μm) for the synthetic model presented in Figure 6.29. The total segments in the model is $N = 286$.

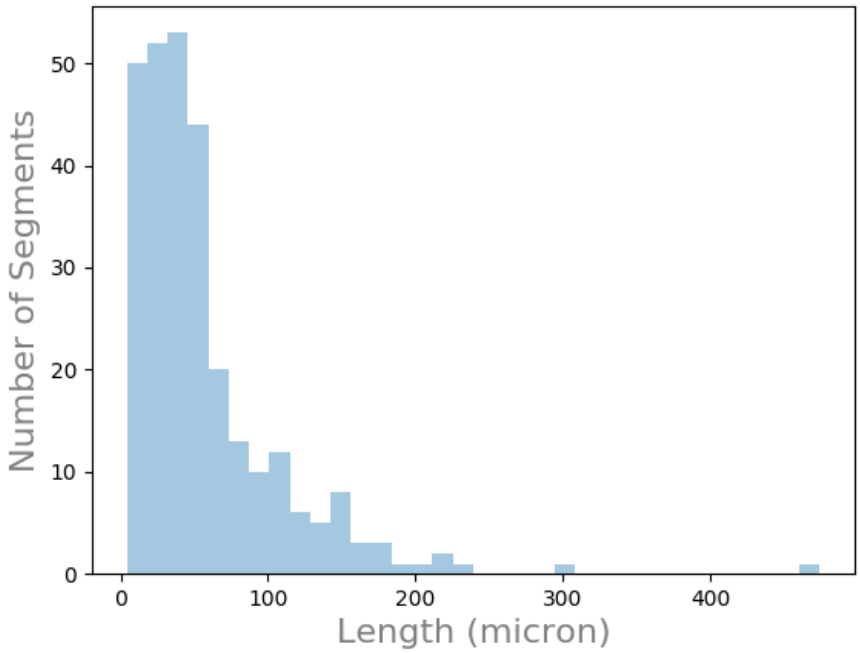


Figure 6.31: Number of segments by length (μm) for the synthetic model presented in Figure 6.29. The total number of segments in the model is $N = 286$.

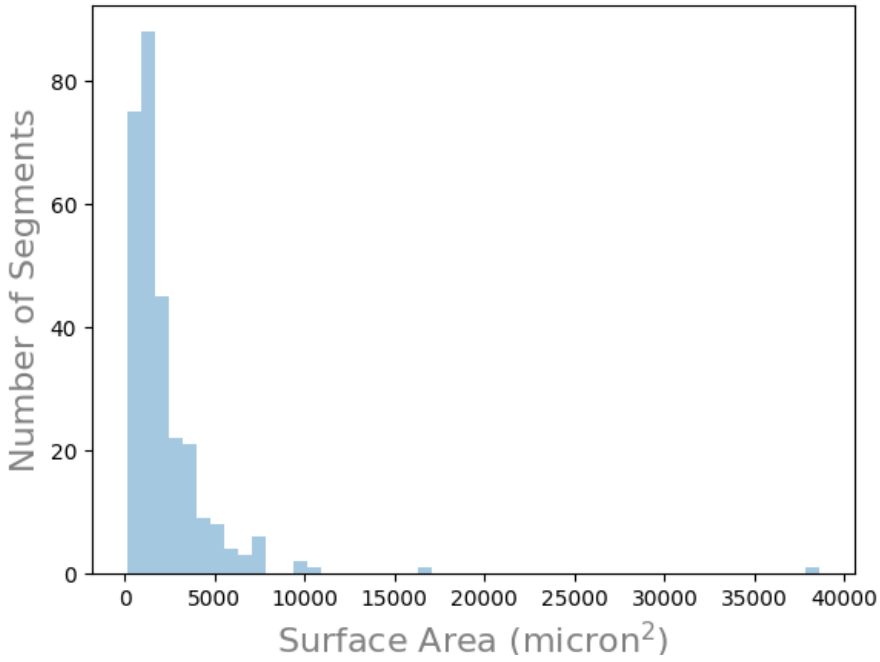


Figure 6.32: Number of segments by surface area (μm^2) for the synthetic model presented in Figure 6.29. The total number of segments in the model is $N = 286$.

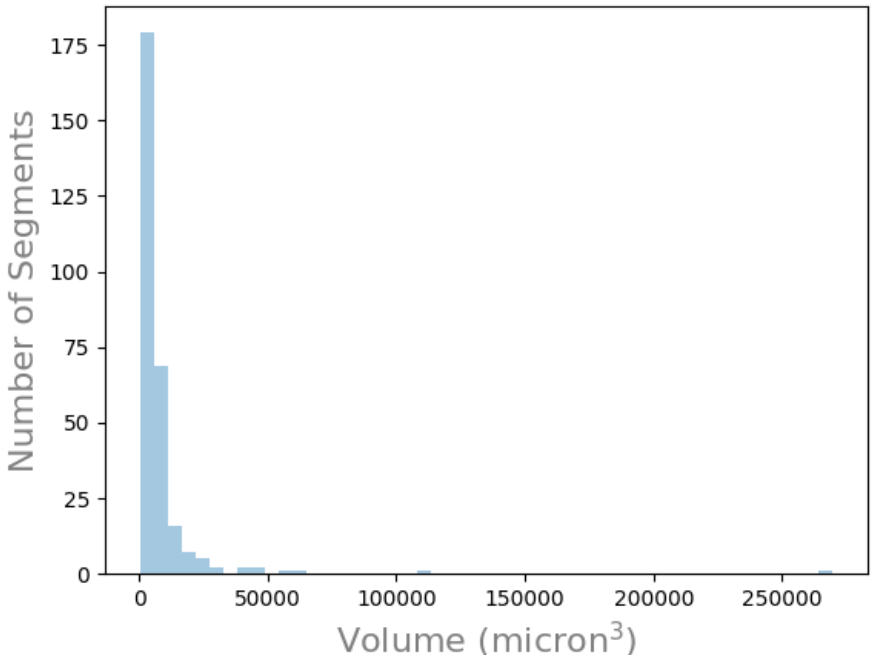


Figure 6.33: Number of segments by volume (μm^3) for the synthetic model presented in Figure 6.29. The total number of segments in the model is $N = 286$.

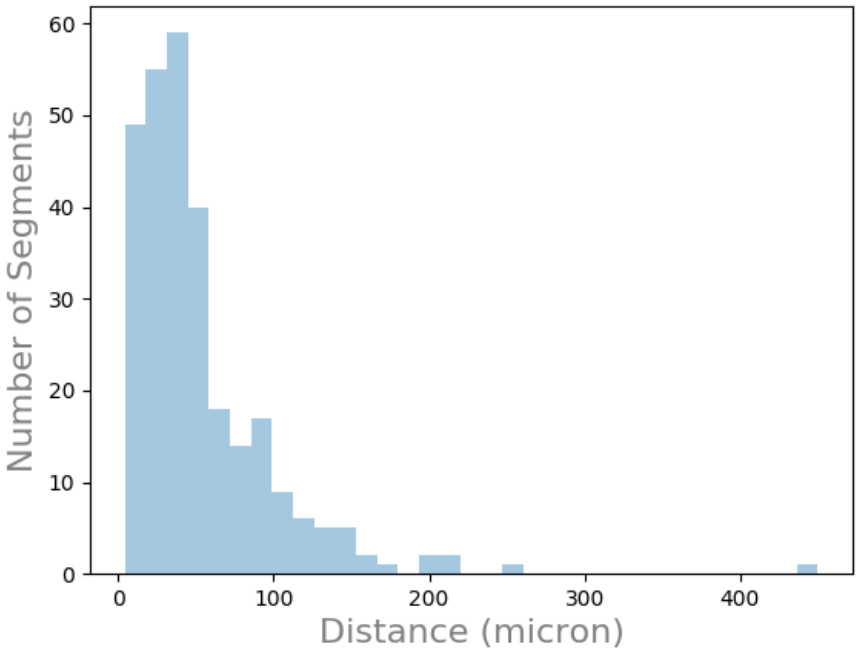


Figure 6.34: Number of segments by distance (μm) between their beginning and terminal end for the synthetic model presented in Figure 6.29. The total number of segments in the model is $N = 286$.

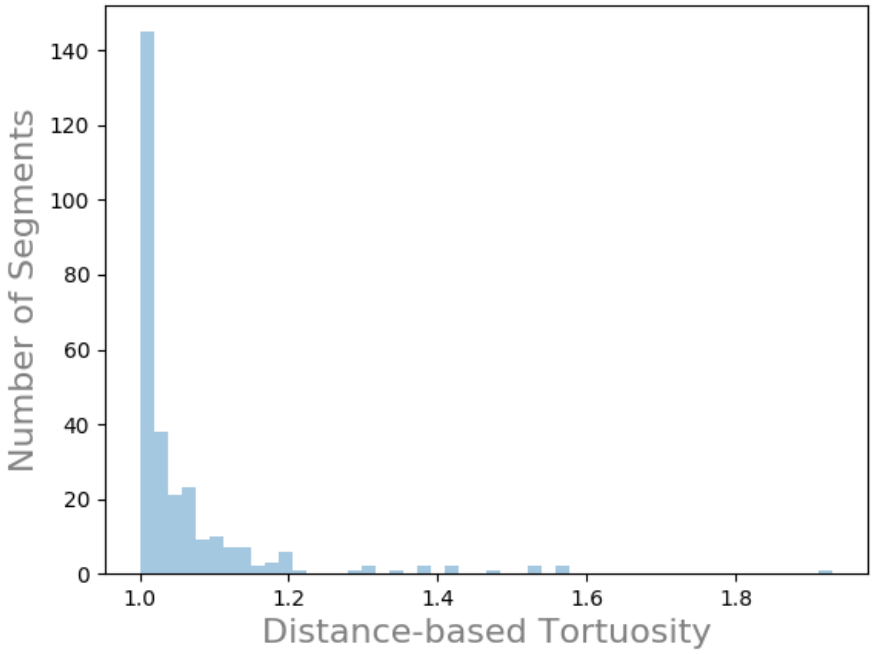


Figure 6.35: Number of segments by distance-based tortuosity for the synthetic model presented in Figure 6.29. The total number of segments in the model is $N = 286$.

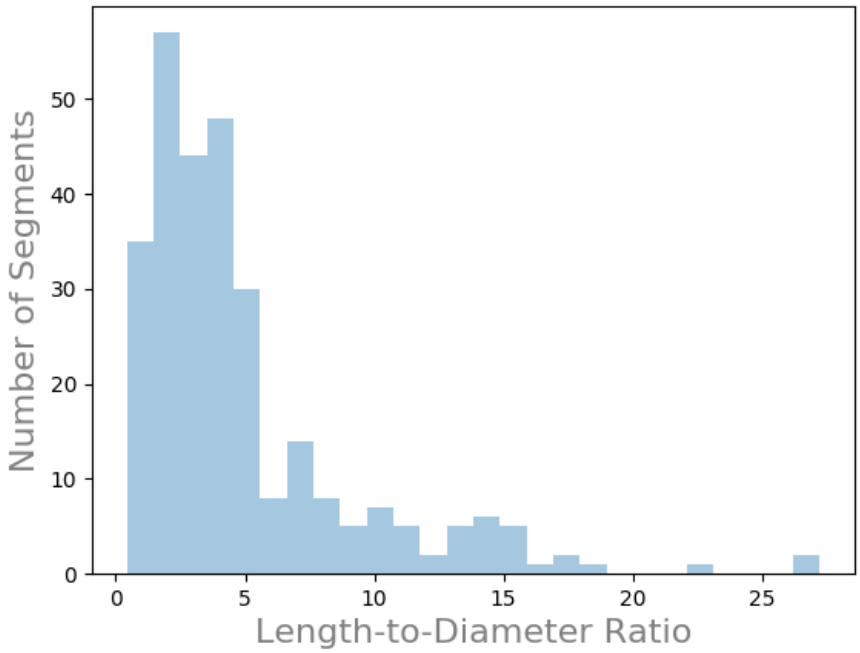


Figure 6.36: Number of segments by length-to-diameter ratio for the synthetic model presented in Figure 6.29. The total number of segments in the model is $N = 286$.

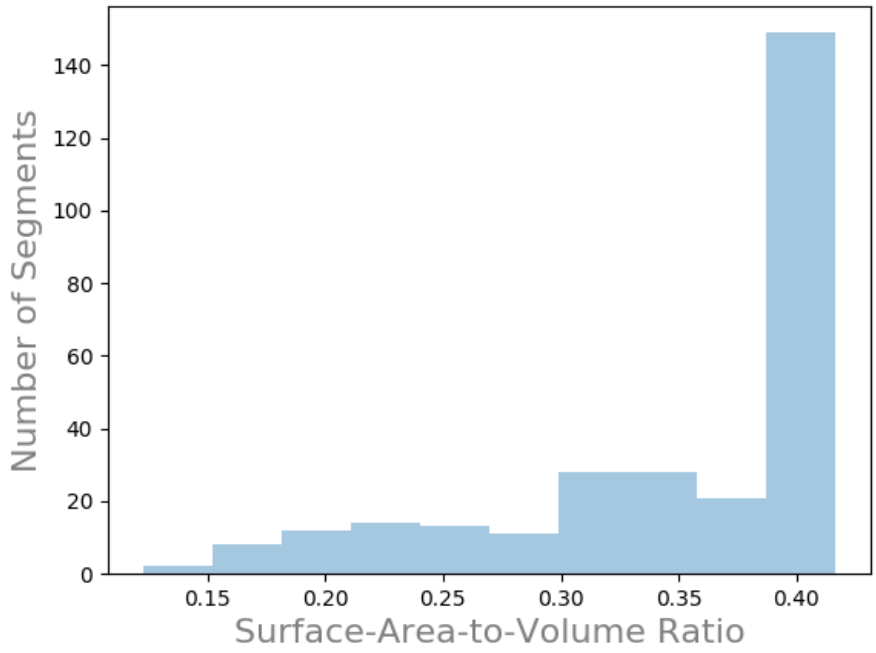


Figure 6.37: Number of segments by surface-area-to-volume ratio for the synthetic model presented in Figure 6.29. The total number of segments in the model is $N = 286$.

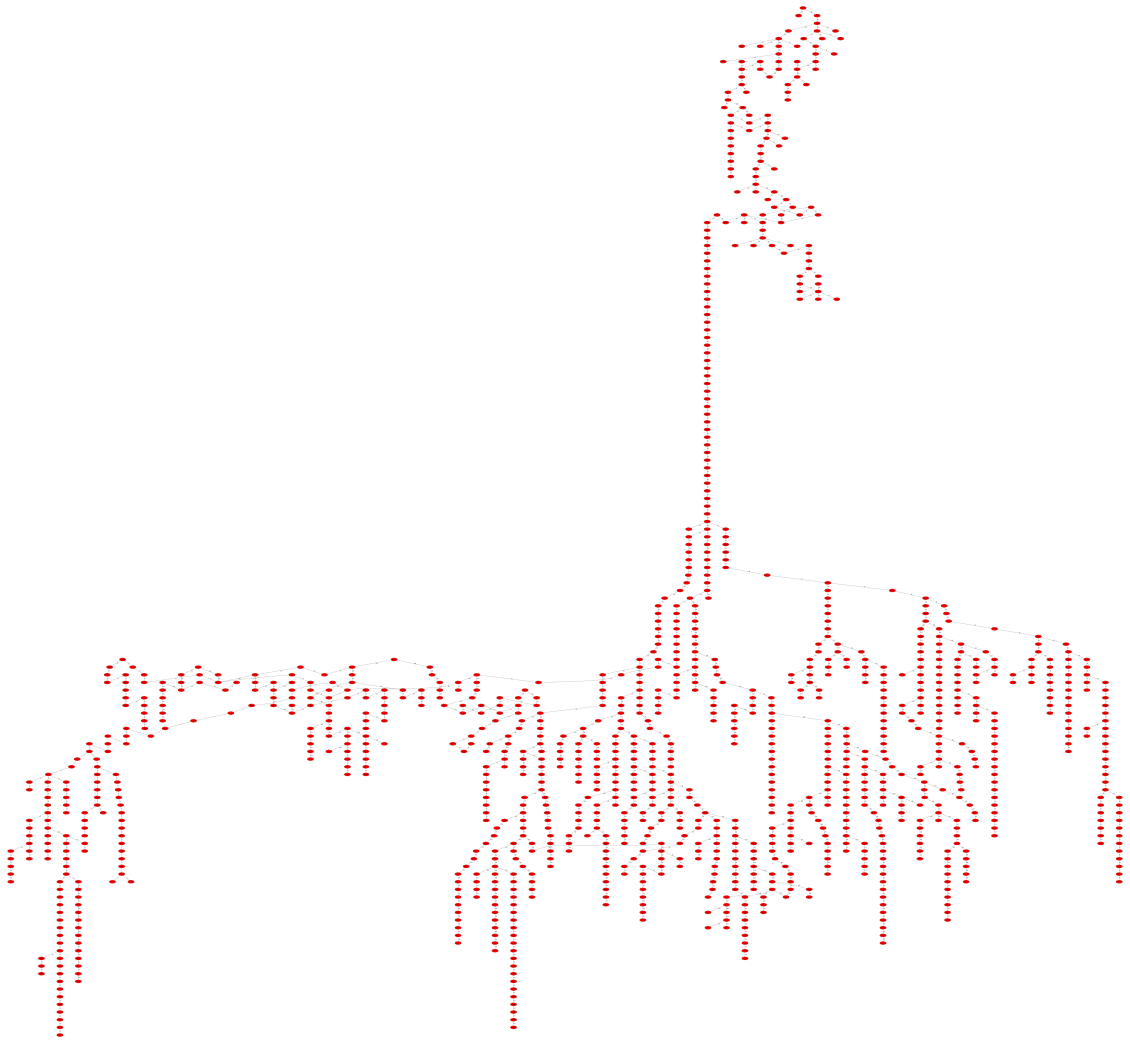


Figure 6.38: “Updated original” graph calculated for the synthetic model shown in Figure 6.29.

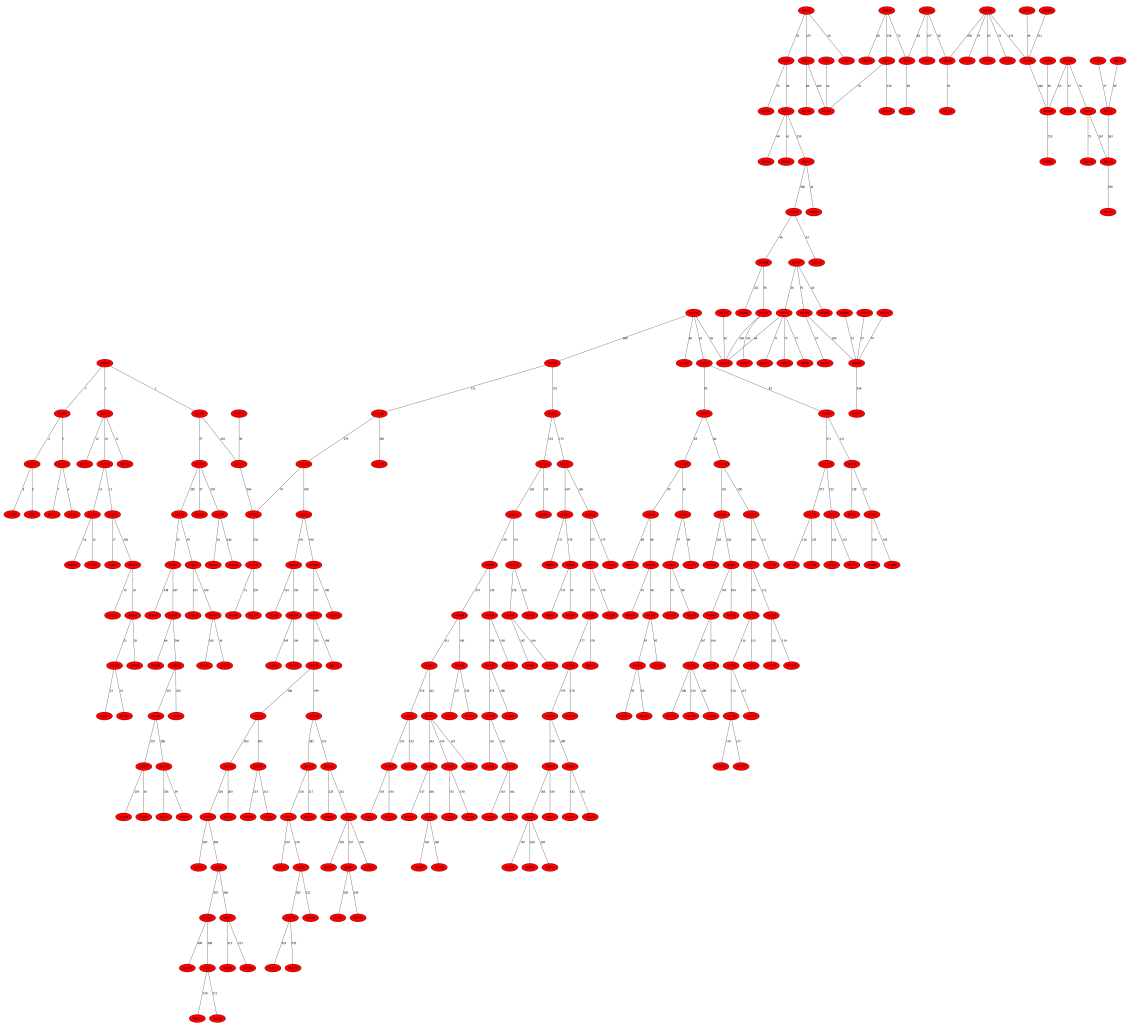


Figure 6.39: “Reduced” graph calculated for the synthetic model shown in Figure 6.29.

6.3.3 Summary

The principal contribution of this chapter is a framework for the analysis of cerebrovasculature acquired through different modalities, resolution, and volume. Our system directly addresses the need for a standardized informatics platform for the quantitative analysis of the high-resolution, high-volume, and multi-scale datasets procured using modern imaging techniques. Our platform can be applied irrespective of the resolution, size, or scale of the dataset and regardless of the modality used to acquire it. The implementation of our informatics platform is novel in that vessel connectivity information is readily available in its graph encoding. Our graph representation facilitates graph-theoretical analyses of the structural properties of the underlying vascular networks. It enables the seamless report of volume characteristics, including:

- Total length
- Total surface area
- Total volume
- Total-surface-area-to-total-volume ratio

In addition to volume characteristics, our framework also has attributes for each vessel segments comprising the cerebrovasculature under analysis. These attributes include:

- Radius
- Length
- Tortuosity, using the distance metric defined in [87]
- Distance (between beginning and terminal end of segment)
- Surface area
- Volume

- Surface-area-to-volume ratio (i.e., SA:V)
- Length-to-diameter ratio (i.e., LEN:DIA)

Given that our informatics platform stores both geometrical and topological data in graph form, we can apply different graph algorithms to isolate subgraphs meeting some criteria to be isolated. All that is required is a node in the vicinity of the region of interest, a springboard from which we traverse the local network. The structural organization of these areas can then be elucidated. We will elaborate more on this capability in the next chapter.

Our informatics platform is expected to have a significant scientific impact as it opens the door to data-driven, quantitative investigation of the brain's vascular system. We believe that our representation and storage of vasculature data in graph form with each vessel segment labeled and quantified through its attributes will provide researchers with a crucial resource towards the discovery of novel biomarkers of processes underlying healthy and diseased brains. In conclusion, our informatics platform will contribute a means for a standardized, quantitative approach to the study of cerebrovasculature datasets acquired using modern imaging techniques.

7. QUERYABLE DATABASE

7.1 Motivation

In chapter 6, we detailed how angiome topological and geometrical information can be stored in a graph representation. We described how our method employs a "reduced" graph that contains an edge for each vessel segment in the vasculature, that stores its attributes (as specified in Section 6.2.2.1) and an index into an "updated original" graph that stores the geometry of the filaments comprising that segment. The filaments comprising a respective segment can be retrieved from the "updated original" graph using a segment's index. Therefore, the "reduced" graph stores the topology and maps into the "updated original" graph for the geometry.

Our graph representation of a vascular volume can be entered into a graph database management system (i.e., graph database) to provide a better interface to the data. Such a database system provides create, read, update, and delete (CRUD) methods that expose the underlying data model; the database can also be accessed through the Internet. Storing vascular connectivity and geometrical information in a graph database, and providing an adequate interface to the data stored within it, will allow researchers to search and interact with that data from anywhere in the world. A researcher would be able to query for vessel segments meeting certain criteria (morphometry, taxonomy, type, etc.), within a specific scale (as organization and morphometry vary depending on size), or using a heuristic algorithm based on vessel characteristics. Furthermore, once data of interest are acquired, the volume can be reconstructed using the geometry information stored in the graph database. This would allow for seamless integration into applications such as blood flow simulations [18].

The Neo4j graph database platform can be used to store the graph representations of vascular data. Neo4j not only allows one to query the database but also provides standard graph algorithms [100] that can be used to analyze the network properties of the encoded data. The system that supports this is the Neo4j analytics platform; it runs atop the Neo4j graph database and provides a

suite of dynamic graph analytic tools through a library of high-performance algorithms. Included are algorithms that can be used for traversals, pathfinding, centrality, and community detection [100]. For vascular data, the traversal algorithms, such as breadth-first search (BFS) and parallel depth-first search (DFS), could be used to reconstruct particular aspects of the vascular network beginning from some starting point of interest. Pathfinding algorithms, such as single-source shortest path and all-pairs shortest path, would be useful in determining direct and alternating routing to different regions or in the analysis of network capacity [id.]. The Neo4j analytics platform also provides measures of centrality. This could provide researchers an ability to ascertain bottlenecks of blood flow into different regions of the volume under consideration [100]. Community detection algorithms, including the Louvain algorithm [109], are also offered. These algorithms could prove helpful in determining a region's robustness to occlusion [id.], given that communities are interconnected groups within a network whose nodes are more connected to one another than to other nodes. The application of these algorithms would be supported by the vessel segments attributes (section 6.2.2.2) computed by our informatics platform and stored in the graph representation.

In this chapter, we detail our procedure for the construction of a queryable database of vasculature data derived from its graph representation. Furthermore, we describe the interface that we developed to interact with that database and how it provides seamless access to the Neo4j analytical platform. Finally, we show how researchers can explore the topology and geometry of vasculature data using through that interface.

7.2 Methodology

7.2.1 Graph database platform set-up

We selected Neo4j (Neo4j, Inc., San Mateo, CA) primarily due to the graph analytical libraries that it provides, along with our familiarity with it; there are many different graph database platforms currently available. We installed the Neo4j community (v3.5.2) database engine following the documentation available on the Neo4j website. Subsequently, we installed The APOC (A Package of Component) library (v3.5.0.1) following the procedure detailed in the APOC User Guide

3.5. This plug-in consists of about 450 procedures and functions that help with data integration, graph algorithms, and data conversion. We also installed the Graph Algorithms library (v3.5.0.1) according to the Neo4j Graph Algorithms User Guide v3.5. This library provides efficiently implemented, parallel versions of many different graph algorithms for Neo4j v3.x. Both libraries allow access to the algorithms that they provide through Cypher procedures (Neo4j's query language). Using the Neo4j Graph Algorithms library, we achieve increased throughput through parallelization. Neo4j is also highly scalable and can be integrated into Amazon Web Services, Microsoft Azure, Google Cloud Platform, and other cloud services; a Neo4j graph database is entirely portable and can be distributed via popular cloud platforms.

7.2.2 Graph data model construction

Once Neo4j has been set-up, the next step is to import the graph representation to the database as the data model. Neo4j supports the importation of nodes or edges from comma-separated values (*CSV*) files with headers describing the data fields on each line. Therefore, we generate three *CSV* files composed of: (1), the nodes from the "updated original" graph; (2), the edges from the "updated original" graph; and (3), the edges from the "reduced" graph (including the segment attributes). The *CSV* file containing the nodes includes each node's ID, radius measure (*r*), *x*-coordinate, *y*-coordinate, and *z*-coordinate. In order to construct this file, we iterate through each node in the "updated original" graph, writing its id, x, y, z, and $radius$ to the nodes *CSV* file. We use Cypher, Neo4j's query language, to perform the insertion. The query used to perform this action is shown in Figure 7.1, which populates the graph data model line-by-line with the nodes from the *CSV* file. For each line read, a node is created in the database with the attributes specified in the file.

Next we construct the *CSV* for the edges from the "updated original" graph. We accomplish this by iterating through that graph's edges, writing out the ID of the *from* node, the ID of the *to* node, along with the segment's unique index (`sid`) for which that filament is apart. We then make a Cypher request to the Neo4j database to construct relationships in the data model for the edges stored in that *CSV* file. This is completed using the query presented in Figure 7.2.

```
LOAD CSV WITH HEADERS FROM "file:///path/to/nodes.csv" AS
 ↪ csvLine
CREATE (n:Node {id: toInt(csvLine.id), x:
 ↪ toFloat(csvLine.x), y: toFloat(csvLine.y), z:
 ↪ toFloat(csvLine.z),r: toFloat(csvLine.r)});
```

Figure 7.1: Cypher query used to construct nodes in the Neo4j database from a *CSV* file that has headers describing the data fields on each line. Each line of the *CSV* file describes a node and its attributes; the query reads line-by-line, populating the new node's attributes with those read from the file.

```
LOAD CSV WITH HEADERS FROM
 ↪ "file:///path/to/updatedOriginalEdges.csv" AS csvLine
MATCH (n1:Node { id: toInt(csvLine.from)}),(n2:Node { id:
 ↪ toInt(csvLine.to)})
CREATE (n1)-[r1:FILAMENT {sid:toInt(csvLine.sid)}]->(n2);
```

Figure 7.2: Cypher query used to construct `FILAMENT` relationships (i.e., edges) in the Neo4j database from a *CSV* file of "updated original" graph edges. This file has headers describing the data fields on each line. Each line of the *CSV* file describes a edge and its properties. The query reads line-by-line, populating the new relationship's properties with those of the edge read from the file.

The relationships in the graph model (i.e., edges) are labeled and can have additional attributes (such as `sid`). We label all the edges imported from the "updated original" graph as `FILAMENT` to recognize that they store the geometry of a particular segment; the segment to which they belong is denoted by their `sid`. We use these edges when reconstructing an individual or series of vessel segments. We also import the edges comprising the "reduced" graph. These edges represent a single vessel segment and are uniquely identified by a segmentation index (`sid`); a `FILAMENT`'s `sid` maps to a unique segment with the corresponding `sid`.

The *CSV* file containing these edges is similarly generated by iterating over each edge of the "reduced" graph, writing out the ID of the *from* node, the ID of the *to* node, the segment's unique index (`sid`), along with all of the attributes calculated for that segment (see Section 6.2.2.2).

Again, we use Cypher to tell Neo4j to construct relationships in the graph data model according to the edges stored in the *CSV* file. We request this using the query presented in Figure 7.3. The relationships formed from these edges receive the label `SEGMENT` in the data model.

```
LOAD CSV WITH HEADERS FROM
 ↪ "file:///path/to/reducedEdges.csv" AS csvLine
MATCH (n1:Node { id: toInt(csvLine.from)}),(n2:Node { id:
 ↪ toInt(csvLine.to)})
CREATE (n1)-[r1:SEGMENT {sid:toInt(csvLine.sid),
 ↪ length:toFloat(csvLine.length),
 ↪ surface_area:toFloat(csvLine.surface_area),
 ↪ volume:toFloat(csvLine.volume),
 ↪ dist:toFloat(csvLine.dist),
 ↪ avg_radius:toFloat(csvLine.avg_radius),
 ↪ surface_to_volume_ratio:
 ↪ toFloat(csvLine.surface_to_volume_ratio),
 ↪ length_to_diameter_ratio:
 ↪ toFloat(csvLine.length_to_diameter_ratio),
 ↪ tortuosity:toFloat(csvLine.tortuosity)}]->(n2);
```

Figure 7.3: Cypher query used to construct `FILAMENT` relationships (i.e., edges) in the Neo4j database from a *CSV* file of "reduced" graph edges. This file has headers describing the data fields on each line. Each line of the *CSV* file describes a edge and its properties. The query reads line-by-line, populating the new relationship's properties with those of the edge read from the file.

After importing the nodes and edges from the "updated original" graph, along with the edges from the "reduced" graph, we have two relationships in the graph data model: `SEGMENT` and `FILAMENT`. Edges labeled `SEGMENT` encodes the connectivity of a vessel segment along with that segment's attributes; those labeled `FILAMENT` comprise the individual twists and turns to comprise the geometry of a respective segment. The `sid` of a `SEGMENT` indexes the `FILAMENT`s comprising that `SEGMENT`. Performing a traversal on edges labeled `SEGMENT` will move one from one vessel segment to another. We isolate segments meeting some criteria by querying for edges with the label `SEGMENT` for that criteria. We then get the geometry of that segment with unique

`sid` by querying the database for all `FILAMENT` with that `sid`.

It is important to note that Neo4j only supports directed edges in its graph data model. Therefore, the edges that we've inserted into the graph are directed. However, we stated in section 6.2.1 that our graph representation would store undirected edges due to the nature of blood flow. The limitation that Neo4j can only store directed edges is easily overcome: when querying the database, we can tell the database engine to treat the edges as non-directed.

7.2.3 Database custom interface

In order to interact with the Neo4j database, we created a custom interface written in the Python programming language. This interface leverages the existing Python library *py2neo* which provides the means to connect to a Neo4j database, execute a query, and receive the result back. Our assumption is the Neo4j database containing vascular data will be hosted remotely; that our users will query the remote database and interact with results locally. We have written four algorithms to aid in this process.

Our first algorithm allows one to query the database with any native Neo4j or Neo4j library query that returns a path and construct a graph representation of its result. The second runs a depth-first traversal to level l from starting node n, traverses out l `SEGMENT`s from n, and returns a *networkx* graph embedding the geometry of the segments traversed. This is done by mapping into the `FILAMENTS` relationships with the `sid` of each segment visited. The third searches for the shortest path across `SEGMENT` relationships between two nodes, using the values of some property (Table 7.1) as the relationship's weight in the process, and returns a *networkx* graph embedding the geometry of the segments traversed. The fourth looks up a `SEGMENT` by its `sid` and returns all properties stored within that relationship; i.e., that vessel segment's attributes (Table 7.1).

Table 7.1: Vessel segment attributes reported (defined in Section 6.2.2.2).

Segment Attribute
Radius
Length
Tortuosity
Distance
Surface Area
Volume
SA:V
LEN:DIA

7.2.3.1 General database query result to graph representation

The majority of queries into our database will return relationships and nodes. For instance, one might want to query the database to select all `FILAMENTS` with a `sid` of 2, 3, or 4 to isolate segments 2, 3, or 4. Or perhaps we initiate a depth-first search query starting from some segment and expanding out n segments from the starting point. In these cases and many others, the result of the query is a list of relationships (i.e., edges) coded with their properties and a list of nodes coded with their attributes. Using the information returned by a query, we construct a *networkx* graph that is compatible with our informatics platform (detailed in chapter 6). This is achieved using algorithm 4.

Algorithm 4 General database query result to graph representation

Input: Database connection and query
Output: Graph representation of query result

Run query on database
Let `nodes`, `edges` be the nodes and relationships returned by the query
Let `G` be initialized to a *networkx* graph
for each node `n` in `nodes` **do**
    Let `id` be `n`'s ID
    Let `x`, `y`, `z`, be `n`'s x-, y-, and z-coordinate points respectively.
    Let `r` be `n`'s radius
    Insert a new node into `G` with attributes id:`id`, x:`x`, y:`y`, z:`z`, and r:`r`
end for
for each edge `e` in `edges` **do**
    Let `n1` be one node comprising the edge
    Let `n2` be the other node comprising the edge
    Insert a new edge into `G` connecting `n1` and `n2`
end for
return `G`

7.2.3.2 *Depth-first traversal of l segments from node n*

Leveraging Neo4j's *APOC* library, we perform a depth-first traversal l levels away from some node n with the library's `expandConfig` algorithm. This is done by traversing relationships in the data model labeled as `SEGMENT`, meaning that the level of the depth-first traversal is the number of segments away from the starting node n. This query is written according to Figure 7.4.

```
MATCH (n:Node)
WHERE n.id = STARTING_NODE_ID
CALL apoc.path.expandConfig(n,
  ↪ {relationshipFilter:"SEGMENT",
  ↪ maxLevel:NO_SEGMENTS_AWAY, uniqueness:"NODE_GLOBAL"})
  ↪ YIELD path
RETURN path, nodes(path) as nodes, rels(path) as rels
```

Figure 7.4: Cypher query used to perform a depth first traversal starting from the node with ID `STARTING_NODE_ID`, moving `NO_SEGMENTS_AWAY` segments from that node, returning the nodes and their attributes, along with relationships and their properties, traversed.

The result of this depth-first traversal query is the nodes and relationships traversed in the graph model. We apply algorithm 4 to construct a *networkx* graph of that result, allowing our informatics platform to be applied only to those segments l levels away from n. The overall procedure is presented as an algorithm 5.

Algorithm 5 Depth-first traversal l segments from node n and graph construction of that result

Input: Database `graph_db` and query `db_query`
Output: Graph representation of query result
1: Execute `db_query` on `graph_db`
2: Let `edges` be the relationships returned after execution of `db_query`
3: Let `sids` be a list of `sids`
4: **for** each edge `e` in `edges` **do**
5: Append `e`'s `sid` to `sids`
6: **end for**
7: Let `sid_query` be a query string that selects all `FILAMENT` relationships having a `sid` in `sids`
8: **return** Output of algorithm 4 provided `graph_db` and `sid_query` as inputs.

7.2.3.3 Shortest path from node $start$ to node end

This algorithm finds the minimum number of segments needed to be traversed to get from node $start$ to node end. This algorithm proceeds across `SEGMENT` labeled relationships in the graph database, using some property (Table 7.1) as the relationship's weight in the process. We achieve this behavior by leveraging Neo4j's *APOC* library and executing its implementation of Dijkstra's algorithm. This query is shown in Figure 7.5.

The query returns the overall weight, nodes, and their attributes, along with relationships and their properties, for the shortest path found. We apply algorithm 4 to construct a *networkx* graph of that result, allowing our informatics platform to be applied to the vasculature of comprising the pathway that was found. Our implementation of this is nearly identical to algorithm 5. The only minor differences are the query being executed and the parameters: (1) graph database; (2) starting node; (3) terminal node; (4) the attribute to be used as a weight. Parameters (2-4) are substituted

```
MATCH (from:Node{id:NODE_FROM}), (to:Node{id:NODE_TO})
CALL apoc.algo.dijkstra(NODE_FROM, NODE_TO, 'SEGMENT',
 ↪ 'WEIGHT_ATTRIBUTE') yield path as path, weight as weight
RETURN path, nodes(path) as nodes, rels(path) as rels,
 ↪ weight
```

Figure 7.5: Cypher query used to search for the shortest path across relationships labeled as `SEGMENT` from `NODE_FROM` to `NODE_TO`. The query returns the nodes and their attributes, along with relationships and their properties, composing the shortest path. In addition, the overall weight across that is also returned.

accordingly into the query (Figure 7.5).

7.2.3.4 Segment property look-up

For a vessel segment encoded in our graph database, it is likely that we would like to query the properties of that segment via the relationship `SEGMENT` storing those properties. We perform this lookup by writing an algorithm that performs a query into the database returning that relationship `SEGMENT` and enters its properties (Table 7.1) into a dictionary that is returned to the user. The actual database query to get the segment is shown in Figure 7.6.

```
MATCH ()-[r:SEGMENT]-()
WHERE r.sid = SEGMENT_ID
RETURN r
```

Figure 7.6: Cypher query used retrieve the relationship `SEGMENT` with specified `sid` value `SEGMENT_ID`.

After executing the query, we simply iterate through the properties for the relationship returned, constructing a dictionary of those properties in the process. That dictionary is then presented to the user.

7.2.4 Reconstruction from graph representation

After querying the database and constructing a graph representation of that result, it is possible to employ our informatics platform to characterize the vascular data returned by that query. Algorithms provided by *networkx* that are not included in our informatics platform can also be applied. However, in most cases, we would like to reconstruct the vascular data embedded in the graph representation constructed from the database query result. To do this, we first construct a *SWC* file (ref: sections 6.2.1 and 4.2.1) from the *networkx* graph according to algorithm 6.

Algorithm 6 Depth-first traversal l segments from node n and graph construction of that result

Input: Graph representation `G`
Output: *SWC* file f of `G`

Let `f` be a file opened for writing
for each connected component `cc` in `G` **do**
    Let `leafs` be a list of all nodes in `cc` with degree less than or equal to 1
    Get the first node `n` from `leafs`
    Let `id` be `n`'s ID
    Let `x`, `y`, `z`, be `n`'s x-, y-, and z-coordinate points respectively.
    Let `r` be `n`'s radius
    Write to new line to `f`: "`id` 3 `x` `y` `z` `r` -1"
    Get list `successors` of all connected successors to `n` from `G` as (parent node `parent`, child node `child`) tuples.
    for each (`parent`, `child`) pair in `successors` **do**
        Let `id` be `child`'s ID
        Let `x`, `y`, `z`, be `child`'s x-, y-, and z-coordinate points respectively.
        Let `r` be `child`'s radius
        Let `parent_id` be `parents`'s ID
        Write to new line to `f`: "`id` 3 `x` `y` `z` `r` `parent_id`"
    end for
end for

After constructing the *SWC* file from the graph representation, we employ Daisuke Miyamoto's `swc2vtk` converter [110]. This creates a `VTK` file encoding the vascular data in vector graphics format, which can be viewed in an application such as ParaView [111].

7.3 Results and analysis

7.3.1 Case study #1, digital phantoms

We first illustrate the utility of encoding a vasculature as a data model in a graph database using two digital phantoms.

7.3.1.1 Digital phantom #1

The first digital phantom that we use is presented in Figure 7.7. This phantom exhibits two branches that are connected through a bridge. The result of the graph construction process on this volume, per our procedure detailed in Section 6.2.1, is shown in Figure 7.8. We present the location of the nodes in the phantom volume in Table 7.2, along with the radius measurements at those node positions. The application of our graph reduction algorithm to this "original" graph (i.e., that in Figure 7.8) is shown in Figure 7.10. Each edge in our "reduced" graph is labeled with a vessel segment number that maps to edges composing that vessel segment in the "updated original" graph (shown in Figure 7.9). This allows for the easy reconstruction of the different vessel segments. For instance, to reconstruct the vessel segment denoted by edge 1 in the "reduced" graph (i.e., 7.10), we would derive all edges labeled with 1 from the "updated original" graph (i.e., Figure 7.9).

The attributes (as defined in Section 6.2.2.2) for each vessel segment (i.e., vessel arising between branch points/endpoints) in the digital phantom are reported in Table 7.3. The overall characteristics (as defined in Section 6.2.2.1) for the vascular volume follow: total volume, $331.28\,\text{pixel}^3$; total surface area, $662.55\,\text{pixel}^2$; and total length, $105.45\,\text{pixel}$. For the vessel segment attributes, we detail the mean, median, min, and max values in Table 7.4

The Neo4j data model constructed from the graph representation of this phantom is shown in Figure 7.11. In the figure, the nodes are shown as light blue circles, with the `FILAMENT` relationships between them in blue and `SEGMENT` relationships in drawn in red. The labels on those "edges" denotes that relationship's `sid` value. These values are used to map from a `SEGMENT` to its `FILAMENTS`. For example, the `FILAMENT`s labeled 1 store the geometry of `SEGMENT` 1.

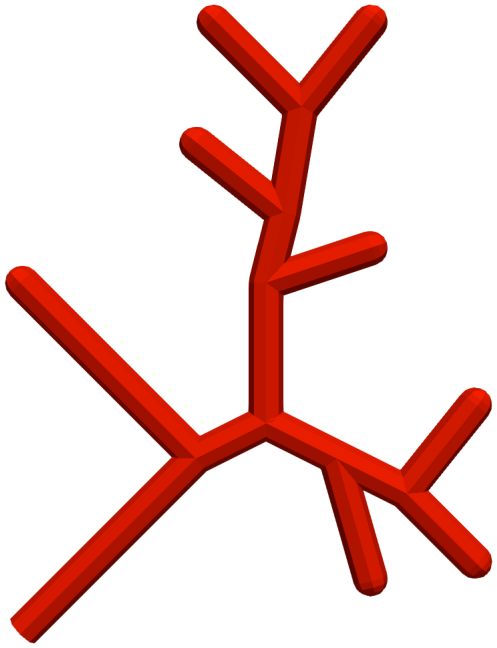


Figure 7.7: Phantom 1 volume.

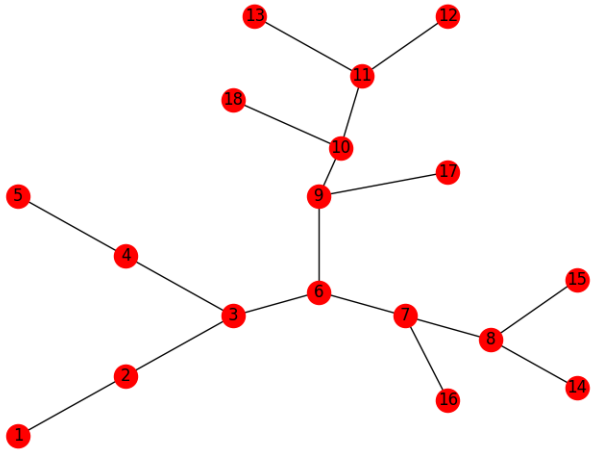


Figure 7.8: Graph representation of phantom 1 volume (Figure 7.7).

Table 7.2: Correspondence of nodes in the graph representation of phantom 1 (Figure 7.8) to their location in the volume (Figure 7.7), along with the radius measurement at that position.

Node No.	***x***	***y***	***z***	**Radius** (pixel)
1	-10	-10	0	1
2	-5	-5	0	1
3	0	0	0	1
4	-5	5	0	1
5	-10	10	0	1
6	4	2	0	1
7	8	0	0	1
8	12	-2	0	1
9	4	10	0	1
10	5	14	0	1
11	6	20	0	1
12	10	25	0	1
13	1	25	0	1
14	16	-6	0	1
15	16	3	0	1
16	10	-7	0	1
17	10	12	0	1
18	0	18	0	1

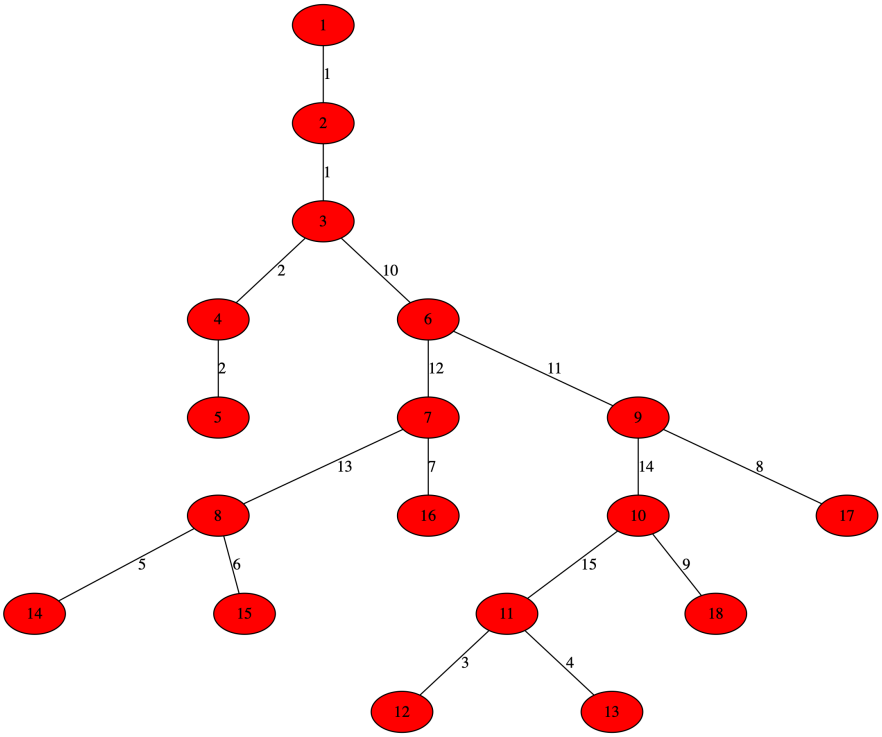


Figure 7.9: Graph representation of phantom 1 (Figure 7.8) updated to contain labeled edges denoting the unique index of the vessel segment that each serves as a filament in.

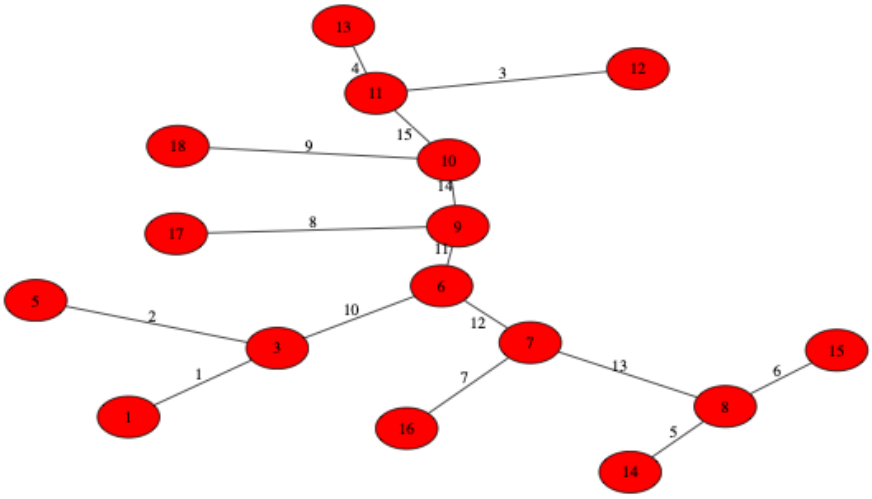


Figure 7.10: "Reduced" graph representation of phantom 1 (Figure 7.8) obtained by algorithm 3. The labels on each edge serve as a unique index for a respective vessel segment presenting in the volume (Figure 7.7).

Table 7.3: Vessel segment attributes (defined in Section 6.2.2.2) calculated across the unique segments composing phantom 1.

Segment ID	**Length** (pixel)	**Surface Area** (pixel^2)	**Volume** (pixel^3)	**Distance** (pixel)	**SA:V**	**LEN:DIA**	**Tortuosity**
1	14.14	88.86	44.43	14.14	2.00	7.07	1.00
2	14.14	88.86	44.43	14.14	2.00	7.07	1.00
3	6.40	40.23	20.12	6.40	2.00	3.20	1.00
4	7.07	44.43	22.21	7.07	2.00	3.54	1.00
5	5.66	35.54	17.77	5.66	2.00	2.83	1.00
6	6.40	40.23	20.12	6.40	2.00	3.20	1.00
7	7.28	45.74	22.87	7.28	2.00	3.64	1.00
8	6.32	39.74	19.87	6.32	2.00	3.16	1.00
9	6.40	40.23	20.12	6.40	2.00	3.2	1.00
10	4.47	28.10	14.05	4.47	2.00	2.24	1.00
11	8.00	50.27	25.13	8.00	2.00	4.00	1.00
12	4.47	28.10	14.05	4.47	2.00	2.24	1.00
13	4.47	28.10	14.05	4.47	2.00	2.24	1.00
14	4.12	25.91	12.95	4.12	2.00	2.06	1.00
15	6.08	38.22	19.11	6.08	2.00	3.04	1.00

Table 7.4: Mean, median, minimum, and maximum values for vessel segment attributes calculated for the synthetic model presented in Figure 7.7.

Vessel Attribute	***Mean***	***Median***	***Min***	***Max***
Radius	1.00 pixel	1.00 pixel	1.00 pixel	1.00 pixel
Length	7.03 pixel	6.40 pixel	4.12 pixel	14.14 pixel
Surface Area	44.17 pixel^2	40.23 pixel^2	25.91 pixel^2	88.86 pixel^2
Volume	22.09 pixel^3	20.12 pixel^3	12.95 pixel^3	44.43 pixel^3
Distance	7.03 pixel	6.40 pixel	4.12 pixel	14.14 pixel
Tortuosity	1.00	1.00	1.00	1.00
LEN:DIA	3.51	3.20	2.06	7.07
SA:V	2.00	2.00	2.00	2.00

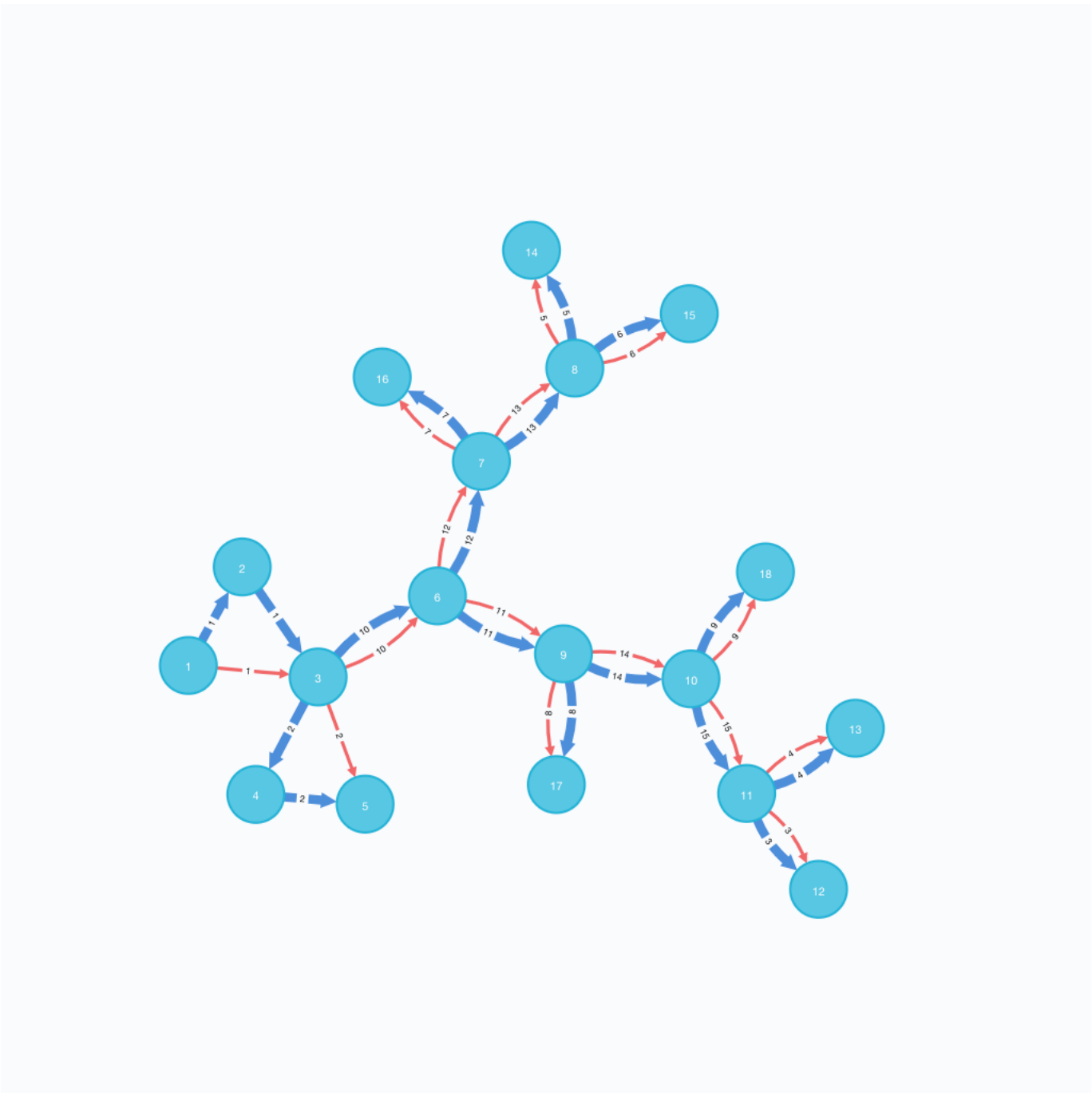


Figure 7.11: Neo4j data model constructed for the phantom shown in Figure 7.7. In this figure, the node light blue circles, with the `FILAMENT` relationships between them in blue and `SEGMENT` relationships in drawn in red. The labels on an "edges" denotes its `sid` value; the labels on a node is its unique ID.

Once the data model has been entered into the database, we can query the database using algorithm 4 and reconstruct the result using our procedure detailed in section 7.2.4. For instance, if we wanted to query the database for the filaments comprising segments 3, 11, 14, and 15 we would pass algorithm 4 the query presented in Figure 7.12. This query (Figure 7.12) returns the `FILAMENT` relationships and nodes shown in Figure 7.13.

```
MATCH path = (start:Node)-[n:FILAMENT]-(end:Node)
WHERE  n.sid=3 or n.sid = 11 or n.sid = 14 or n.sid = 15
RETURN path, nodes(path) as nodes, rels(path) as rels
```

Figure 7.12: Cypher query used retrieve the relationship `FILAMENTS` with specified `sid` value 3, 11, 14, or 15.

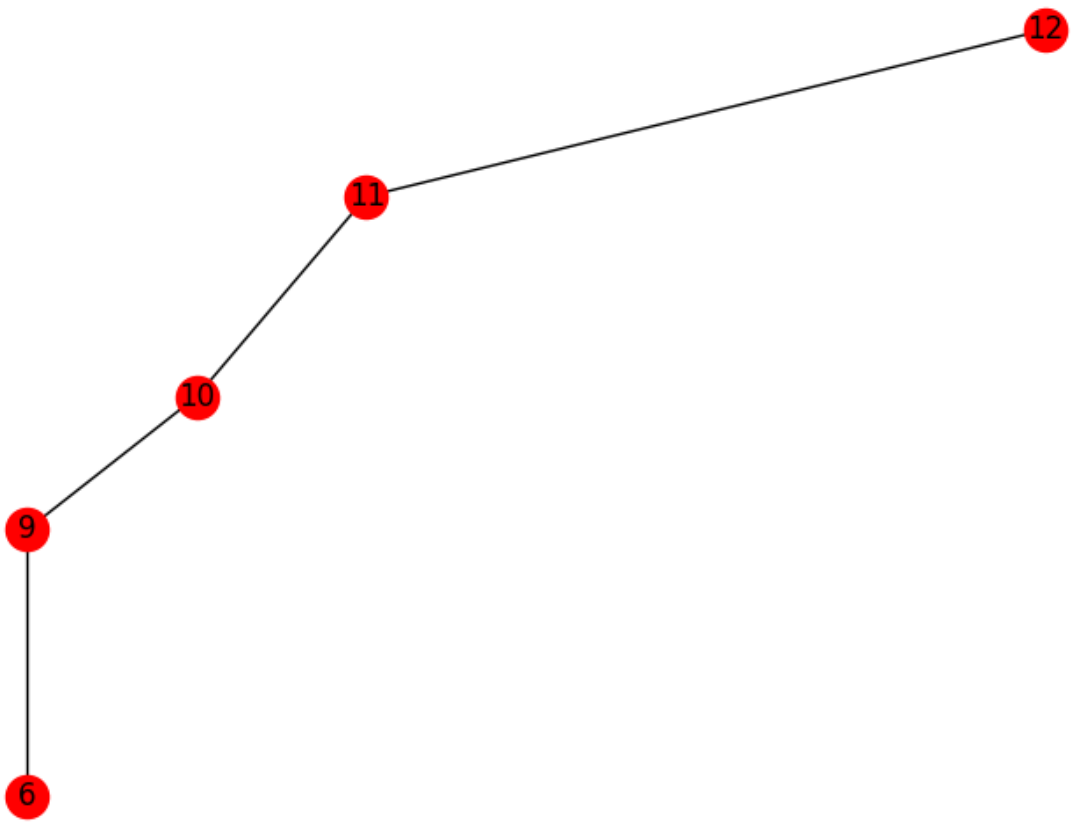


Figure 7.13: Subgraph isolated after executing the query presented in Figure 7.12.

We reconstructed the volume from the query (Figure 7.12) result according to the procedure detailed in section 7.2.4. We present this volume in Figure 7.14 and overlaid the original volume (Figure 7.7) in Figure 7.15.

Using our informatics platform, we calculated the total volume, total surface area, total length for the vasculature isolated by the query (Figure 7.12) as $77.31\,\text{pixel}^3$, $154.62\,\text{pixel}^2$, and $24.61\,\text{pixel}$ respectively.

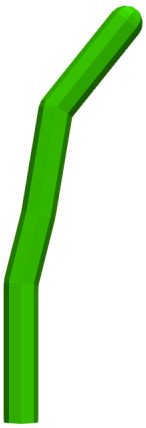


Figure 7.14: Vasculature reconstructed from result of query shown in Figure 7.12.

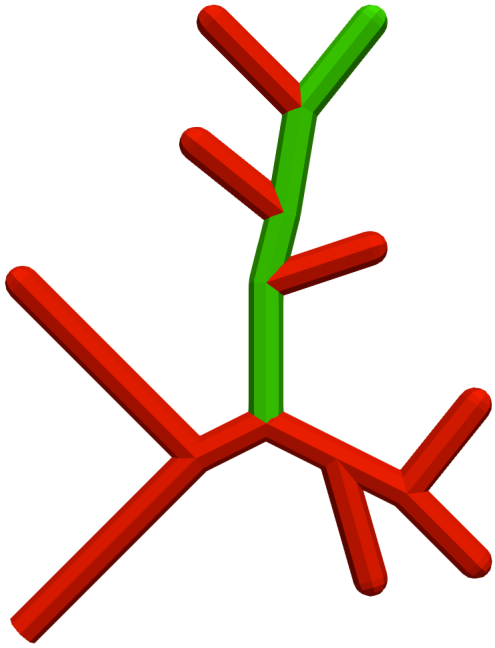


Figure 7.15: Vasculature reconstructed (green) from result of query shown in Figure 7.12 overlaid the original volume (red) (Figure 7.7).

We can also query the database for vessel segments that fit some criteria. For example, perhaps we'd like to look for all LEN:DIA greater than, or equal to, the mean LEN:DIA. We would write the query to that presented in Figure 7.16. This query (Figure 7.16) returns the `FILAMENT` relationships and nodes shown in Figure 7.13. We reconstructed the volume from the result of this query (Figure 7.16) according to the procedure detailed in section 7.2.4. We present this volume in Figure 7.17 and overlaid the original volume (Figure 7.7) in Figure 7.19.

```
MATCH path = ()-[n:SEGMENT]-()
WHERE n.length_to_diameter_ratio >= 3.51
RETURN path, nodes(path) as nodes, rels(path) as rels
```

Figure 7.16: Cypher query used retrieve the `SEGMENTS` with LEN:DIA greater than

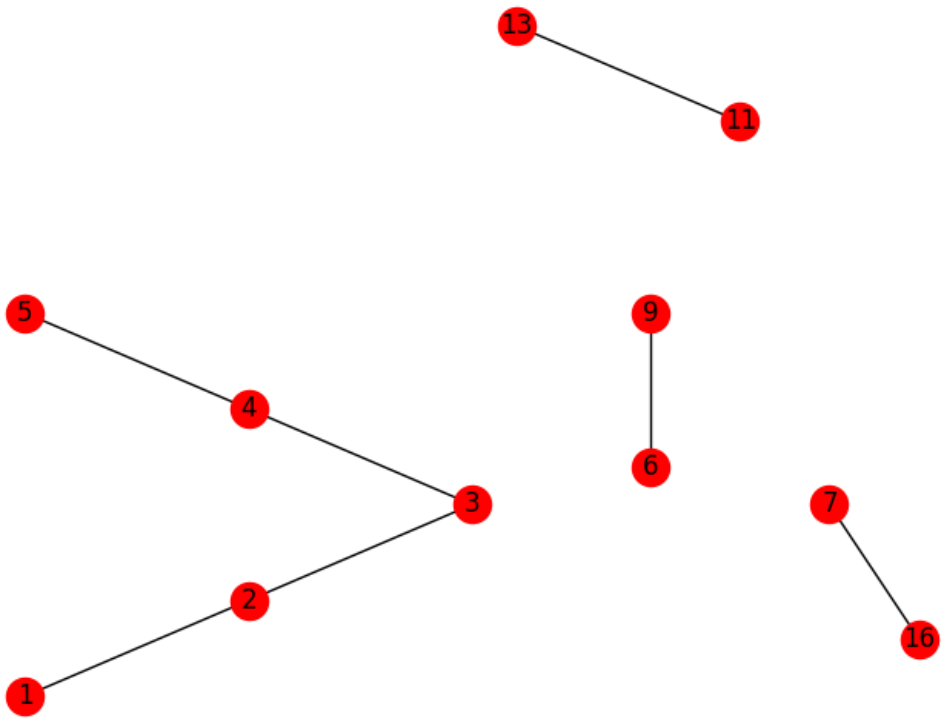


Figure 7.17: Subgraph isolated after executing the query presented in Figure 7.16.

Using our informatics platform, we also calculated the total volume, total surface area, total length for the vasculature isolated by this query (Figure 7.16) as $159.08\,\text{pixel}^3$, $318.15\,\text{pixel}^2$, and 50.64 pixel respectively.

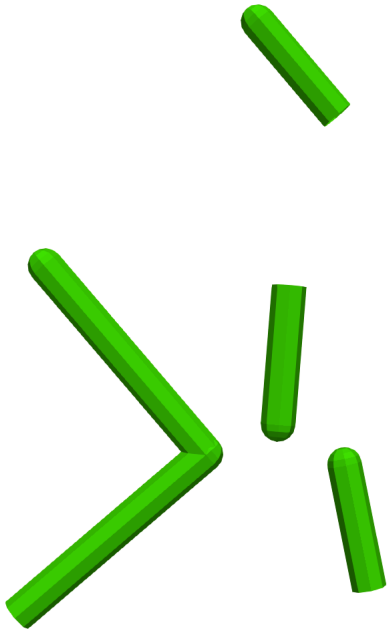


Figure 7.18: Vasculature reconstructed from result of query shown in Figure 7.12.

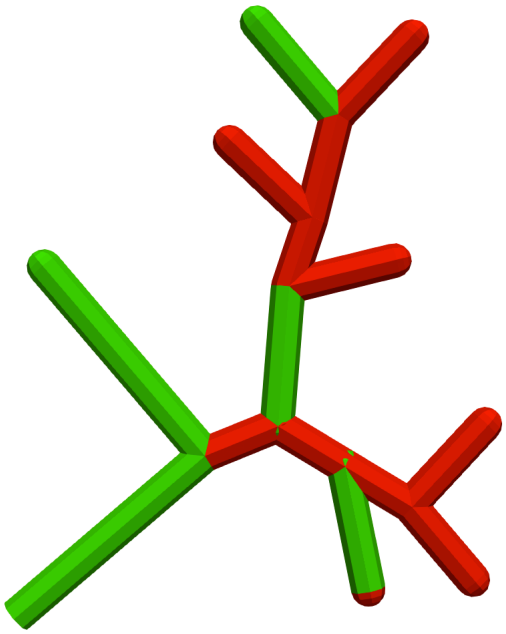


Figure 7.19: Vasculature reconstructed (green) from result of query shown in Figure 7.12 overlaid the original volume (red) as presented in Figure 7.7.

We will now illustrate our depth first traversal that was detailed in section 7.2.3.2 on the graph database storing this digital phantom. We elected to begin all of our traversals from node 6 of the data model (Figure 7.11) and show the results of traversing to different levels, beginning with 1. The subgraph generated by this traversal is shown in Figure 7.20. We reconstructed the volume from the result of this traversal according to the procedure detailed in section 7.2.4. We present this volume in Figure 7.21 and overlaid the original volume (Figure 7.7) in Figure 7.22. Using our informatics platform, we calculated the total volume, total surface area, total length for the vasculature isolated by the query as $53.23\,\text{pixel}^3$, $106.46\,\text{pixel}^2$, and $16.94\,\text{pixel}$ respectively.

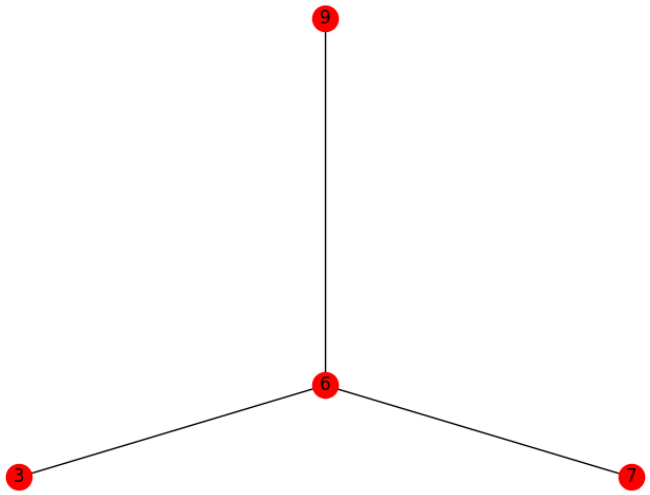


Figure 7.20: Subgraph isolated after executing a depth-first traversal to level 1 from node 6 of the data model (Figure 7.11) to level 1.

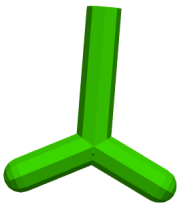


Figure 7.21: Vasculature reconstructed after executing a depth-first traversal to level 1 from node 6 of the data model shown Figure 7.11 of the phantom shown in Figure 7.7.

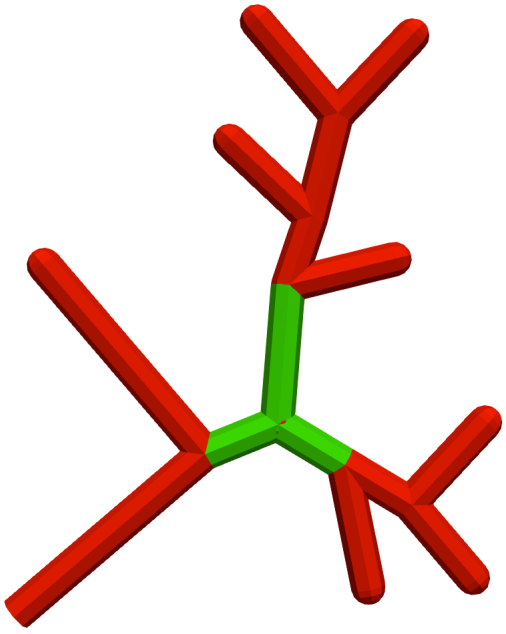


Figure 7.22: Vasculature reconstructed (green) after executing a depth-first traversal to level 1 from node 6 of the data model shown Figure 7.11 overlaid on the original volume (red) as presented in Figure 7.7.

We now will present the results of traversing from node 6 to a level of 2. The subgraph generated by this traversal is shown in Figure 7.23. We reconstructed the volume from the result of this traversal according to the procedure detailed in section 7.2.4. We present this volume in Figure 7.24 and overlaid the original volume (Figure 7.7) in Figure 7.25. Using our informatics platform, we calculated the total volume, total surface area, total length for the vasculature isolated by the query as $211.83\,\text{pixel}^3$, $423.67\,\text{pixel}^2$, and $67.43\,\text{pixel}$ respectively.

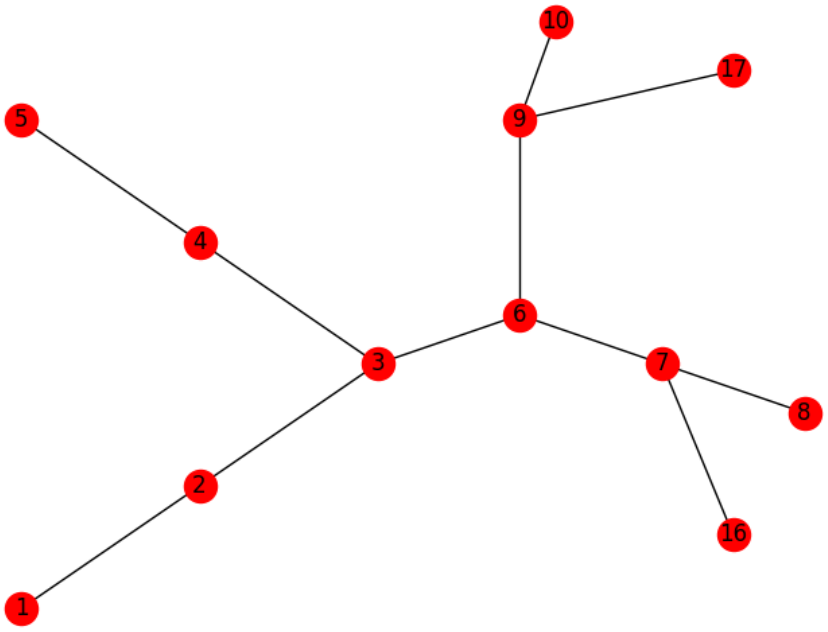


Figure 7.23: Subgraph isolated after executing a depth-first traversal from node 6 of the data model (Figure 7.11) to level 2.

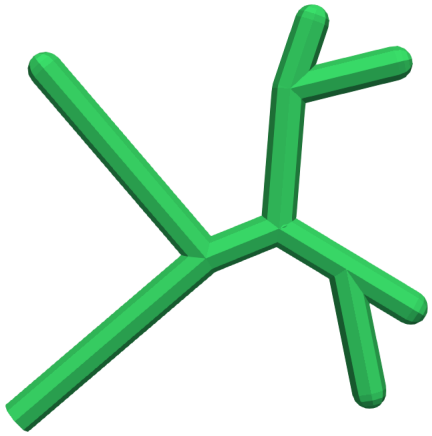


Figure 7.24: Vasculature reconstructed after executing a depth-first traversal to level 2 from node 6 of the data model shown Figure 7.11 of the phantom as shown in Figure 7.7.

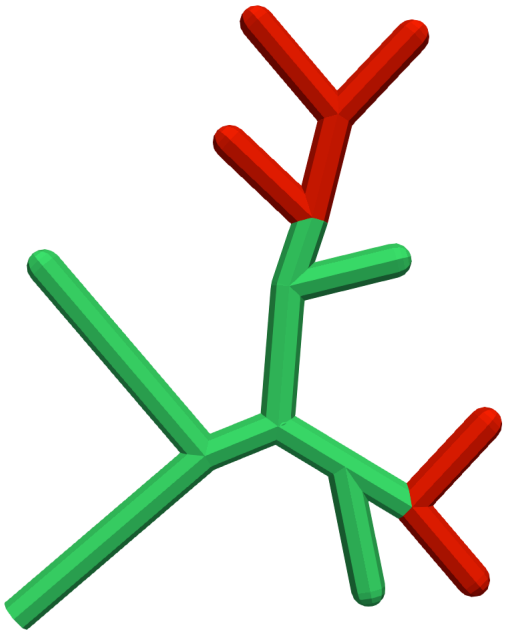


Figure 7.25: Vasculature reconstructed (green) after executing a depth-first traversal to level 2 from node 6 of the data model shown Figure 7.11 overlaid on the original volume (red) presented in Figure 7.7.

Finally, we present the results of traversing from node 6 to a level of 3. This traversal almost captures the entire vasculature of the phantom. The subgraph generated by this traversal is shown in Figure 7.26. We reconstructed the volume from the result of this traversal according to the procedure detailed in section 7.2.4. We present this volume in Figure 7.27 and overlaid the original volume (Figure 7.7) in Figure 7.28. Using our informatics platform, we calculated the total volume, total surface area, total length for the vasculature isolated by the query as $288.95\,\text{pixel}^3$, $577.89\,\text{pixel}^2$, and $91.97\,\text{pixel}$ respectively.

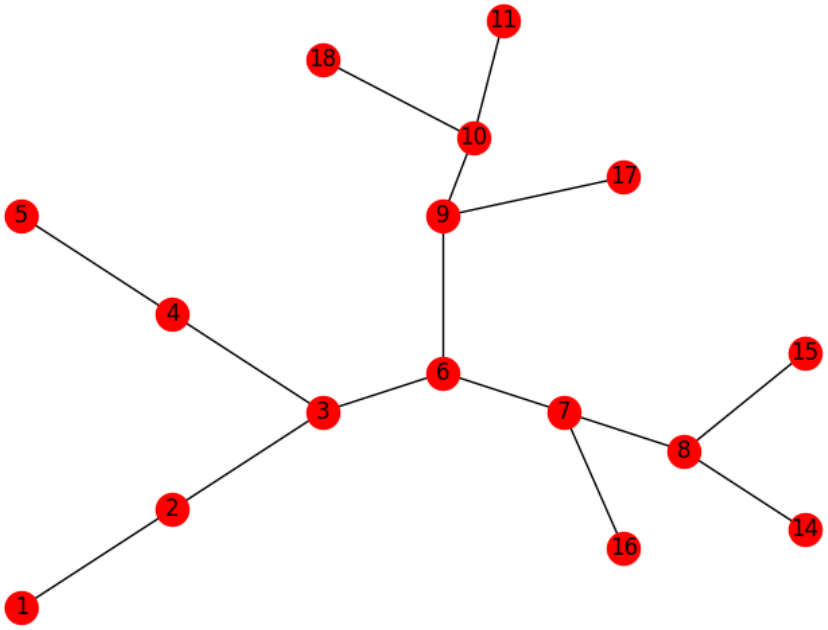


Figure 7.26: Subgraph isolated after executing a depth-first traversal from node 6 of the data model (Figure 7.11) to level 3.

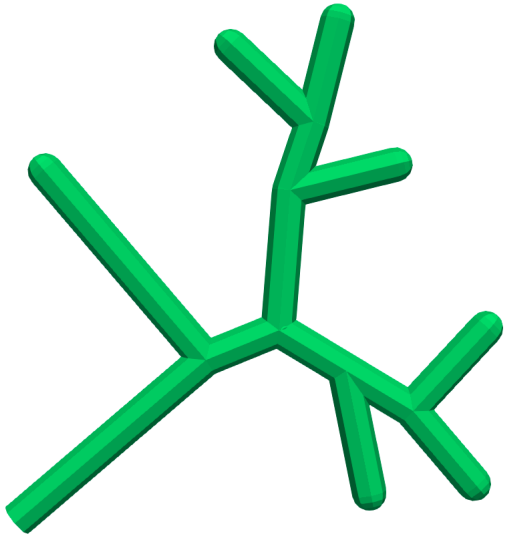


Figure 7.27: Vasculature reconstructed after executing a depth-first traversal to level 3 from node 6 of the data model shown Figure 7.11 of the phantom shown in Figure 7.7.

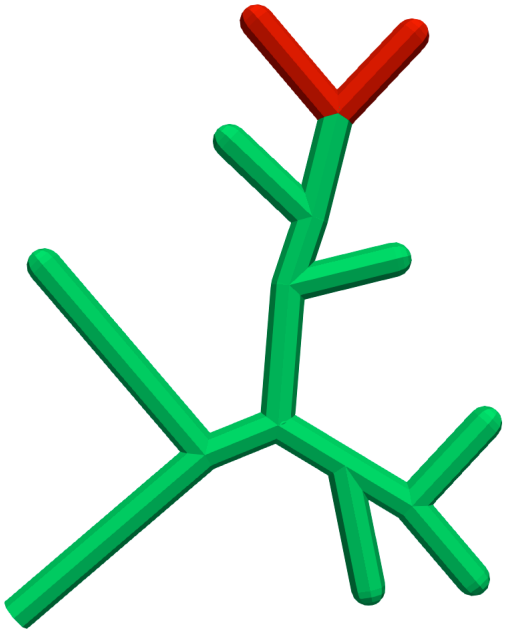


Figure 7.28: Vasculature reconstructed (green) after executing a depth-first traversal to level 3 from node 6 of the data model shown Figure 7.11 overlaid on the original volume (red) as presented in Figure 7.7.

7.3.1.2 Digital phantom #2

The second digital phantom that we use is presented in Figure 7.29. This phantom exhibits a simple loop structure. The result of the graph construction process on this volume, per our procedure detailed in Section 6.2.1, is shown in Figure 7.30. We present the location of these nodes in the phantom volume in Table 7.5, along with the radius measurements at those node positions. The application of our graph reduction algorithm to this "original" graph (i.e., that in Figure 7.30) is shown in Figure 7.32. Each edge in our "reduced" graph is labeled with a vessel segment number that maps to edges composing that vessel segment in the "updated original" graph (shown in Figure 7.31). This allows for the easy reconstruction of the different vessel segments. As stated earlier, to reconstruct the vessel segment denoted by edge 1 in the "reduced" graph (i.e., 7.32), we would derive all edges labeled with 1 from the original graph (i.e., Figure 7.31).

The attributes (as defined in Section 6.2.2.2) for each vessel segment (i.e., vessel arising between branch points/endpoints) in the digital phantom are reported in Table 7.6. The overall characteristics (as defined in Section 6.2.2.1) for the vascular volume follow: total volume, $331.28\,\text{pixel}^3$; total surface area, $662.55\,\text{pixel}^2$; and total length, $105.45\,\text{pixel}$. For the vessel segment attributes, we detail the mean, median, min, and max values in Table 7.7

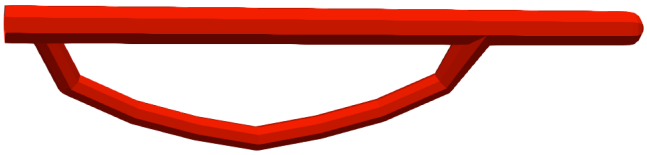


Figure 7.29: Phantom 2 volume.

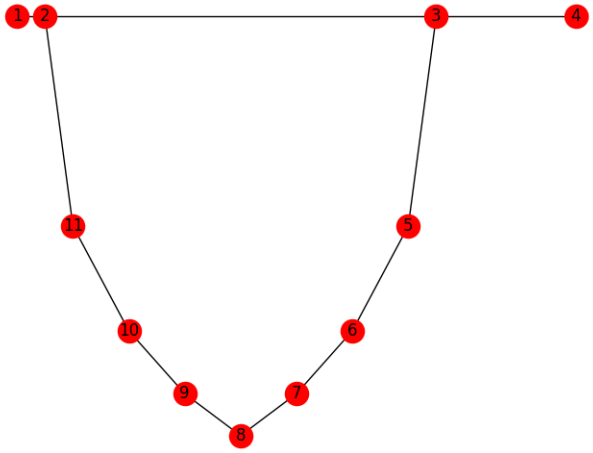


Figure 7.30: Graph representation of phantom 2 volume (Figure 7.29).

Table 7.5: Correspondence of nodes in the graph representation of phantom 2 (Figure 7.30) to their location in the volume (Figure 7.29), along with the radius measurement at that position.

Node No.	***x***	***y***	***z***	**Radius** (pixel)
1	0	0	0	3
2	5	0	0	3
3	75	0	0	3
4	100	0	0	3
5	70	-10	0	2
6	60	-15	0	2
7	50	-18	0	2
8	40	-20	0	2
9	30	-18	0	2
10	20	-15	0	2
11	10	-10	0	2

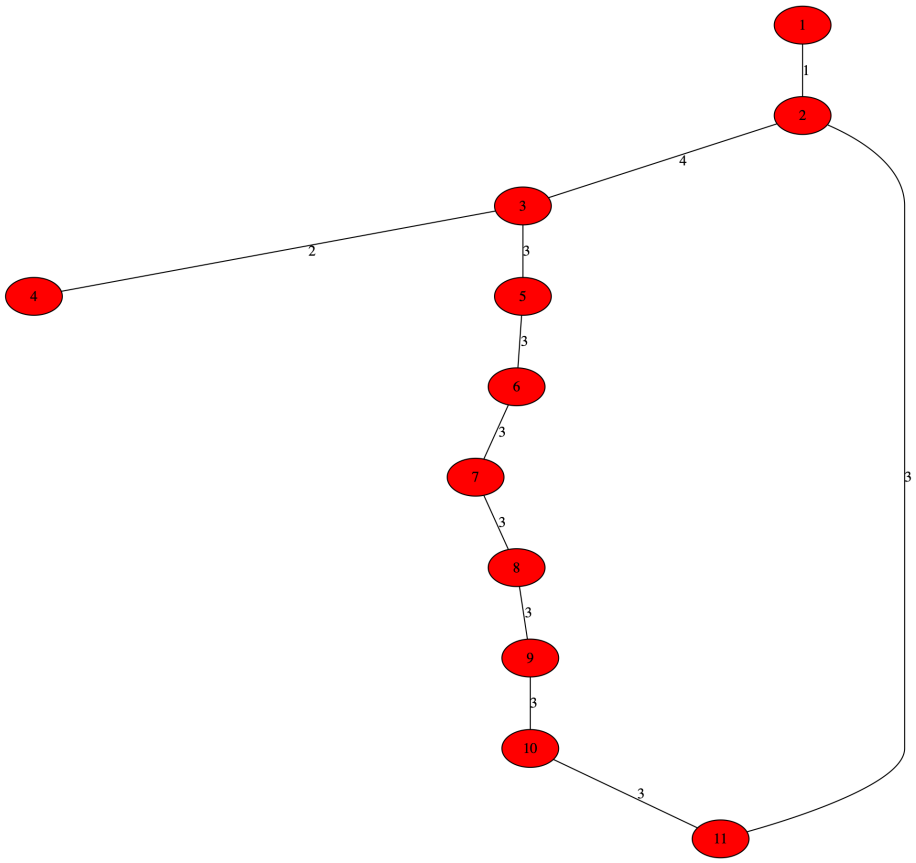


Figure 7.31: Graph representation of phantom 2 (Figure 7.30) updated to contain labeled edges denoting the unique index of the vessel segment that each serves as a filament in.

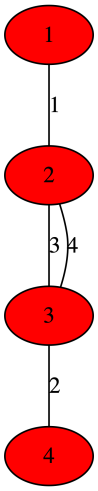


Figure 7.32: "Reduced" graph representation of phantom 2 (Figure 7.30) obtained by algorithm 3. The labels on each edge serve as a unique index for a respective vessel segment presenting in the volume (Figure 7.29).

Table 7.6: Vessel segment attributes (defined in Section 6.2.2.2) calculated across the unique segments composing phantom 2.

Segment ID	**Length** (pixel)	**Surface Area** (pixel^2)	**Volume** (pixel^3)	**Distance** (pixel)	**SA:V**	**LEN:DIA**	**Tortuosity**
1	5.0	94.25	141.37	5.0	0.6667	0.83	1.0
2	25.0	471.24	706.86	25.0	0.6667	4.17	1.0
3	86.0	1150.93	1238.74	70.0	0.9291	14.33	1.23
4	70.0	1319.47	1979.2	70.0	0.6667	11.67	1.0

Table 7.7: Mean, median, minimum, and maximum values for vessel segment attributes calculated for the synthetic model presented in Figure 7.29.

Vessel Attribute	***Mean***	***Median***	***Min***	***Max***
Radius	2.81 pixel	3.00 pixel	2.22 pixel	3.00 pixel
Length	46.50 pixel	47.50 pixel	5.00 pixel	86.00 pixel
Surface Area	758.97 pixel^2	811.09 pixel^2	94.25 pixel^2	1,319.47 pixel^2
Volume	1,016.54 pixel^3	972.80 pixel^3	141.37 pixel^3	1,979.20 pixel^3
Distance	42.50 pixel	47.50 pixel	5.00 pixel	70.00 pixel
Tortuosity	1.06	1.00	1.00	1.22
LEN:DIA	7.75	7.92	0.83	14.33
SA:V	0.73	0.67	0.67	0.93

The Neo4j data model constructed from the graph representation of this phantom is shown in Figure 7.33. In that figure, the nodes are shown as light blue circles, with the `FILAMENT` relationships between them in blue and `SEGMENT` relationships is drawn in red. The labels on those "edges" denotes that relationship's `sid` value. Again, these values are used to map from a `SEGMENT` to its `FILAMENTS`.

We will now illustrate our ability to derive a shortest path between two nodes, with a relationship property as the weight, per our procedure in section 7.2.3.3. We will determine the shortest path between nodes 1 and 4 (see Figure 7.30) using length as the edge weight. After executing algorithm 4 with this query, the subgraph containing the shortest path that was found is presented in Figure 7.34. We reconstructed the volume from the result according to the procedure detailed in section 7.2.4. We present this volume in Figure 7.35 and overlaid the original volume (Figure 7.29) in Figure 7.36. Using our informatics platform, we calculated the total volume, total surface area, total length for the vasculature isolated by the query as $2{,}827.43\,\mathrm{pixel}^3$, $1{,}884.96\,\mathrm{pixel}^2$, and $100.00\,\mathrm{pixel}$ respectively.

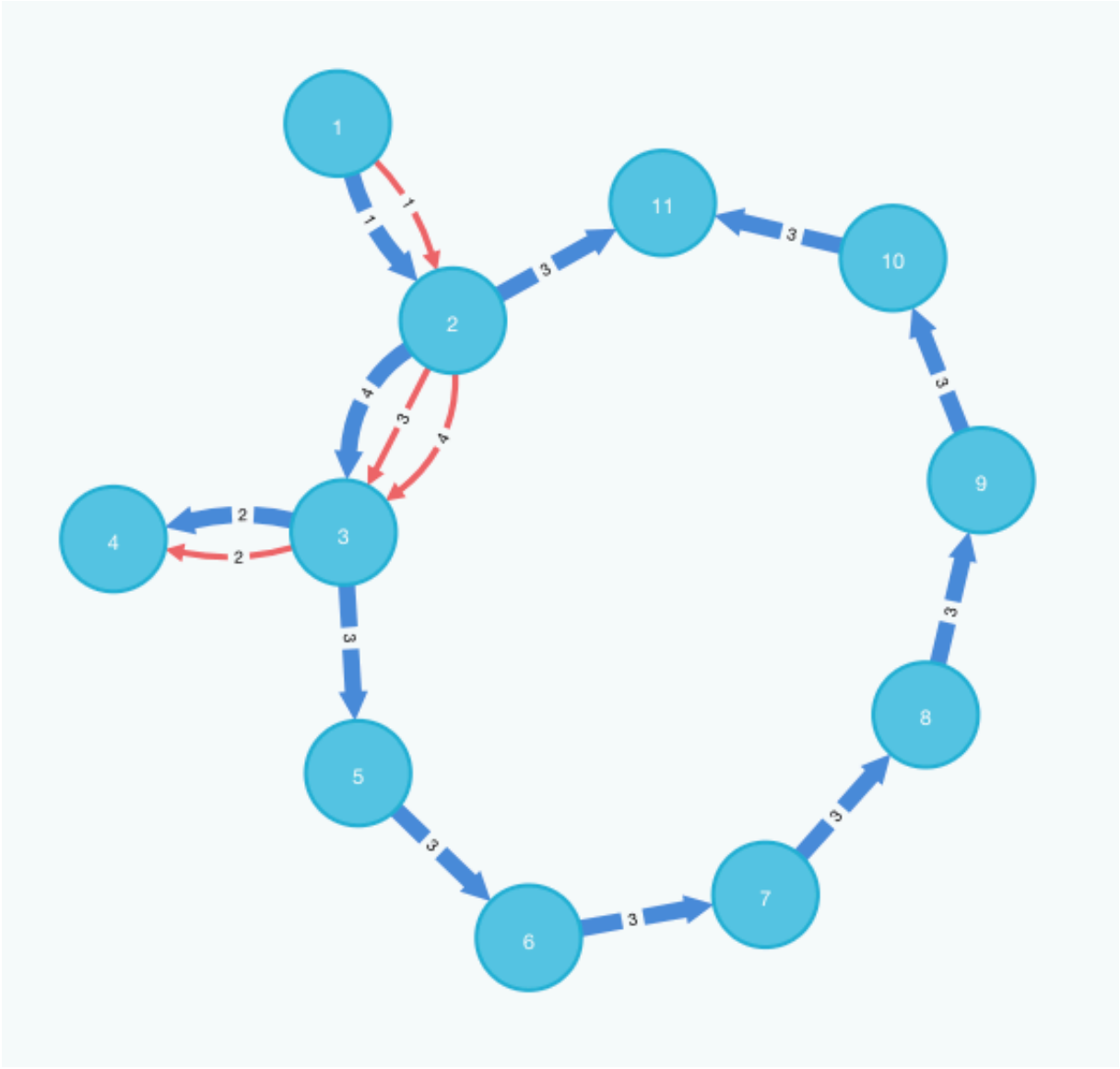


Figure 7.33: Neo4j data model constructed for the phantom shown in Figure 7.29. In this figure, the node light blue circles, with the `FILAMENT` relationships between them in blue and `SEGMENT` relationships in drawn in red. The labels on an "edges" denotes its `sid` value; the labels on a node is its unique ID.

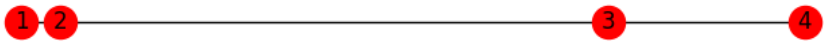


Figure 7.34: Subgraph isolated after finding the shortest path between nodes 1 and 4 (see Figure 7.30) with length as the edge weight.

Figure 7.35: Vasculature reconstructed after finding the shortest path between nodes 1 and 4 (see Figure 7.30) with length as the edge weight.

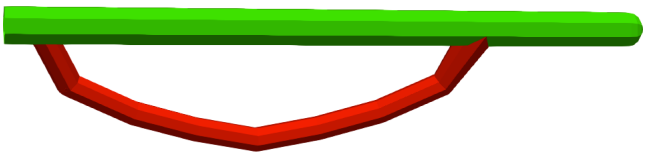


Figure 7.36: Vasculature reconstructed (green) after finding the shortest path between nodes 1 and 4 (see Figure 7.30) with length as the edge weight, overlaid on the original volume (red) as presented in Figure 7.29.

7.3.2 Case study #2, data-driven synthetic model

To further illustrate the utility of our queryable database, we will construct a data model for one of our synthetic models constructed in Chapter 4. The model that we selected for this case study is presented in Figure 7.37 (for full-page image, see Figure 4.5) and was treated as isotropic in resolution at $1.0\,\mu\text{m}$. It is the same model that was selected for analysis using our informatics platform in Chapter 6.

Figure 7.37: Cerebrovasculature model acquired from KESM India Ink dataset. This voxel-based model is $1158 \times 1158 \times 1158$ pixels derived from imaging data acquired at a voxel resolution of $0.6\,\mu\text{m} \times 0.7\,\mu\text{m} \times 1.0\,\mu\text{m}$. The total size of this model is $1.4\,\text{GB}$.

The graph representation for this model was shown previously in Figure 6.38 and its "reduced" graph in Figure 6.39. The overall characteristics for the vascular volume are reported as follows: total volume, $2{,}286{,}872.45\,\mu\text{m}^3$; total surface area, $636{,}994.41\,\mu\text{m}^2$; and total length, $16{,}293.23$ μm. We also found that this volume was comprised of $N = 286$ vessel segments. For the vessel segment attributes, we detailed the mean, median, min, and max values in Table 6.13. We also presented the distributions of these attributes for the vessel segments embedded in the model in Figures 6.30, 6.31, 6.32, 6.33, 6.34, 6.35, 6.36, and 6.37.

The Neo4j data model constructed from the graph representation of this synthetic model is shown in Figure 7.38. In this figure, the nodes are shown as light blue circles, with the `FILAMENT` relationships between them in blue and `SEGMENT` relationships is drawn in red. Once the data model is constructed in the database, we can query the database using algorithm 4. As before, we can reconstruct the result of any supported query using the procedure detailed in section 7.2.4. All of the queries illustrated in Section 7.3.1.1 are applicable to this (and any) data model. We can also write more elaborate queries that isolate vessel segments by attribute.

For instance, if we were interested in all segments whose radius is greater than the median radius of the synthetic volume ($5.08\,\mu\mathrm{m}$) *and* whose length is greater than the median length ($40.67\,\mu\mathrm{m}$), we'd write the query presented in Figure 7.39. This query (Figure 7.39) returns the `SEGMENT` relationships and nodes shown in Figure 7.40. The reconstructed volume is shown overlaid the original volume in Figure 7.41.

Using our informatics platform, we can quantitatively describe the segments we isolated, whose radius is greater than the median radius of the synthetic volume ($5.08\,\mu\mathrm{m}$) and whose length is greater than the median length ($40.67\,\mu\mathrm{m}$). The overall characteristics for the isolated volume are reported as follows: total volume, $1{,}437{,}201.76\,\mu\mathrm{m}^3$; total surface area, $328{,}575.46\,\mu\mathrm{m}^2$; and total length, $6{,}921.51\,\mu\mathrm{m}$. The volume, surface area, and length of the isolated volume comprised 62.85%, 51.58%, and 42.48% of the overall volume, surface area, and length respectively. Descriptive statistics of the vessel segment attributes are presented in Table 7.8.

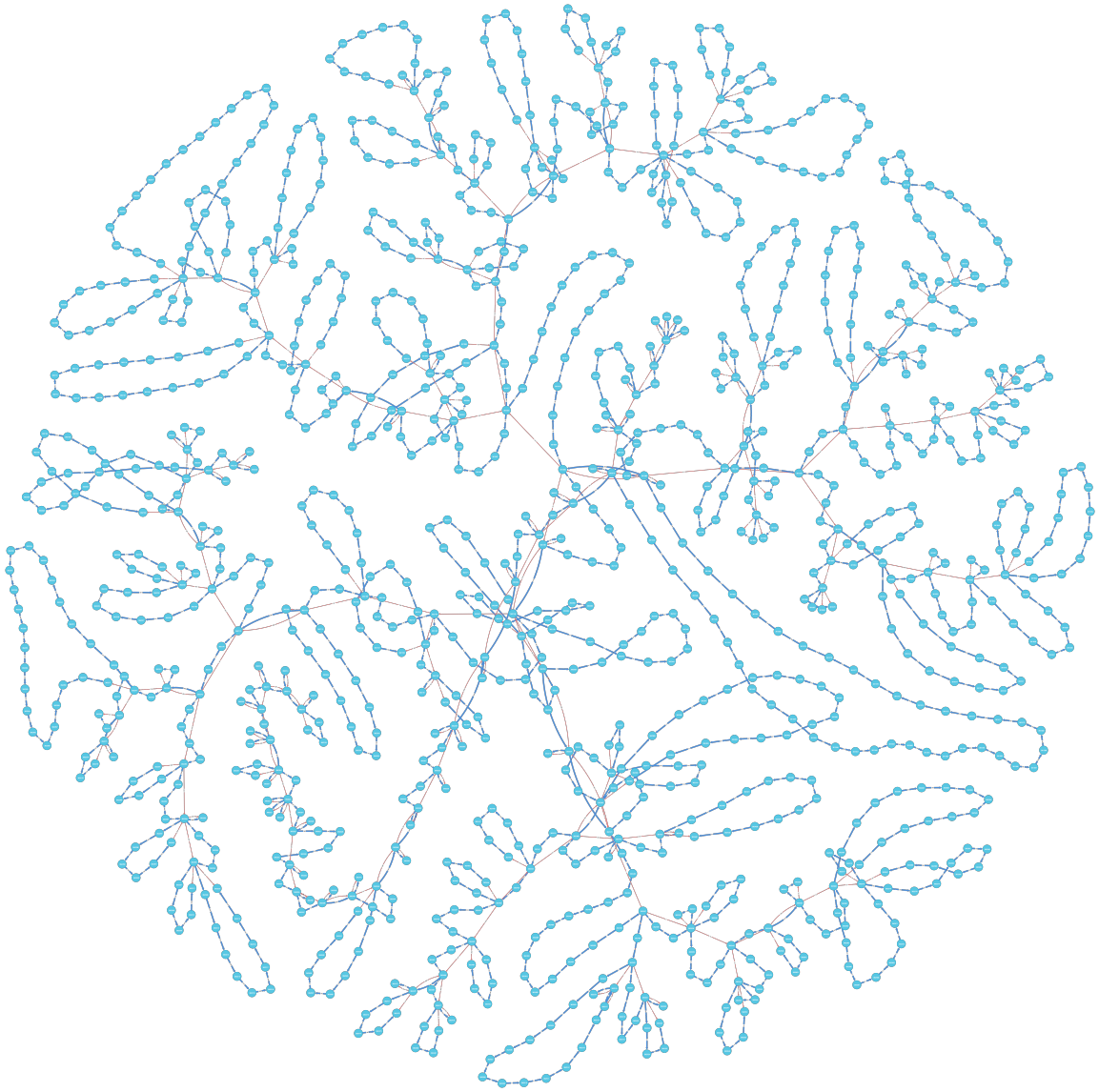


Figure 7.38: Neo4j data model constructed for the synthetic model in Figure 7.37. In this figure, the node light blue circles, with the `FILAMENT` relationships between them in blue and `SEGMENT` relationships in drawn in red.

```
MATCH path = (start:Node)-[f:SEGMENT]-(end:Node)
WHERE  f.avg_radius > 5.08 and f.length > 40.67
RETURN path, nodes(path) as nodes, rels(path) as rels
```

Figure 7.39: Cypher query used retrieve the relationship `SEGMENTS` whose radius is greater than $5.08\,\mu\text{m}$ and length greater than $40.67\,\mu\text{m}$.

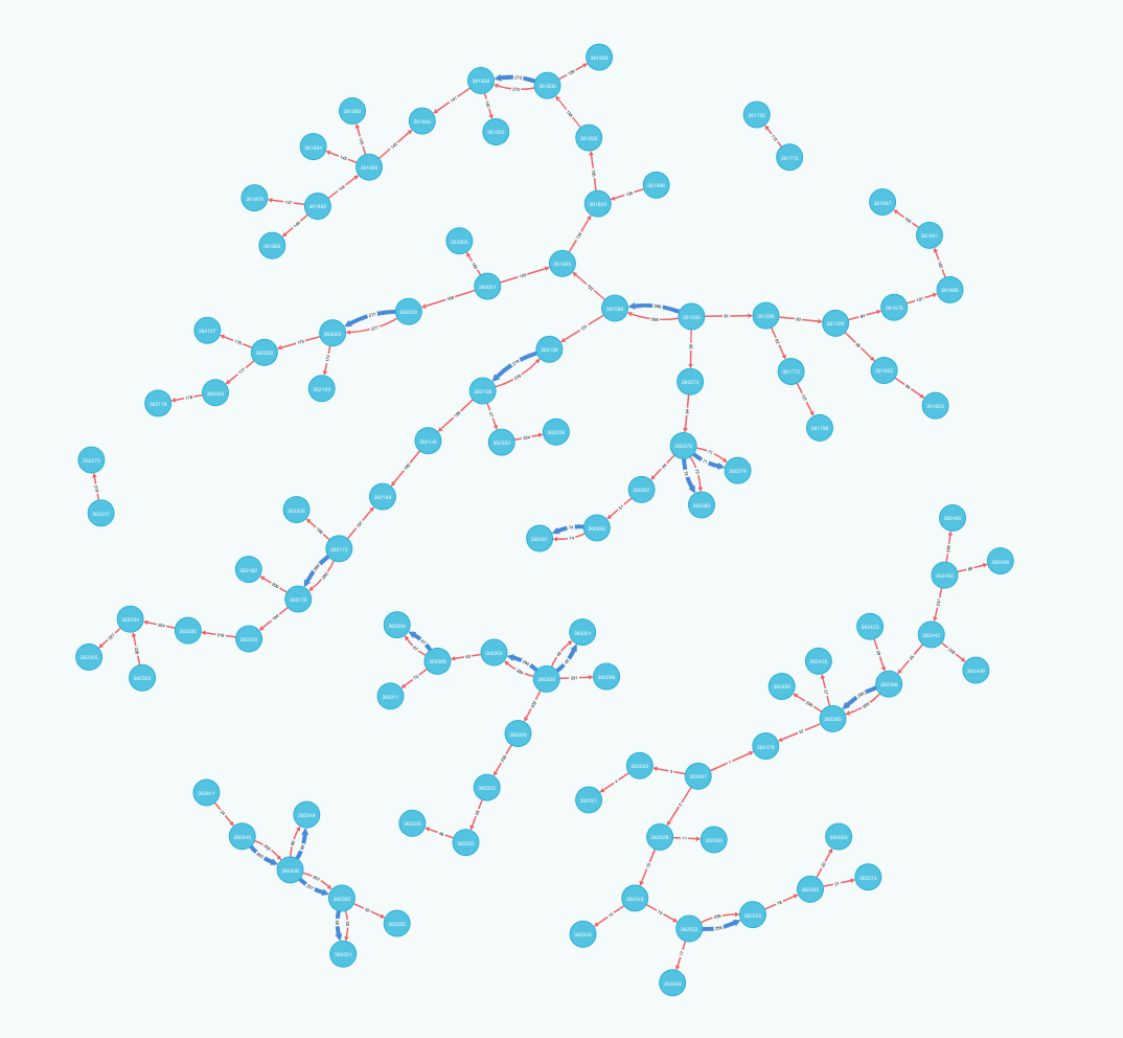


Figure 7.40: Subgraph isolated after executing the query presented in Figure 7.39.

Figure 7.41: Vasculature reconstructed (green) after executing the query presented in Figure 7.39 on the data model. This result is overlaid upon the original volume (red).

Table 7.8: Mean, median, minimum, and maximum values of the attributes calculated for the vessel segments isolated by the query in Figure 7.39 performed on the data model in Figure 7.38.

Vessel Attribute	***Mean***	***Median***	***Min***	***Max***
Radius	$6.93\,\mu m$	$5.91\,\mu m$	$5.09\,\mu m$	$12.33\,\mu m$
Length	$133.11\,\mu m$	$99.43\,\mu m$	$42.75\,\mu m$	$823.25\,\mu m$
Surface Area	$6{,}318.76\,\mu m^2$	$3{,}758.12\,\mu m^2$	$1{,}417.49\,\mu m^2$	$56{,}880.97\,\mu m^2$
Volume	$27{,}638.50\,\mu m^3$	$13{,}600.75\,\mu m^3$	$3{,}640.85\,\mu m^3$	$351{,}110.22\,\mu m^3$
Distance	$98.66\,\mu m$	$76.76\,\mu m$	$17.99\,\mu m$	$466.69\,\mu m$
Tortuosity	1.33	1.06	1.0	5.67
LEN:DIA	10.54	6.08	2.45	60.99
SA:V	0.30	0.33	0.16	0.40

One may be interested in the segments occupying a specific region of the volume instead of vessels exhibiting specific attribute(s). In Figure 7.42, we present a modified Figure 7.37 to include the x-axis with respect to the geometry of the synthetic model. We calculated the mid-point of a bounding-box encompassing that geometry along its x-axis as 920.84 based on a minimum x-value of 346.47 and maximum x-value of 1495.20. To illustrate our ability to extract all `FILAMENT` segments within a specific region, we decided to query for all `FILAMENT`s whose nodes have x-coordinates less than 920.84. We write the query presented in Figure 7.43 to accomplish just that. After executing algorithm 4 with that query, the subgraph returned by the graph database is shown in Figure 7.44. The reconstructed volume is shown overlaid the original volume in Figure 7.45.

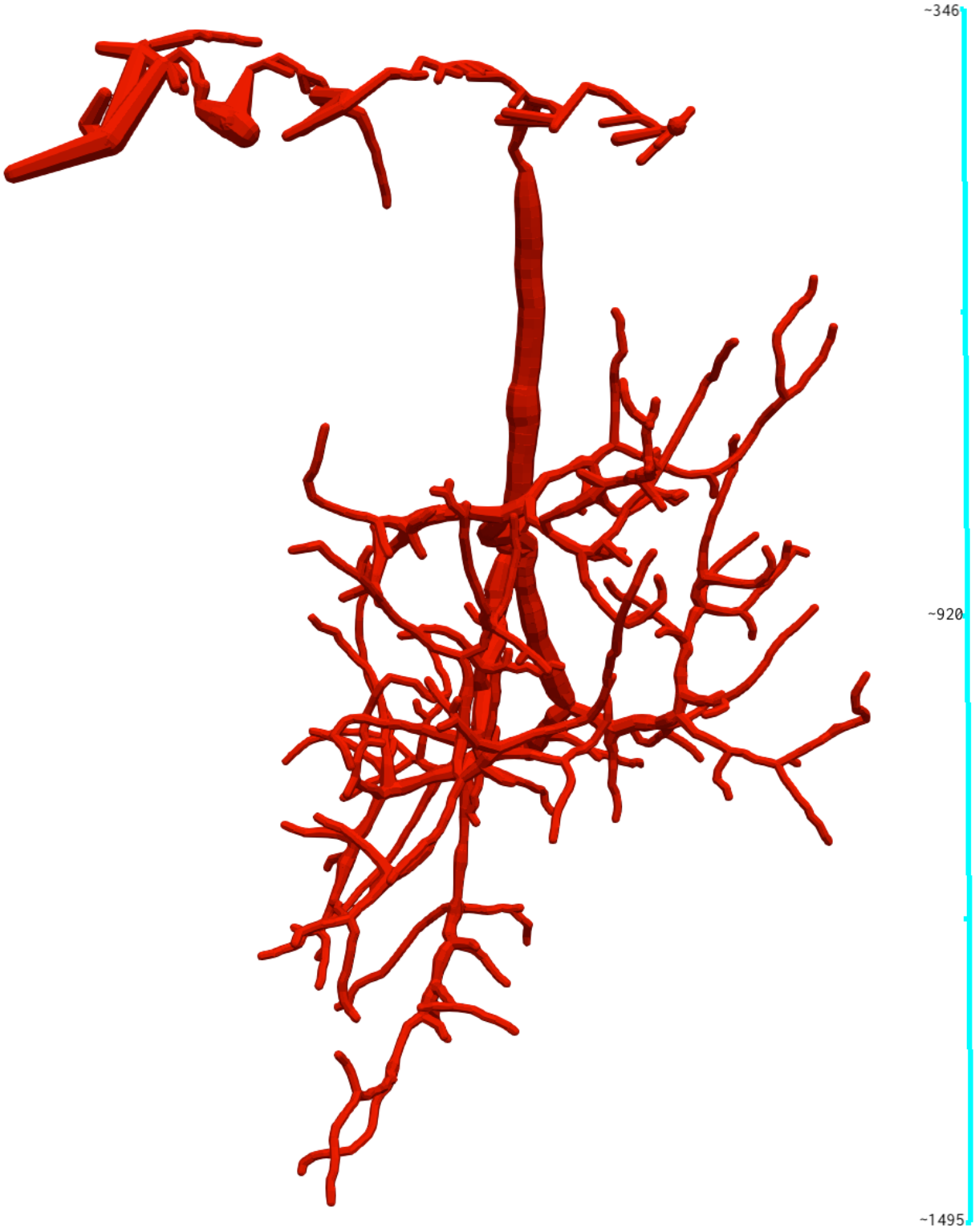


Figure 7.42: We present a modified Figure 7.37 to include the x-axis (shown in cyan) with respect to the geometry of the synthetic model.

```
MATCH path = (n1:Node)-[f:FILAMENT]-(n2:Node)
WHERE n1.x <= 920.84 and n2.x <= 920.84
RETURN path, nodes(path) as nodes, rels(path) as rels
```

Figure 7.43: Cypher query used retrieve all `FILAMENT`s whose endpoint's x-coordinates are less than 920.84.

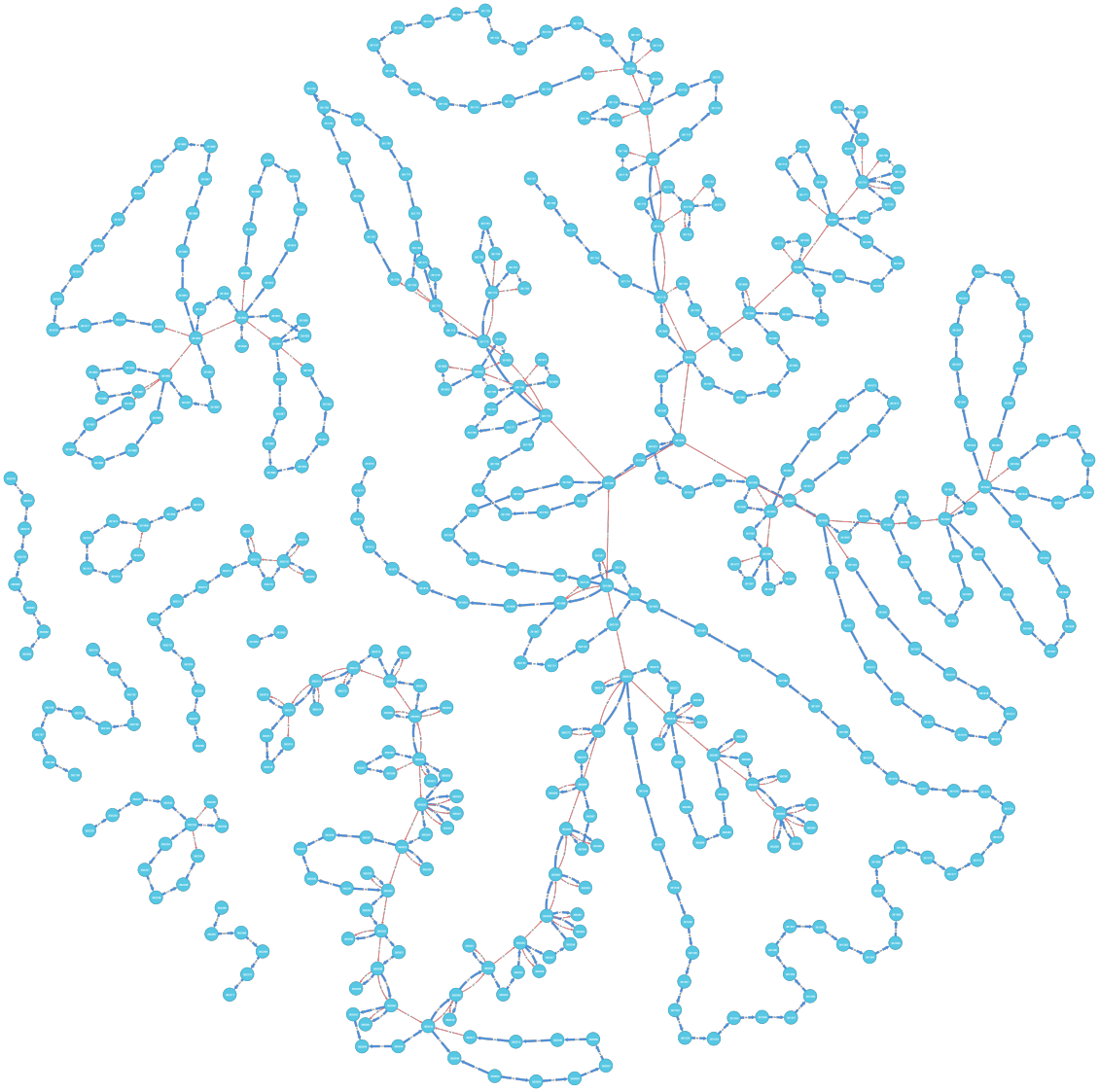


Figure 7.44: Subgraph isolated after executing the query presented in Figure 7.43.

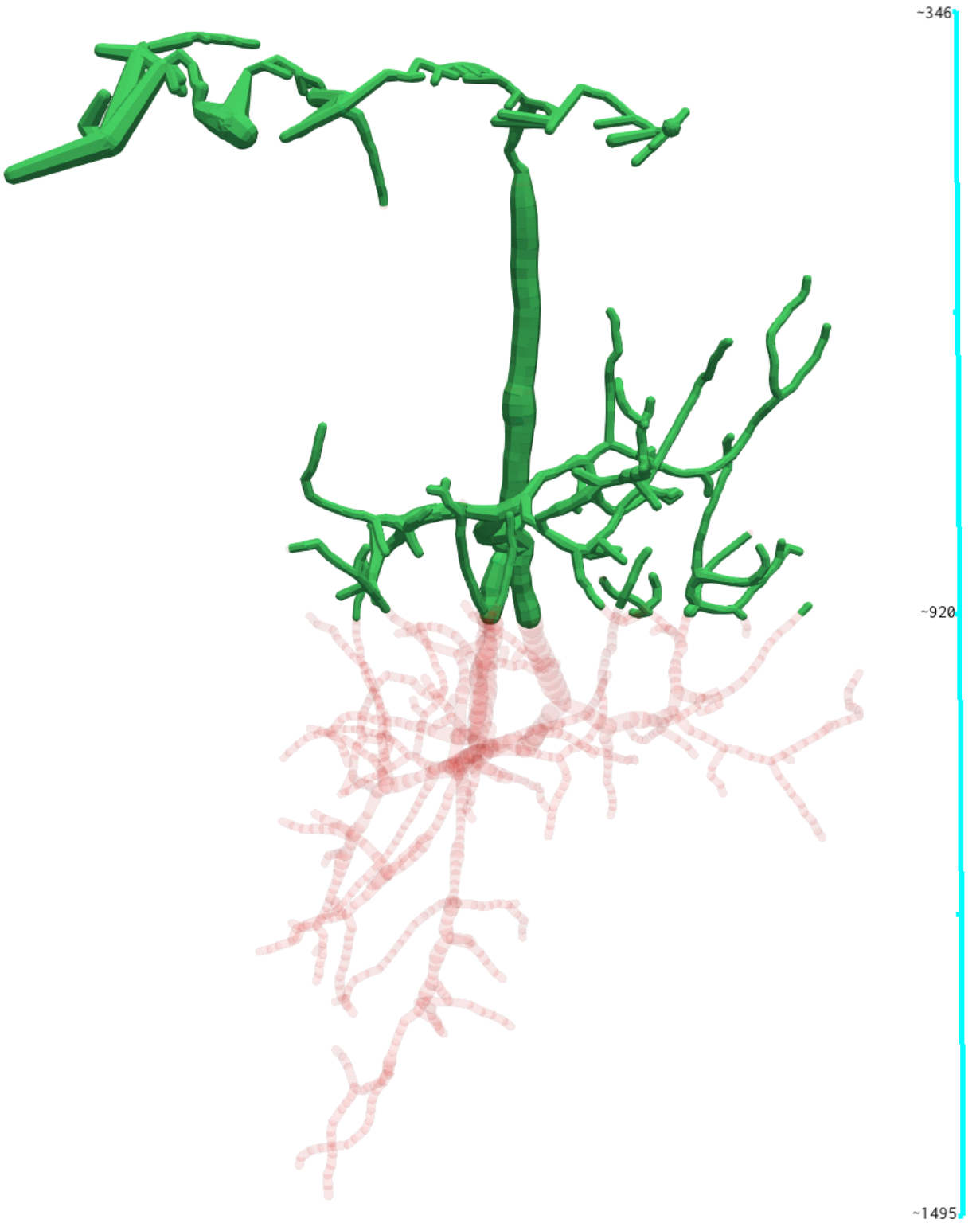


Figure 7.45: Vasculature reconstructed (green) after executing the query presented in Figure 7.43 on the data model. This result is overlaid upon the original volume (red). The x-axis with respect to the underlying geometry of the synthetic model (and not our viewing perspective) is shown in cyan.

Using our informatics platform, we can quantitatively describe the filaments that were returned by this query (i.e., Figure 7.43). The overall characteristics for the isolated filaments are reported as: total volume, $1{,}222{,}739.68\,\mu m^3$; total surface area, $308{,}978.47\,\mu m^2$; and total length, $7{,}435.96$ μm. The volume, surface area, and length of the isolated volume comprised 53.47%, 48.51%, and 45.64% of the overall volume, surface area, and length respectively. Descriptive statistics of the vessel segment attributes are presented in Table 7.9.

Table 7.9: Mean, median, minimum, and maximum values of the attributes calculated for the vessel segments isolated by the query in Figure 7.43 performed on the data model in Figure 7.38.

Vessel Attribute	***Mean***	***Median***	***Min***	***Max***
Radius	$6.00\,\mu m$	$4.99\,\mu m$	$4.8\,\mu m$	$16.11\,\mu m$
Length	$51.28\,\mu m$	$36.95\,\mu m$	$4.20\,\mu m$	$474.25\,\mu m$
Surface Area	$2{,}130.88\,\mu m^2$	$1{,}253.56\,\mu m^2$	$126.59\,\mu m^2$	$38{,}630.19\,\mu m^2$
Volume	$8{,}432.69\,\mu m^3$	$3{,}640.85\,\mu m^3$	$303.83\,\mu m^3$	$269{,}709.58\,\mu m^3$
Distance	$47.60\,\mu m$	$35.18\,\mu m$	$4.20\,\mu m$	$450.14\,\mu m$
Tortuosity	1.06	1.01	1.00	1.56
LEN:DIA	4.44	3.28	0.42	27.25
SA:V	0.36	0.40	0.12	0.42

7.4 Summary

The contributions of the work detailed in this chapter are multifaceted. First, we implement the necessary infrastructure to translate a graph representation of vascular connectivity and geometry into a Neo4j graph database model. At the time of writing, and to the best of our knowledge, this is the first time that vascular topological and geometrical data has been warehoused in a highly optimized database for graph data models. Second, we provide a platform to interact with that data. We offer users a means to execute a user-defined query on the database, as well as the ability to perform a depth-first traversal or determine the shortest path between two points (for any vessel attribute). Queries for vessel segments that meet certain criterion (morphometry, taxonomy, type, etc.) and/or that are within a specific scale (as organization and morphometry vary depending on

size) are possible as user-defined queries. It is important to note that there are many user-defined queries that can be executed than those listed in this chapter. Such queries include (but are not limited to) pathfinding algorithms (such as single-source shortest path and all-pairs shortest path), measures of centrality, and community detection algorithms. We refer the reader to the Neo4j documentation on how to construct these (and many other) queries. Finally, our platform translates the result of any of these queries returning a path to a networkx graph representation. The networkx graph representation integrates with our informatics platform (chapter 6) for quantitative analysis and also affords seamless reconstruction of the vasculature in vector graphics format.

These contributions are important in cerebrovasculature research. Using a graph database to store cerebrovasculature topological and geometrical database allows researchers to query the database for vascular segments with some attribute. For instance, all segments with a radius measurement less than x, but greater than or equal to y, could be retrieved, and our informatics platform could be applied to quantify the vessels within that range. Alternatively – or in conjunction with constraint(s) on segment attribute(s) – we could query the database for vascular filaments presenting within some region of the brain. This would allow vessels presenting within an ROI – or within some distance of an anatomical marker of interest – to be isolated. These vessels could then be quantified using our informatics platform to characterize the cerebrovasculature within a given area.

We expect the frameworks and platforms developed in this chapter and the last (chapter 6) to have a significant scientific impact. Specifically, we believe that they will open the door for data-driven, quantitative investigation of the brain's vascular system. The analytical metrics provided for quantitative analysis coupled with high-performance graph analytics and reconstruction ability in these resources will aid in the discovery of novel biomarkers of processes underlying healthy and diseased brains. We believe that these resources will contribute a standardized means for the quantitative study of whole-brain vascular datasets acquired using modern imaging techniques.

8. WHOLE-BRAIN CEREBROVASCULATURE DATABASE AND ANALYSIS

8.1 Motivation

To better understand the development and function of cerebral vasculature in the brain, it is essential to map out its anatomical structure in the healthy brain and compare differences in the disordered mind. For instance, local variations in vascular morphometry can have a profound effect on the neuronal composition in surrounding areas. Therefore, charting the alterations in cerebrovasculature across the brain will help elucidate how these changes affect the neuronal structure. Datasets acquired by the modern technology described in Section 2.2 will provide the means necessary for data-driven, quantitative investigations at the whole-brain scale.

The Brain Networks Laboratory at Texas A&M University has developed a high-resolution 3D microscopy method capable of imaging whole small animal brains at sub-micrometer resolution The Knife-Edge Scanning Microscope (KESM). KESM fills the critical gap in neuroimaging between large scales, lower-resolution methods such as magnetic resonance imaging and small-scale, higher-resolution methods like Serial Block-Face Scanning Electron Microscopy. KESM imaging can achieve $0.7\,\mu\mathrm{m}$ resolution laterally and $1\,\mu\mathrm{m}$ axially, sufficient to perform accurate morphometry of neuronal and vascular networks. This imaging modality (KESM) enables the reconstruction of key microstructures, such as the full vascular network of the mouse brain, for quantitative analysis across the whole small animal organ [39].

The KESM has acquired the complete cerebrovasculature system of the C57BL/6J mouse model [13]. This dataset comprises $15{,}580 \times 12{,}100 \times 8{,}560$ pixels, at a resolution of $0.6 \times 0.7 \times 1.0\ \mu\mathrm{m}^3$. In his Ph.D. dissertation [6], Junseok Lee reconstructed this dataset: he split it into ten parts, used Otsu thresholding followed by the voxel scooping method [9] to the extract filaments embedded in each part. Lee measured diameter and length across the whole-volume on binarized image raster.

In this chapter of my dissertation, I plan to extend the whole-brain results reported by [6]. I will

accomplish this using my informatics platform (Chapter 6), which includes additional quantitative metrics describing the vascular segments. I will also construct a whole-brain graph database of the cerebrovasculature embedded within [6]'s dataset (using my method from Chapter 7). Some of the questions motivating this work to follow:

- How accurate is the segmentation algorithm that was used to extract the whole-brain cerebrovasculature data in [6]?
  - Validation of the voxel scooping method was performed on an expert-labeled volume containing a single three-dimensional neuron [9].
- How does the vascular morphometry present across the whole brain?
  - Are the diameter and length measurements reported in [6] reproducible?
  - What are the attributes of the capillaries (Diameter $\leq 10\,\mu\mathrm{m}$), medium-sized vessels ($10\,\mu\mathrm{m} <$ Diameter $\leq 20\,\mu\mathrm{m}$), medium-large-sized vessels ($20\,\mu\mathrm{m} <$ Diameter ≤ 40 $\mu\mathrm{m}$), and large-sized vessels ($40\,\mu\mathrm{m} <$ Diameter)?

Studies into the brain's angioarchitecture have long focused on the cerebral cortex, where the morphometry of vessels residing near the cortical surface and in the intracortical region have been characterized (ref: Section 2.1.7). Meanwhile, investigations into the angioarchitecture of subcortical nuclei, such as the basal ganglia, have not been examined to the same extent. The whole-brain analysis proposed in this section will be amongst the first studies to characterize the cerebrovasculature across the entire brain, from the cerebral cortex to that of the nuclei residing deep within. This is extremely important as it is necessary to understand cerebrovasculature organization within the healthy brain to know how it changes within the diseased brain. Since local variations in vascular morphometry have a profound effect on the neuronal composition in surrounding areas, a connectivity map of the vascular variations across the entire brain will be beneficial to cerebrovasculature research. Our database will provide such a connectivity map (topology), along with the

geometry and attributes noted for every vessel segment presenting within the KESM cerebrovasculature dataset.

8.2 Methodology

8.2.1 Dataset acquisition

We received the completed segmentation of the KESM cerebrovasculature data from [6]. As mentioned in Section 8.1, this dataset comprises $15{,}580\,\text{pixel} \times 12{,}100\,\text{pixel} \times 8{,}560\,\text{pixel}$ at a resolution of $0.6\,\mu\text{m} \times 0.7\,\mu\text{m} \times 1.0\,\mu\text{m}$. Due to the computational complexity of segmenting filaments from this large-scale dataset, [6] elected to split the volume into ten parts and perform the segmentation for each part individually. We will refer you to [6] for the precise methodology; however, we do summarize it briefly in Section 8.2.2. Ultimately, what we've received from [6] are ten `SWC` files describing the filament structure for each part and ten vector graphics files that store geometric primitives illustrating the cerebrovasculature of each part.

8.2.2 Compose single, connected whole-brain dataset

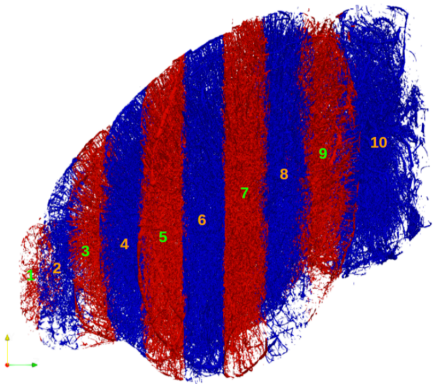


Figure 8.1: The whole-mouse brain, as segmented by [6] into 10 parts. Reprinted, with permission, from [6].
(←: Anterior, →: Posterior)

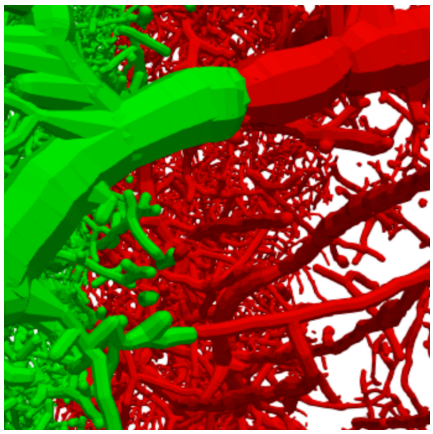


Figure 8.2: Close-up of the boundary between two of the adjacent parts shown in Figure 8.1. The green vasculature occupies the part anterior to the part containing the red vasculature. Reprinted, with permission, from [6].

The vascular filaments embedded in the KESM whole-mouse brain cerebrovasculature dataset [39] were extracted by [6]. This was accomplished by splitting the volume into ten parts, which are shown in Figure 8.1. The voxel scooping method [9] was then applied on a part-by-part basis to extract the embedded vasculature. [6]'s reported measures of vascular diameter and length were then calculated across each part individually. Therefore, vessels traversing multiple parts were not considered in full as they were partitioned into two vessels (one in each part) at the part-boundary divide. The fact that the whole-brain was partitioned into parts means that even by merging the parts, the vector graphics encodings of vessels along the boundary will be disconnected (and encoded as two separate cylinders). We illustrate this disconnectedness in Figure 8.2, where we've opened two raster files encoding adjacent regions of the brain. In Figure 8.2, the green vasculature occupies the part anterior to the part containing the red vasculature. We can see that although they align, they are not connected but instead are generated from two unique sets of geometric primitives. To conduct a more comprehensive analysis, we need to connect up the vasculature presenting along the boundaries of adjacent parts into single encoded segments. That is, we need to go from two separate cylinders at the boundary into one continuous cylinder across that boundary. This will ensure that our informatics platform calculates the attributes for the entire segments and not piecewise data for "two segments" that are one. The problem with the piecewise approach is that it skews the descriptive statistics for metrics such as length, surface area, volume, etc. as the calculated segment attributes would be incorrect. To illustrate this, imagine that we have a segment that is split equally between parts and left unconnected. If the overall length is 10 units, we would observe two segments of length 5 units when we really should have observed a single segment of length 10 units. Therefore, we must resolve the issue of having two segments that should be analyzed as one continuous segment. Our procedure for constructing a connected whole from these disconnected parts follows.

Before proceeding, we require that the adjacent parts have the vasculature topology and geometry stored as a series of nodes and edges. We selected the *SWC* format for this encoding [102] and elaborate on this scheme in Section 4.2.1. In short, the idea behind this format is that each

vascular segment (vessel between two bifurcations) can be embedded as a series of line segments. The filaments are stored in the *SWC* format as a hierarchy of line segments for every connected vascular component in the dataset.

We subsequently construct a graph representation of the two adjacent parts according to our procedure in section 6.2.1, ensuring that the node `IDs` are unique between and within the graphs. The next step is to calculate the bounds of the encoded vessels along the axis of partition. In this case, [6] partitioned the whole-brain along the z-axis. We will refer to the graph containing nodes with the smaller maximum z-axis value as `Gs` and that with the larger minimum z-axis value as `Gl`. We calculate the largest observed z-axis value for `Gs` and the smallest observed z-axis value for `Gl`, and take the difference. We then offset the z-axis of every node in `Gl` by that difference. This process closes any potential gaps between the adjacent parts. This gap corresponds to about four units, where we define a unit as being the z-axis length of a voxel.

After this adjustment, we construct a k-D tree `t` initialized with the nodes from `Gl` that are within four units of the boundary; the k-dimensions are the x-, y-, and z- coordinates. Next, we construct a list `l` of nodes from `Gs` that are within four units of the boundary. We then compose `Gs` and `Gl` into a single graph `G`. For each node `n1` in `l`, we look up the node `n2` closest to it in `t`, and create a new edge in `G` between `n1` and `n2`. We store `G` as the "stitched" result of the two adjacent parts under consideration. We will use this result as `Gs` for the next part. That is, once we've stitched parts 1 and 2 together, we will stitch the $1-2$ result with 3. Following this process for all of the parts, we compose a single, connected whole-brain dataset.

8.2.3 Quality assessment of the whole-brain dataset

The next step towards quantitative analysis of the whole-mouse brain cerebrovasculature dataset [39] is the validation of the segmentation algorithm used to extract the vascular filaments embedded within it. In [6], the voxel scooping method [9] was used. The validation of this algorithm was conducted on a single neuron [i.d.]. To assess the quality of the methods used by [6] to extract the cerebrovasculature, including the accuracy of the segmentation algorithm, we will first inspect the dataset qualitatively. If the dataset appears coherent and the vascular connectivity looks reason-

able, we will follow-up this qualitative assessment with a quantitative study using our model-based system (detailed in Chapter 5) against [6]'s extraction methodology of the cerebrovasculature from the imaging volume.

8.2.4 Whole-brain analysis

After stitching the segmented parts from [6] together and into a single, connected dataset (Section 8.2.2) and assessing the quality of the dataset, we will apply our informatics platform to the stitched result per our procedure in Chapter 6 and construct a graph database according to Chapter 7. The application of our informatics platform will extend the results obtained by [6] by incorporating additional attributes calculated for each vessel segment in our analysis. We will then construct a graph database of the whole-brain cerebrovasculature, which will provide a means for storage and dissemination of those results. We will verify [6] results against our own and also determine the implications of using a single, connected dataset (done here) as opposed to performing analysis on a part-by-part basis (as completed by [6]). In toto, we will report descriptive statistics for the following vessel segment attributes across the whole-brain KESM cerebrovasculature dataset prepared by [6]:

- Radius
- Length
- Tortuosity, using the distance metric defined in [87]
- Distance (between beginning and terminal end of segment)
- Surface area
- Volume
- Surface-area-to-volume ratio (i.e., SA:V)
- Length-to-diameter ratio (i.e., LEN:DIA)

We will also report the following vascular volume characteristics:

- Total length
- Total surface area
- Total volume
- Total-surface-area-to-total-volume ratio

Thereafter, using our database we will focus our study to describe these characteristics and attributes within the: capillaries (Diameter $\leq 10\,\mu\mathrm{m}$), small-sized vessels ($10\,\mu\mathrm{m} <$ Diameter $\leq 20\,\mu\mathrm{m}$), medium-sized vessels ($20\,\mu\mathrm{m} <$ Diameter $\leq 40\,\mu\mathrm{m}$), and large-sized vessels ($40\,\mu\mathrm{m} <$ Diameter). Vessels meeting these criteria will be isolated through the relevant database queries and following the procedures detailed in Chapter 7. We will analyze the results using our informatics platform described in Chapter 6.

8.2.5 Validation against piecemeal data in the literature

While our whole-brain analysis will report new results with no comparable data at the whole-brain scale, we will compare our results against the relevant piecemeal data in the literature. These sources will include the comparison of the morphological results obtained from our whole-brain against those reported in the region of the Middle Cerebral Artery by [92], [29], and [112]. We will also include the relevant cerebrovasculature characteristics reported by [2] in the forebrain and cerebellum, along with that of the neocortex and cerebellum by [7]. Such 'validation' will ensure that our method computes results in the 'right ballpark.'

8.2.6 Dissemination of connectivity data

We will disseminate the raw whole-brain cerebrovasculature connectivity data online, through the Neo4j graph database [100]. To do this, we first make the database read-only and then disable authentication (so that anyone can access it, but no one can write to it). We also limit the time that a query can run. Interactions with the connectivity data are provided through the Neo4j Browser, which provides a means of visualizing the nodes and their relationships across the dataset. Queries

will be issued in Cypher, Neo4j's query language, on the graph data model. The results of a query will be displayed as a visual graph or in tabular form. Those who wish to interact more intimately with our data will be able to do so by using our informatics platform (Chapter 6), which supports database queries and the analysis and reconstruction of query results.

8.3 Results

8.3.1 Compose single, connected whole-brain dataset

The KESM whole-brain cerebrovasculature dataset was partitioned along the z-axis of the imaging plane as shown in Figure 8.1. For each of these parts, our procedure calculated the minimum and maximum z-axis of a bounding box encompassing the data. These values are shown in Table 8.1. Based on this data, our method calculated the offset between each partition as 4 vector graphic units. Accordingly, as our procedure went through each part, the correct multiple of 4 vector graphic units was used to offset the part being composed with everything already composed. For a refresher on our composition method, please see Section 8.2.2.

Table 8.1: Minimum and maximum z-axis bounds for each part (as partitioned in Figure 8.1) of the KESM whole-brain cerebrovasculature dataset.

Segment Number	Min (z-axis bound)	Max (z-axis bound)
1	800	1596
2	1600	2396
3	2400	3196
4	3200	3996
5	4000	4796
6	4800	5596
7	5600	5996
8	6000	6396
9	6400	7196
10	7200	7996

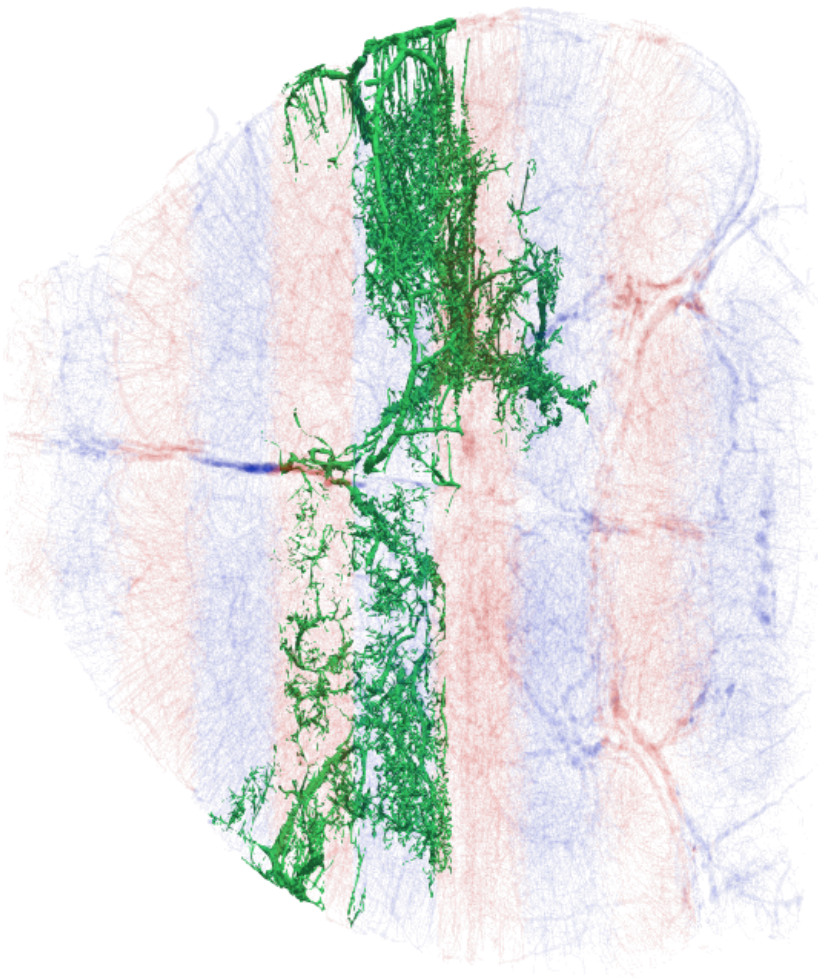


Figure 8.3: Largest connected component (presented in green) found after stitching [6]'s parts (Figure 8.1) into a connected whole along the part-boundary divide. This component is overlaid upon the original, individual parts (shown in red and blue).

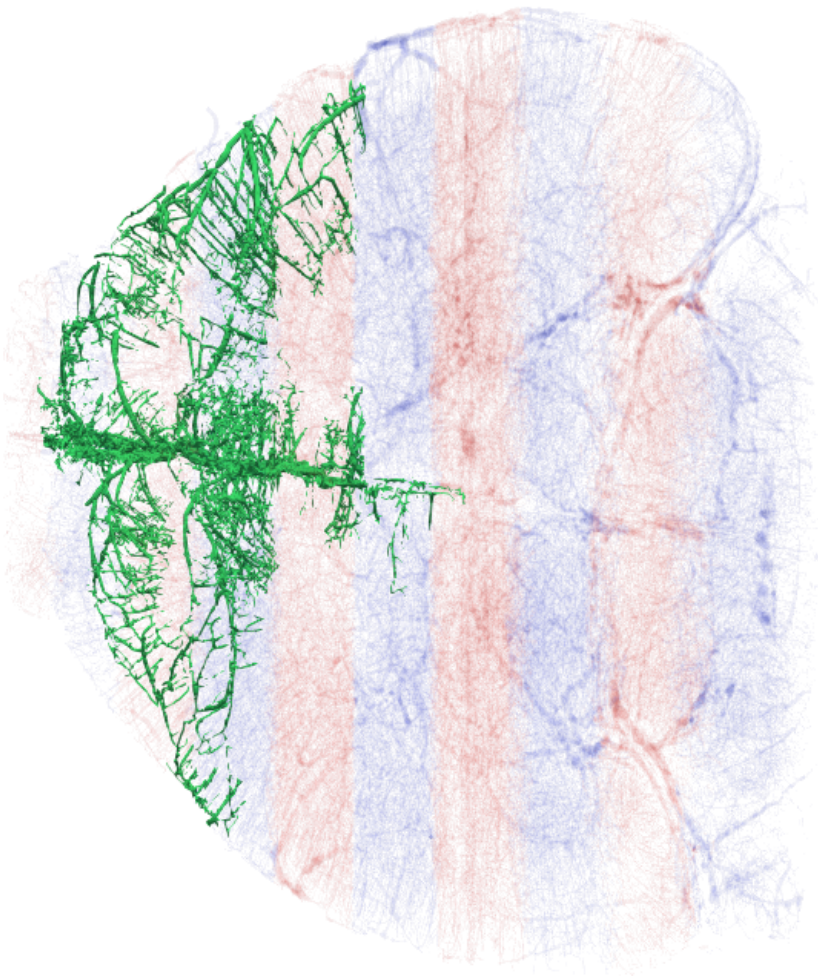


Figure 8.4: Second largest connected component (presented in green) found after stitching [6]'s parts (Figure 8.1) into a connected whole along the part-boundary divide. This component is overlaid upon the original, individual parts (shown in red and blue).

While one would expect the volume to be a single, fully connected component at this stage, this is not the case: there are many discontinuities of cerebrovasculature presenting across the volume, causing the volume to be composed of many connected components. We will address this later in Section 8.3.2. The focus here is that both of these components span multiple parts, thereby illustrating that our composed result stitches the individual parts together across the part-boundary divide.

8.3.2 Quality assessment of the whole-brain dataset

The cerebrovasculature acquired from a C57BL/6J mouse by Mayerich et al. ([13]) was extracted by Lee ([6]) using a series of morphological operations followed by the application of the voxel scooping method [9]. The morphological operations transformed the raw images slices into a sequence of binary images contrasting the foreground (i.e., vessels) from the background. Application of the voxel scooping method simply extracted the foreground from these binary images, an easy procedure given the sharp contrast. Lee's procedure is broadly described in [6]; however, not specific enough to implement their procedure. Thus, we could not replicate [6]'s process to quantitatively assess the quality of the dataset using our model-based validation system detailed in Chapter 5; however, this was not needed.

During our qualitative assessment, we visually inspected the dataset. We found the dataset to be extremely discontinuous, with isolated segments displaced throughout the volume. This spaghetti-like appearance is shown in Figure 8.5; the quality of the dataset is not very good from our perspective. While it is hard to determine the amount of loss across this dataset, we can confidently state where it has come from. The raw images acquired by the KESM are imperfect, with various artifacts disrupting the continuous nature of the cerebrovasculature. This has been recognized since the advent of the device [13].

Compounding the imaging artifacts are the results of the morphological processes carried out by [6]. The downsampling and upsampling of the dataset inevitably resulted in some data being drowned out. Furthermore, the application of different filters to the raw images, including the median filter, would have also resulted in some data loss. In Chapter 5, we illustrated that even

with the smallest median filter there is a degeneration of quality of the surface; the smallest filter that we tested (3×3) resulted in a 20% reduction in the true positive rate for the voxel scooping method [9] from baseline. Therefore, due to imaging artifacts in the raw images and the side effects of the morphological processes that were applied to those images, we expect that this dataset is missing a large percentage of data. We understand that false positives may "make up" for some of the low true positive rate contributing to the vessel isosurfaces. We imagine this scenario taking place when a vessel pixel is labeled as the background, but a background pixel along the periphery is also absorbed into the vessel. These two competing factors could balance one another out. The problem with this is that the density of the vessel is no longer centered where it should be, and this can have profound implications when considering blood flow to the surrounding tissues (see Section 2.1).

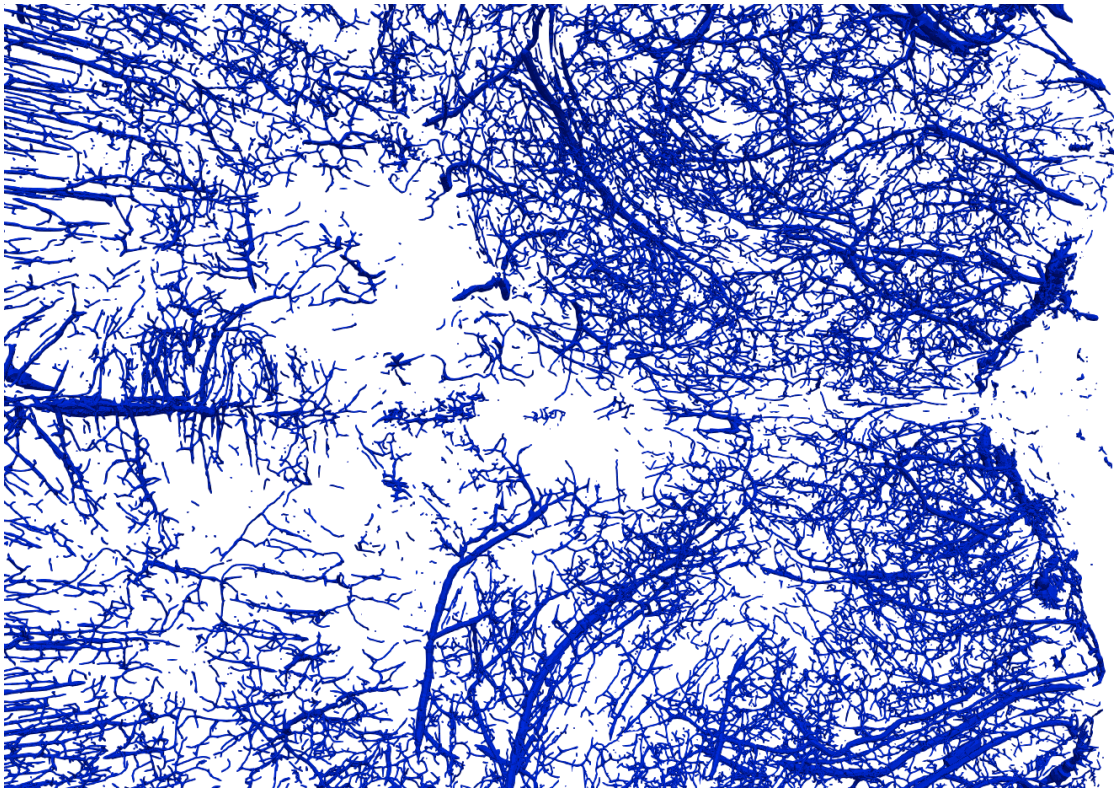


Figure 8.5: View into arbitrarily selected region of part six (Figure 8.1) of [6]'s segmentation of the KESM cerebrovasculature dataset [13]. Throughout the volume, vessel discontinues are clearly present.

8.3.3 Whole-brain analysis

8.3.3.1 Whole-brain graph representation

We constructed the graph representation of the connected whole-brain dataset (that we composed) according to our procedure detailed in Chapter 7. The construction of the graph `G` containing the geometry of the cerebrovasculature went smoothly and was completed within minutes. Recall, `G` encodes the twists and turns of each vessel segment, and is frequently used during the reconstruction of vascular segments of interest. Furthermore, that we reduce this `G` to a graph `G_reduced` which has an edge for each individual vascular segments presenting in the dataset. `G_reduced` stores the attributes of each segment as properties within that segment's edge. We say that `G_reduced` also encodes the topology, or connectivity, of the cerebrovasculature. The construction of `G` resulted in a graph with $1{,}750{,}823\,\mathrm{nodes}$ and $1{,}616{,}258\,\mathrm{edges}$. The size of the graph `G` serialized was $264\,\mathrm{MB}$. We present an illustration of `G` in Figure 8.6 and zoomed into an arbitrary region of that graph in Figure 8.7

In Chapter 6, we illustrate the construction of `G_reduced` according to Algorithm 3. The application of Algorithm 3, which was written without parallelization and multi-threading. Due to the massive scale of our cerebrovasculature dataset, this algorithm would run for days without finishing had we applied it to `G`. Therefore, we decided to write a graph reduction algorithm that "solves" each segment in parallel while simultaneously calculating the attributes for that segment. We achieved this by writing Algorithm 7, which "solves" each segment in a dedicated thread. The construction of `G_reduced` resulted in a graph with $635{,}219\,\mathrm{nodes}$ and $500{,}654\,\mathrm{edges}$. The size of the graph `G_reduced` serialized was $198\,\mathrm{MB}$. We present an illustration of `G_reduced` in Figure 8.8 and zoomed into an arbitrary region of that graph in Figure 8.9. The overall appearance of this graph is redder by nature of its reduction in nodes that exposes the red coloring of its edges. This is in contrast to the whole-brain graph (Figures 8.6 and 8.7), where blue coloring dominates due to its high density of blue nodes (encoding the geometry) covering its red edges.

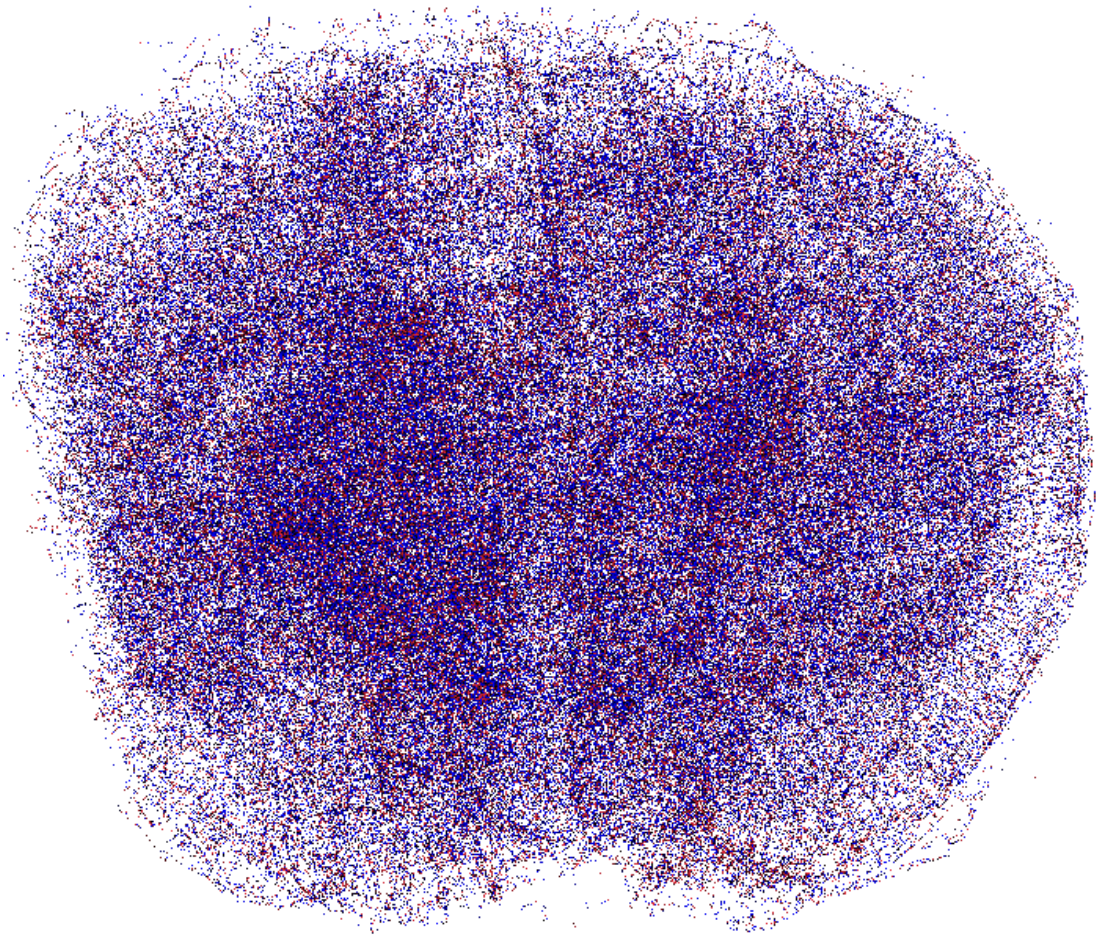


Figure 8.6: Illustration of the graph encoding the geometry of the cerebrovasculature of the mouse. We've elected to place the nodes (blue) and edges (red) at their position in the imaging volume and draw the graph across three-dimensions.

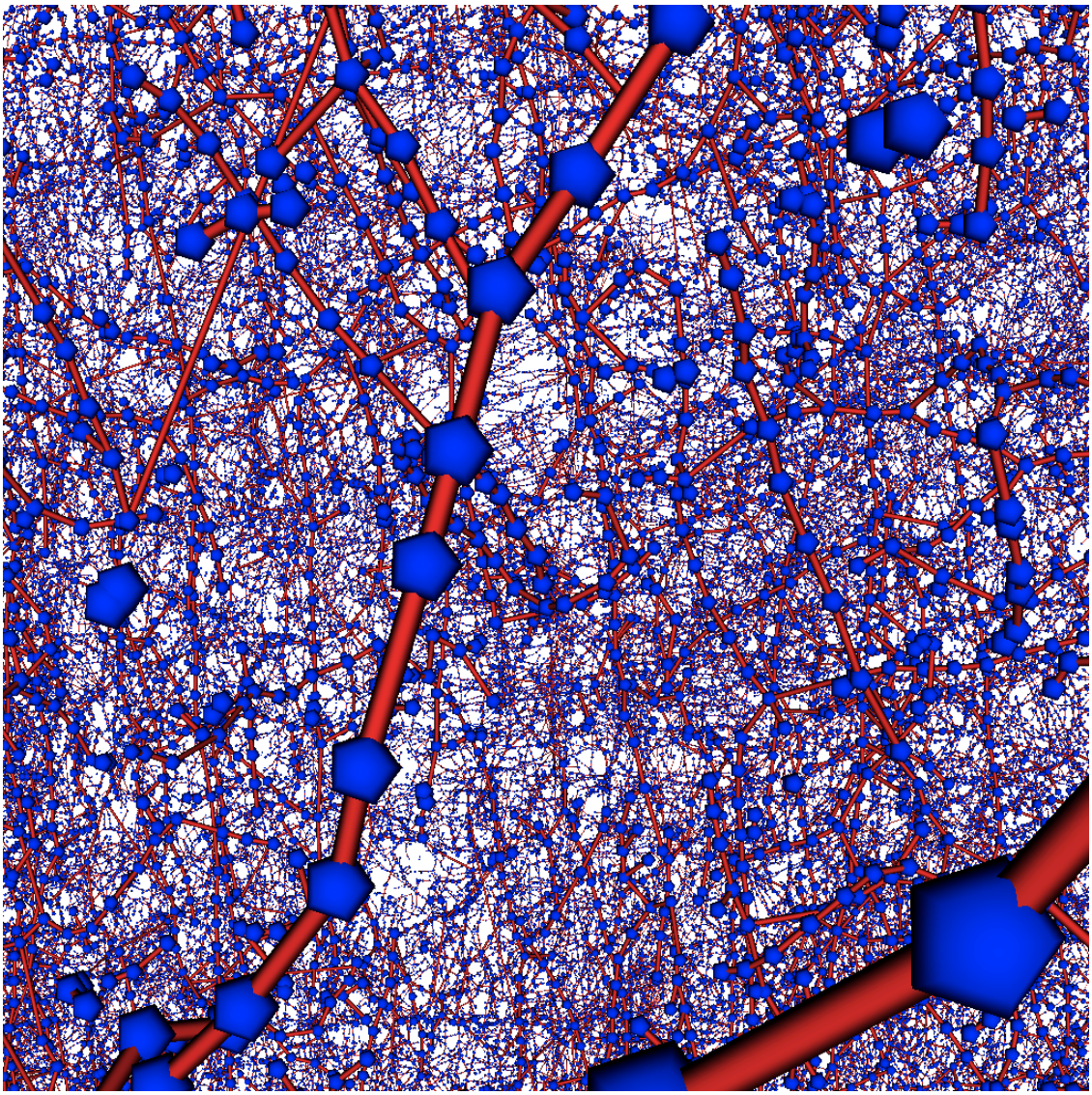


Figure 8.7: Zoomed perspective into an arbitrarily selected region of the whole-brain graph from Figure 8.6. The nodes (blue) and edges (red) in this graph can be used during reconstruction as they encode the individual filaments comprising the twists and turns of the vessel segments.

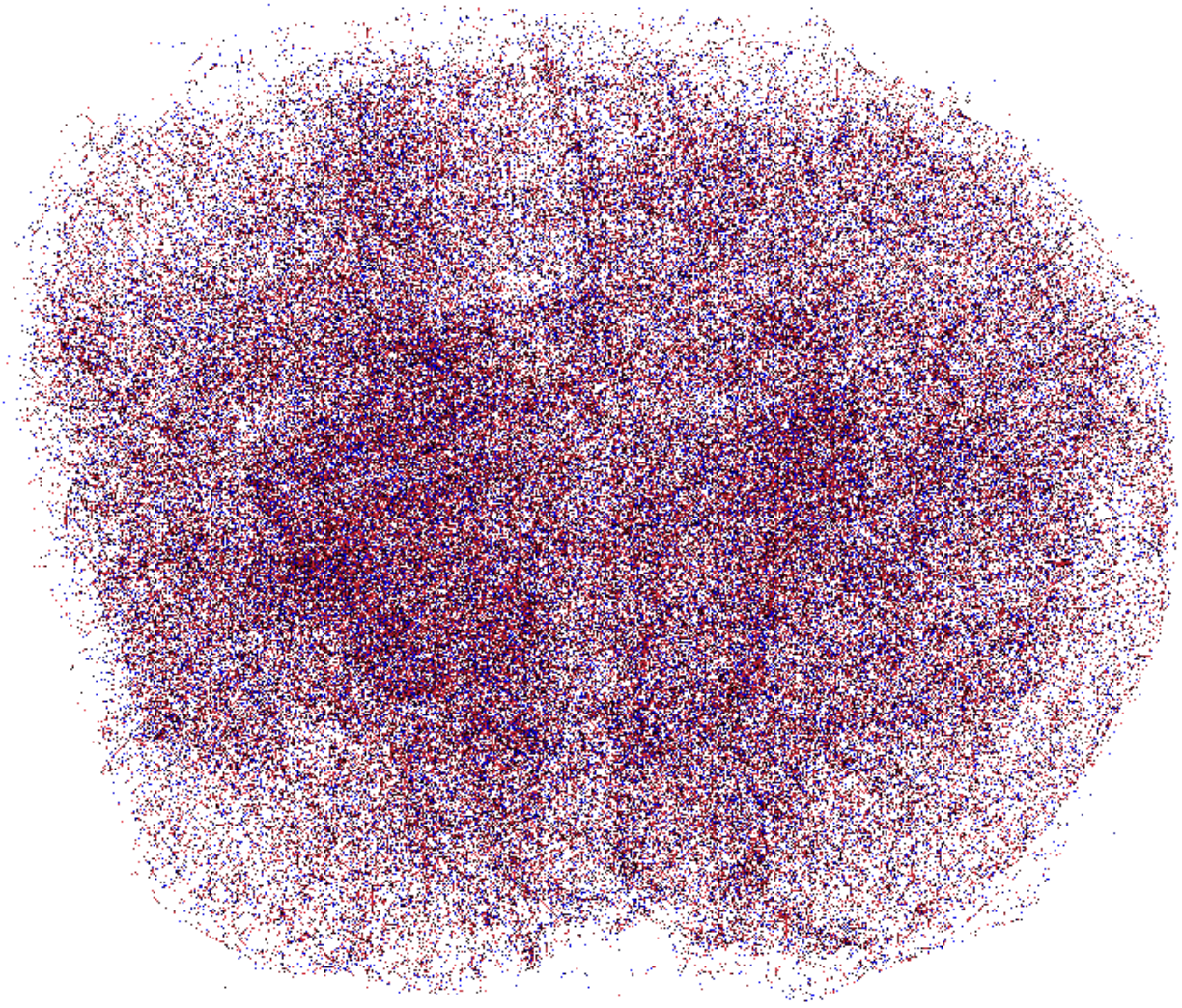


Figure 8.8: Illustration of the reduced graph derived from Figure 8.6. Each edge (red) in this graph represents a vessel segment in the cerebrovasculature system and stores that segment's properties as calculated by our informatics platform. The nodes (blue) in this graph are points of bifurcation between two segments or the terminal end of one segment.

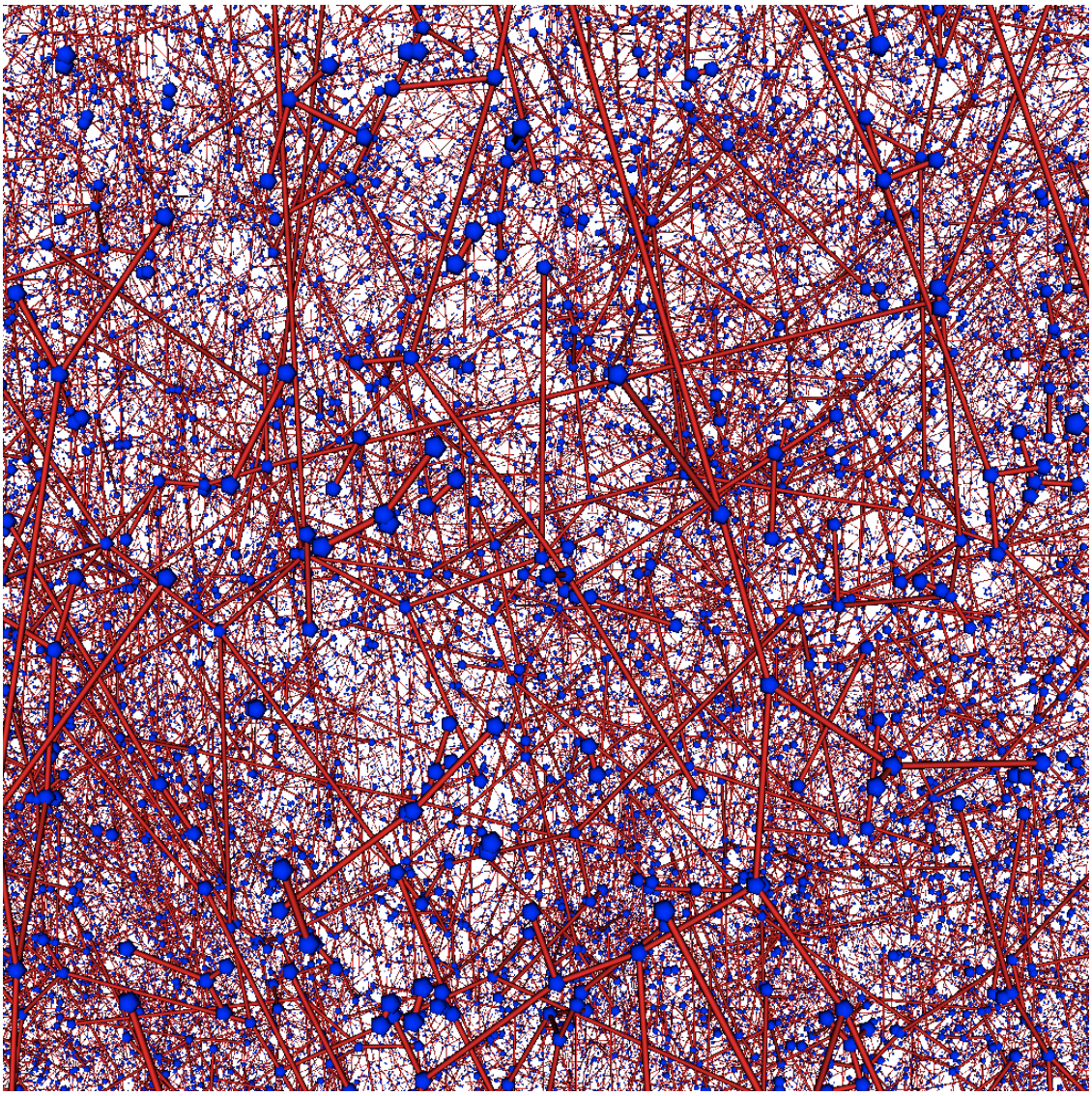


Figure 8.9: Zoomed perspective into an arbitrarily selected region of the whole-brain reduced graph presented in Figure 8.8.

Algorithm 7 Parallel graph reduction and vessel segment attributes calculation algorithm

Input: Graph G
Output: Graph $G\_reduced$ is G reduced

let R be a copy of G
let *G_reduced* be an empty graph
add all nodes from G to *G_reduced*
isolate all segments in R by removing all nodes with degree ≥ 3
let *graphsCCs* be subgraphs for each isolated segment in R
for each segment *gcc* in *graphsCCs* in parallel **do**
    for each node n of degree 1 in *gcc* **do**
        Look up n's edges in G
        Add edges found to gcc, ignoring duplicate edges but adding nodes if necessary to form the connections
    end for
    calculate attributes for each edge in gcc
    let *id* be a unique ID for isolated vessel segment
    let $n1$ and $n2$ be the start and end nodes for the segment (possible that n1 $=$ n2)
    connect $n1$ and $n2$ in $G\_reduced$ with an edge e
    label e with *id*
    store accumulated vessel segment attributes in edge e
    label edges for *gcc* in G with *id*
end for
return G, R

8.3.3.2 Whole-brain graph database

We constructed the whole-brain graph database from the constructed graph and reduced graph representations of the cerebrovasculature dataset. We found a performance issue when importing with the CSV file method described in Section 7.2.2 and resolved this issue with a slight modification of the CSV file being constructed and using `neo4j-admin` to perform the batch import opposed to the cypher query. After making these changes, we executed the command in Figure 8.10 to construct a data model from the nodes and relationships using `neo4j-admin`. The import operation completed in $25\,\mathrm{s}$, which resulted in the construction of approximately 1.88 M nodes and 2.12 M relationships resulting in a database consuming about 564 MB of storage.

It is important to emphasize that once the files have been prepared for import, the data model of the whole-brain is *completely* constructed through the command referenced in Figure 8.10.

```
./neo4j-admin import --nodes "file:///path/to/nodes.csv"\
--relationships "file:///path/to/updatedOriginalEdges.csv" \
--relationships "file:///path/to/reducedEdges.csv"
```

Figure 8.10: `neo4j-admin` command and its arguments used to construct the data model from our graph representations encoding the cerebrovasculature and its attributes across the whole brain.

Relevant optimization is performed once the data model has been constructed, a process similar to indexing in a relational database.

8.3.3.3 Whole-brain study

For our whole-brain study, we classified the blood vessels by their radius to differentiate them into four groups. These groups are shown in Table 8.2 and were also used by [6] in their whole-brain analysis. We determined the distribution of vessels across these groups and present our results in Figure 8.11. In toto, $500{,}654\,\text{segments}$ were isolated, of which $355{,}734\,\text{segments}$ ($\approx 71\%$) were capillaries, $113{,}930\,\text{segments}$ ($\approx 23\%$) were small vessels, $25{,}416\,\text{segments}$ ($\approx 5\%$) were medium vessels, and $5{,}574\,\text{segments}$ ($\approx 1\%$) were large vessels. We illustrate this variation qualitatively across the transverse, coronal, and saggittal in Figures 8.12, 8.13, and 8.14 respectively.

Table 8.2: Criteria used to classify blood vessels segments by their radius in our study.

Group name	**Size of vessels**
Large-sized vessels	$20\,\mu\text{m} <$ Radius
Medium-sized vessels	$10\,\mu\text{m} <$ Radius $\leq 20\,\mu\text{m}$
Small-sized vessels	$5\,\mu\text{m} <$ Radius $\leq 10\,\mu\text{m}$
Capillaries	Radius $\leq 5\,\mu\text{m}$

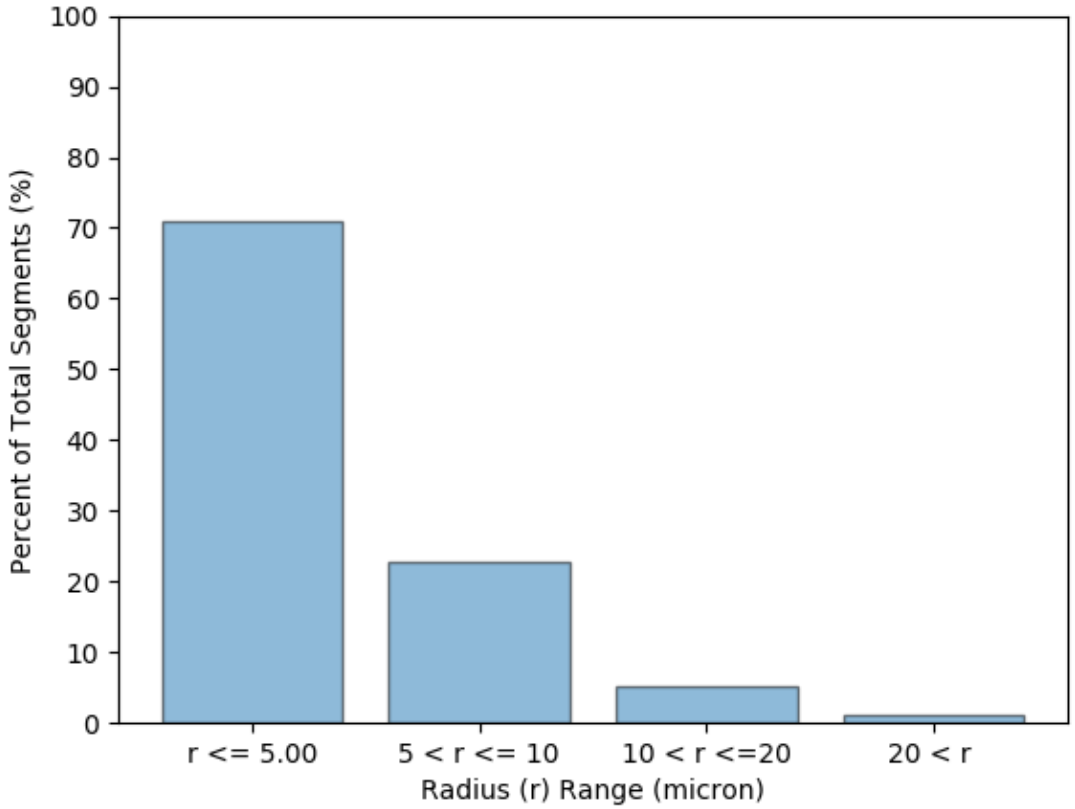


Figure 8.11: Histogram of the percent of total segments classified as each vessel type: capillaries (radius $\leq 5\,\mu\mathrm{m}$), small vessels ($5\,\mu\mathrm{m} <$ radius $\leq 10\,\mu\mathrm{m}$), medium vessels ($10\,\mu\mathrm{m} <$ r $\leq 20\,\mu\mathrm{m}$) and large vessels ($20\,\mu\mathrm{m} <$ radius).

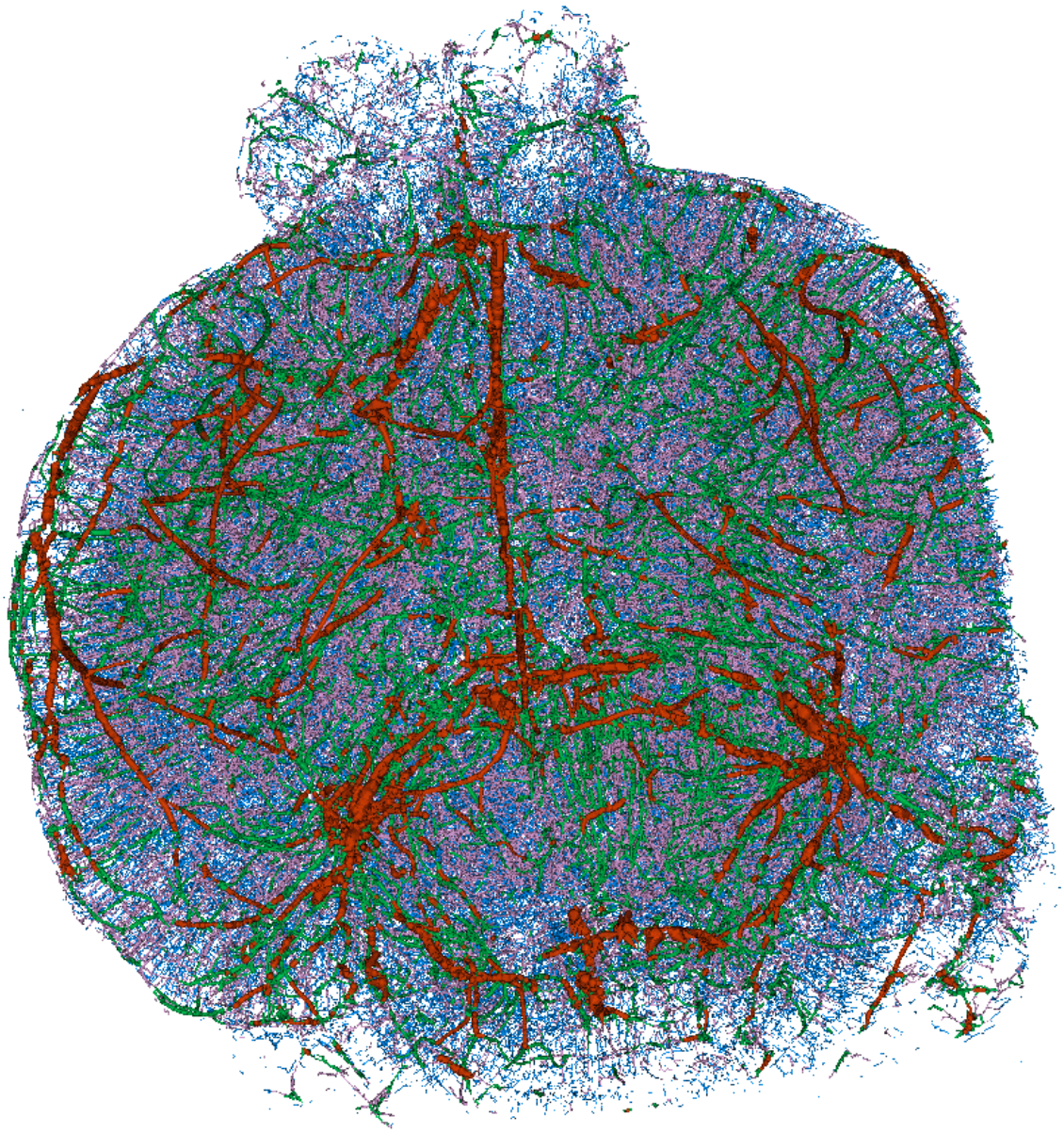


Figure 8.12: Visualization of vessels by category (Table 8.2) across the whole-mouse brain. Capillaries are shown in blue; small-, medium-, and large-sized vessels are shown in purple, green, and red respectively. The perspective of this figure is from the ventral aspect of the brain across its transverse plane (↑, anterior; ↓, posterior).

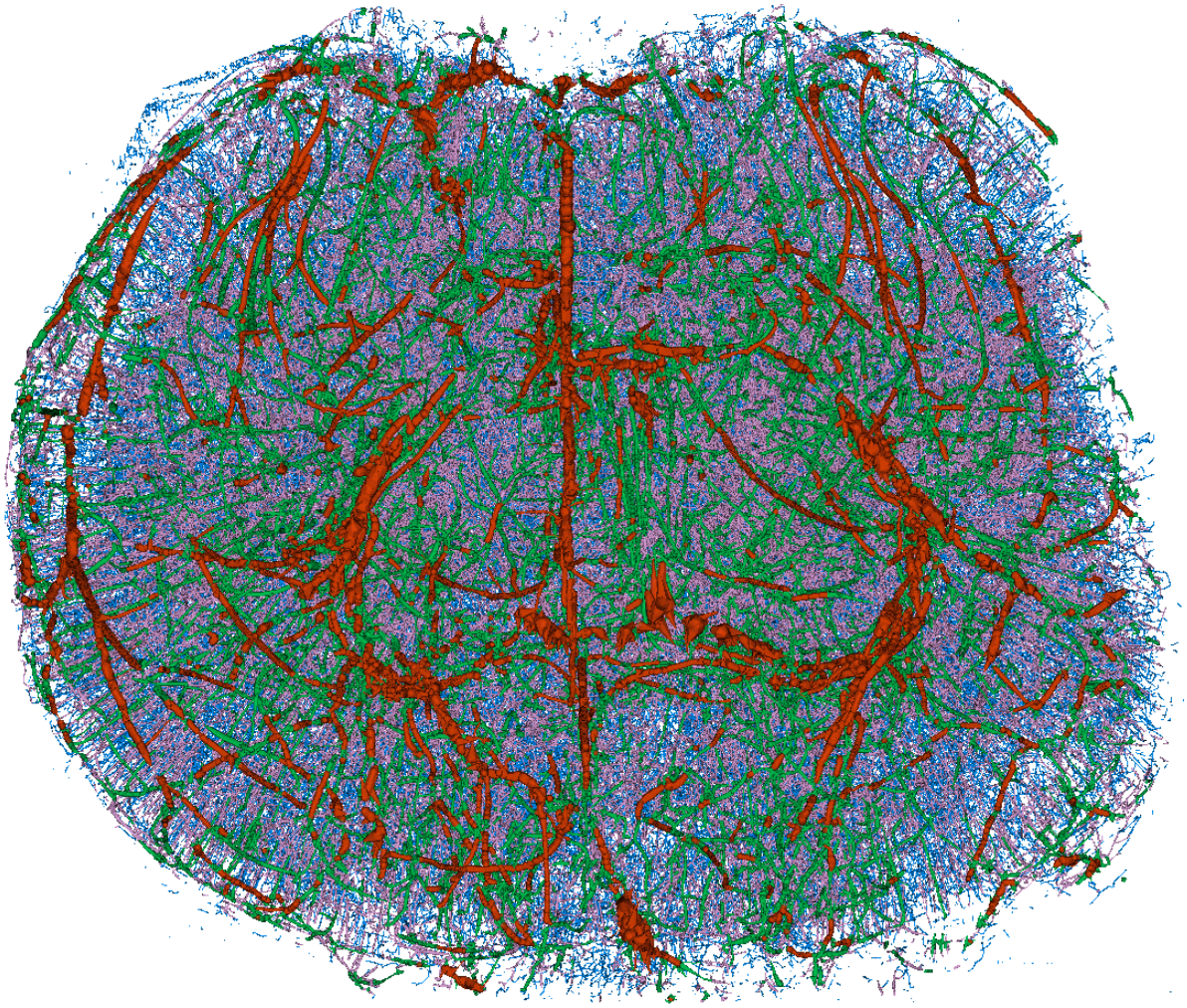


Figure 8.13: Visualization of vessels by category (Table 8.2) across the whole-mouse brain. The perspective of this figure is from the ventral aspect of the brain through its coronal plane (↑, superior; ↓, inferior). Capillaries are shown in blue; small-, medium-, and large-sized vessels are shown in purple, green, and red respectively.

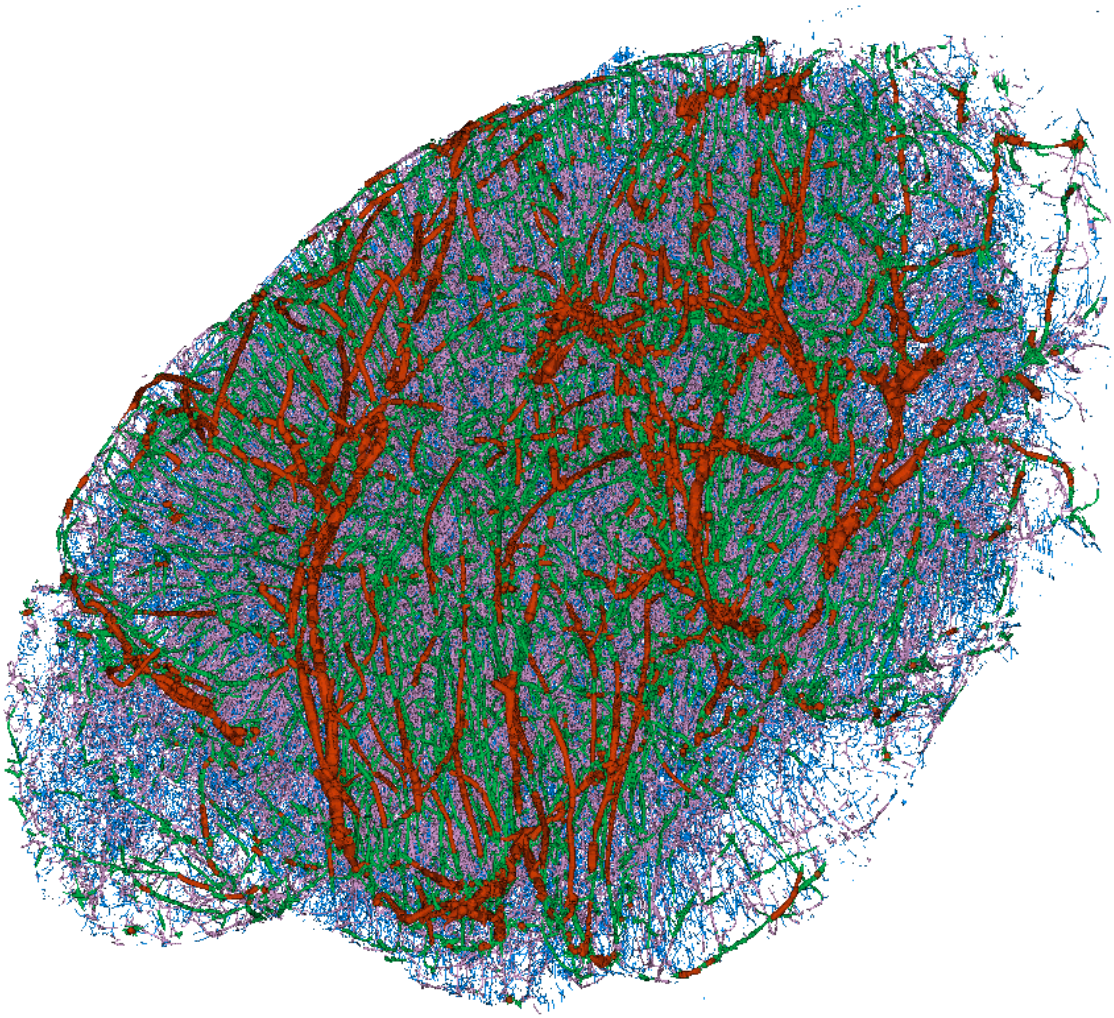


Figure 8.14: Visualization of vessels by category (Table 8.2) across the whole-mouse brain. The perspective of this figure is along the lateral aspect of the left-side of the brain (↙, anterior; ↗, posterior), about through the sagittal plane. Capillaries are shown in blue; small-, medium-, and large-sized vessels are shown in purple, green, and red respectively.

In Chapter 7 we detailed how our database encodes both FILAMENTs and SEGMENTs; the SEGMENTs store the quantitative data describing each respective segment from the cerebrovasculature dataset. Using this data, we report the vascular volume characteristics total volume as $3{,}672{,}591{,}422.51\,\mu m^3$, total surface area as $866{,}686{,}947.68\,\mu m^2$, and total length as $21{,}985{,}119.12$ μm. Assuming that the brain is $1\,cm^3$ in volume, we estimate the fractional volume of the vasculature as 0.367% of the whole-brain. We describe the vessel segment attributes across the whole brain in Table 8.3. We also present the distribution of vessel segment attributes across the entire dataset. The distribution of radii measurements across the dataset is shown in 8.15, with the log number of segments taken by radius. This histogram shows a power function of log number of segments by radius measurement, with smaller radii values appearing with greater frequency than larger values. The distributions for segment length (Figure 8.16), surface area (Figure 8.17), volume (Figure 8.18), distance (Figure 8.19), and surface-area-to-volume ratio follow a similar trend to that of radius (for log number of segments). Meanwhile, the tortuosity of log segments is dominated by those that exhibit only slight twists and turns, with those curved vessels still presenting albeit smaller in number and interspersed in value (Figure 8.20). Finally, the log number of segments grows as a fractional power function of length-to-diameter ratio (Figure 8.21).

Table 8.3: Mean, median, minimum, and maximum values for vessel segment attributes calculated across the whole-mouse brain.

Vessel Attribute	***Mean***	***Median***	***Min***	***Max***
Radius (DIA/2)	$5.92\,\mu m$	$4.80\,\mu m$	$4.80\,\mu m$	$72.00\,\mu m$
Length (LEN)	$43.92\,\mu m$	$29.27\,\mu m$	$0.18\,\mu m$	$1{,}167.92\,\mu m$
Surface Area (SA)	$1{,}731.28\,\mu m^2$	$991.15\,\mu m^2$	$7.54\,\mu m^2$	$165{,}663.09\,\mu m^2$
Volume (V)	$7{,}336.31\,\mu m^3$	$2{,}545.96\,\mu m^3$	$18.10\,\mu m^3$	$2{,}430{,}766.41\,\mu m^3$
Distance	$40.45\,\mu m$	$27.89\,\mu m$	$0.18\,\mu m$	$1{,}075.49\,\mu m$
Tortuosity	1.06	1.01	1.00	171.10
LEN:DIA	3.94	0.01	2.59	73.59
SA:V	0.38	0.42	0.03	0.42

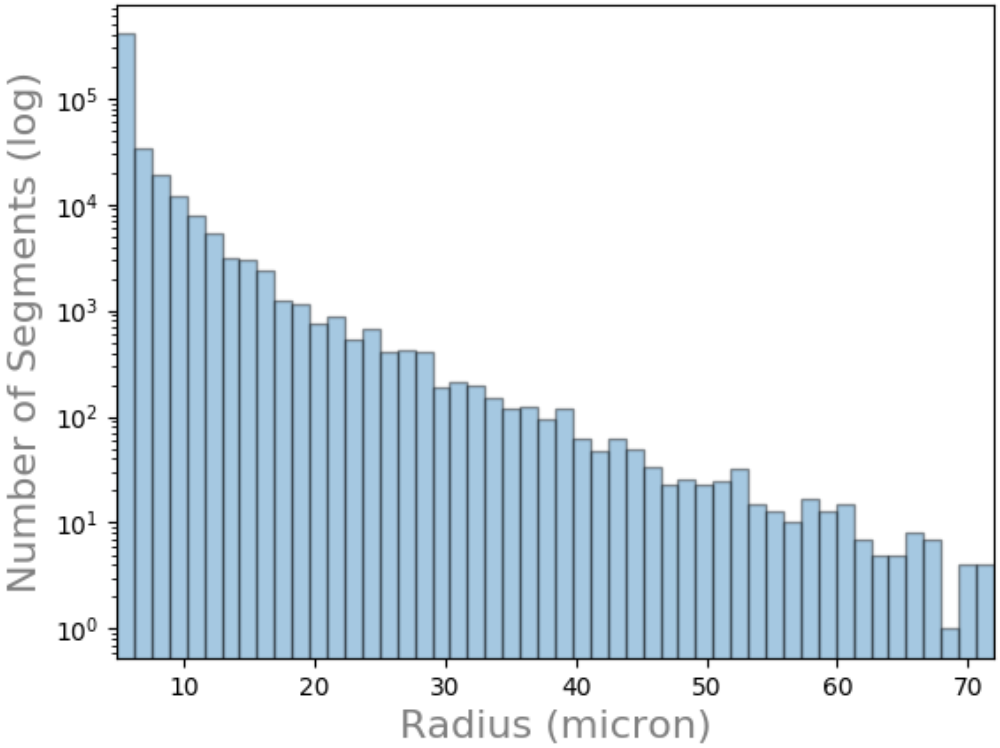


Figure 8.15: Whole brain analysis: number of vessel segments (log) by radius.

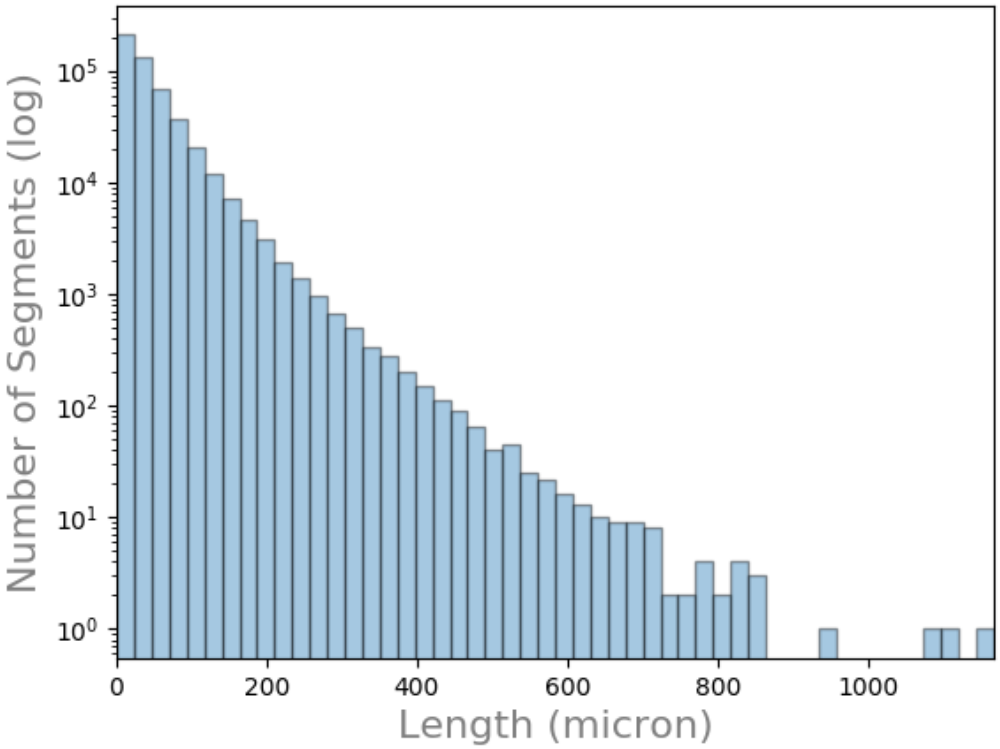


Figure 8.16: Whole brain analysis: number of vessel segments (log) by length.

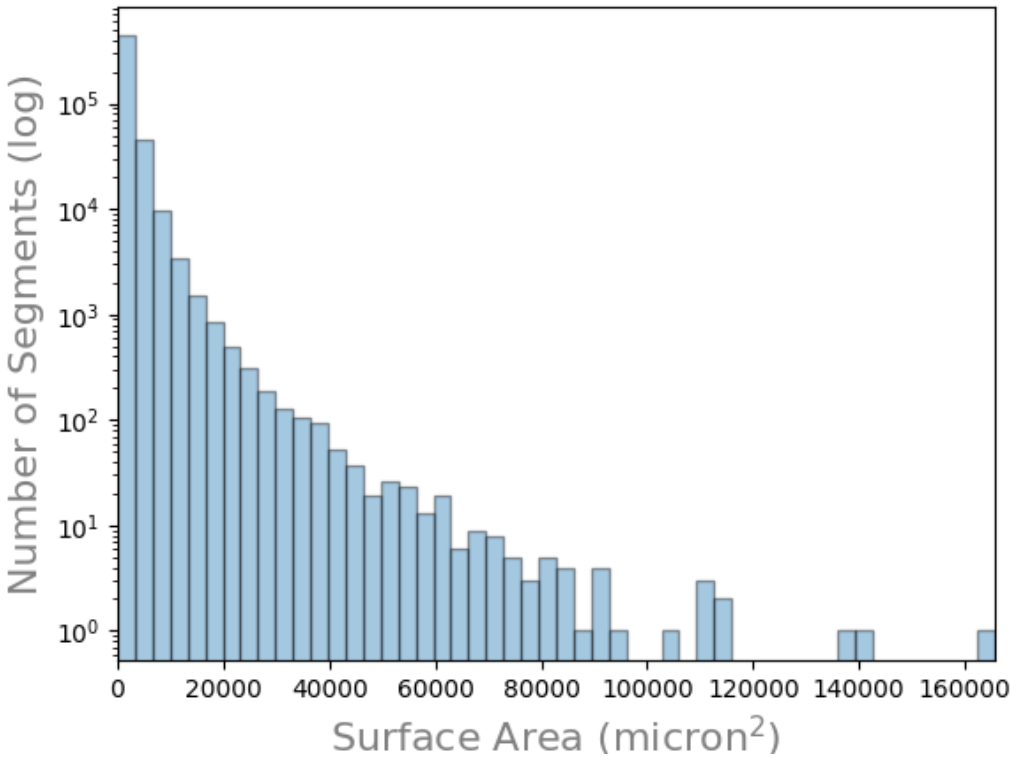


Figure 8.17: Whole brain analysis: number of vessel segments (log) by surface area.

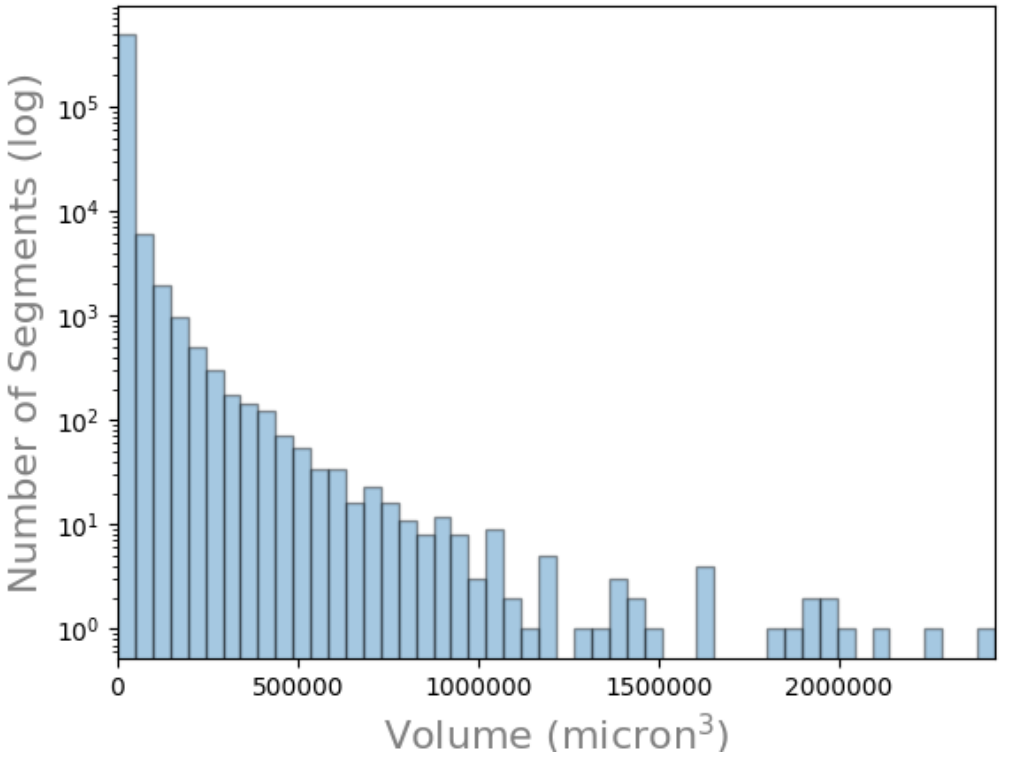


Figure 8.18: Whole brain analysis: number of vessel segments (log) by volume.

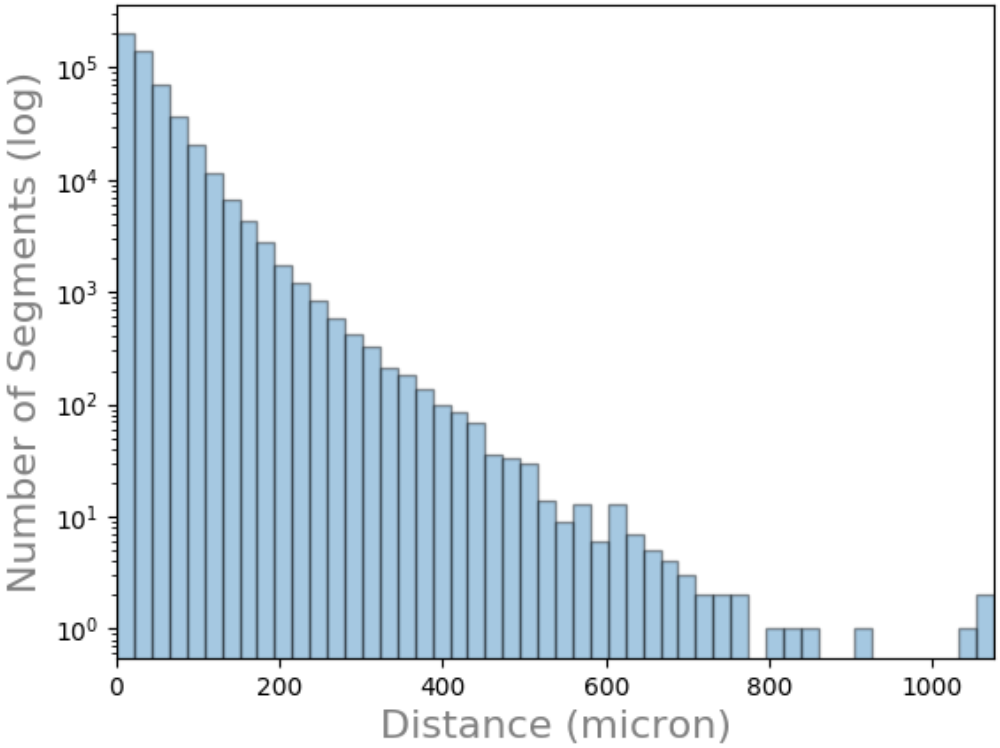


Figure 8.19: Whole brain analysis: number of vessel segments (log) by distance.

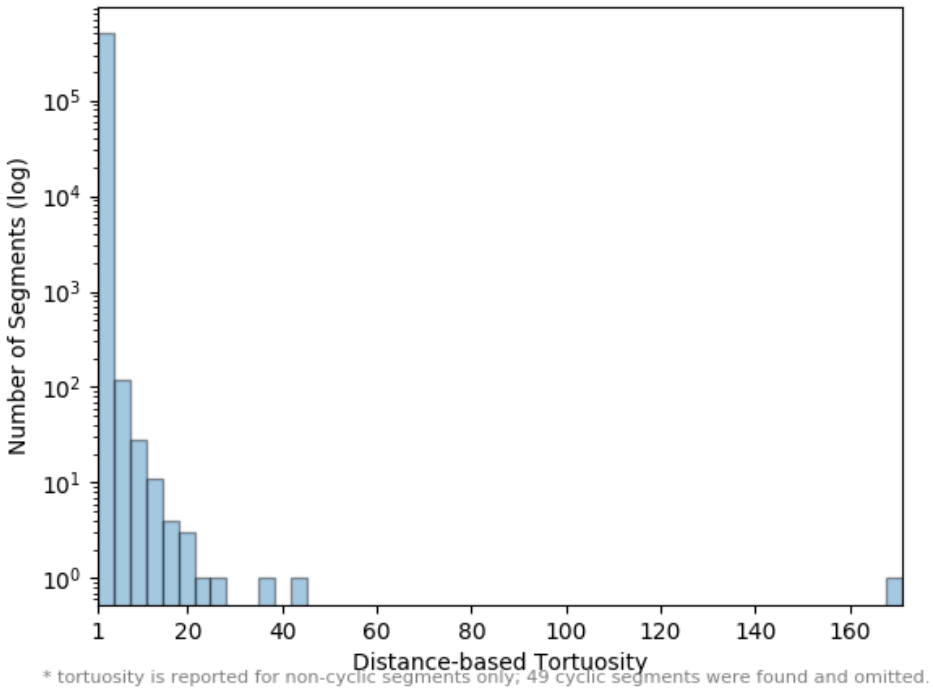


Figure 8.20: Whole brain analysis: number of vessel segments (log) by tortuosity.

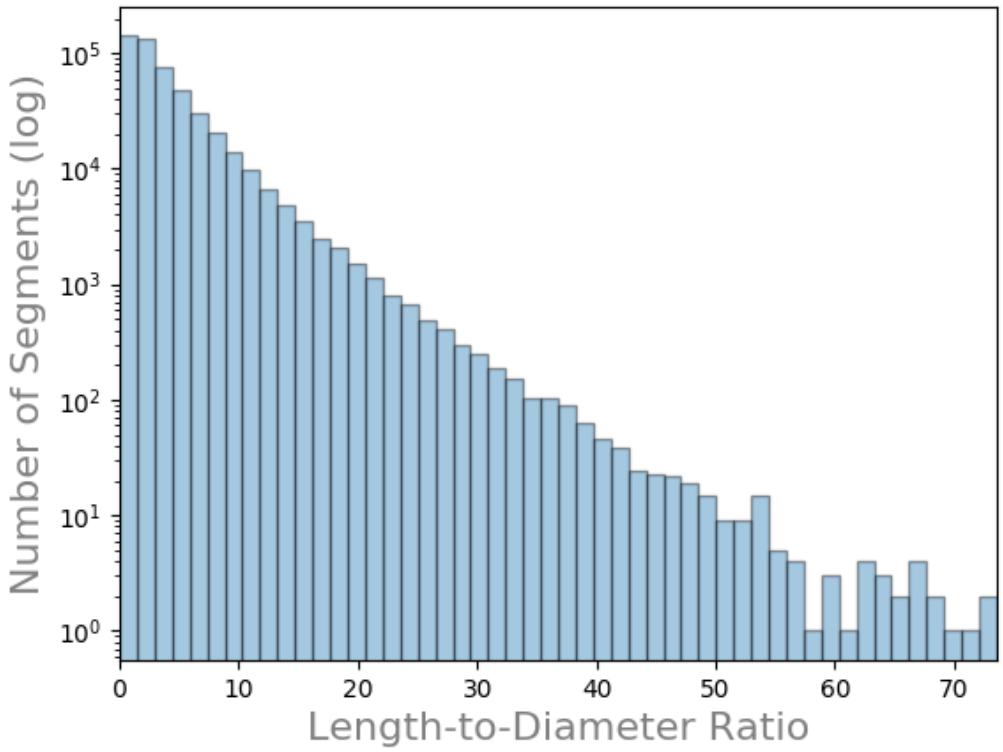


Figure 8.21: Whole brain analysis: number of vessel segments (log) by length-to-diameter ratio.

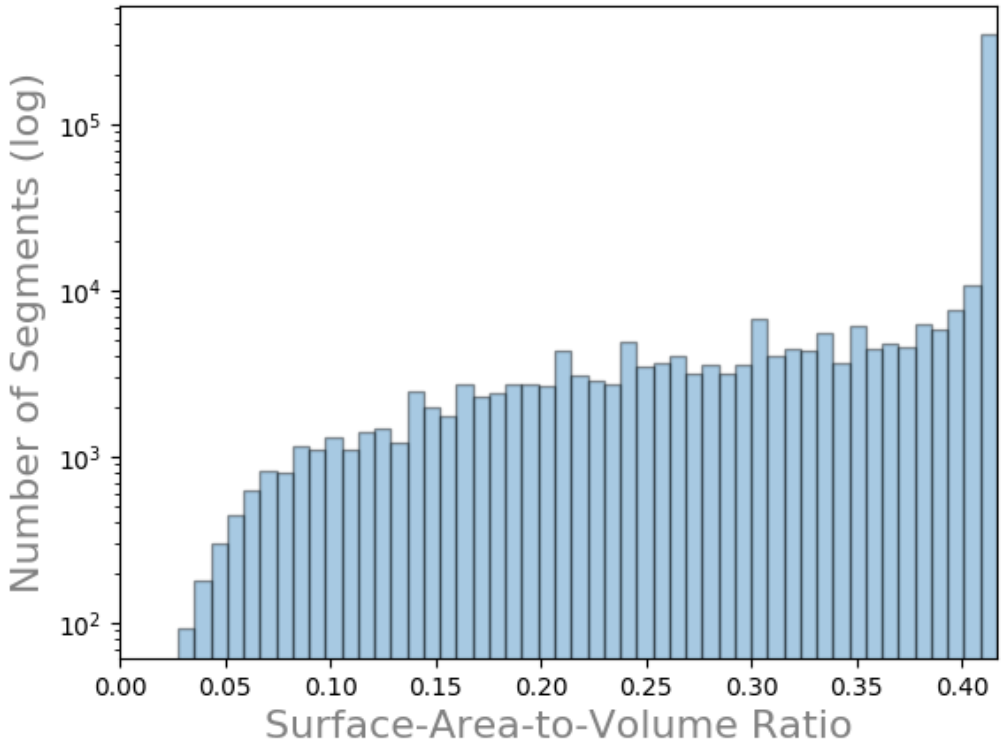


Figure 8.22: Whole brain analysis: number of vessel segments (log) by surface-area-to-volume ratio.

8.3.3.3.1 Analysis of the Large-sized Vessels ($20\,\mu\mathrm{m} <$ *radius)*

Figure 8.23: Large-sized vessels ($20\,\mu\mathrm{m} <$ radius) presenting in the whole-brain dataset (coronal view: ↑, superior; ↓, inferior).

We isolated the large vascular segments by querying our database for segments with an average radius greater than 20 (Figure 8.23). This query is presented in 8.24. The point of showing the query here is to reinforce the notion that all that is required to isolated segments for analysis is the correct database query. Using this result we report the vascular characteristics for all large segments: total volume as $756{,}958{,}128.53\,\mu\mathrm{m}^3$, total surface area as $52{,}782{,}719.59\,\mu\mathrm{m}^2$, and total length as $328{,}727.76\,\mu\mathrm{m}$. We estimate the fractional volume of the vasculature as 0.076% of a $1\,\mathrm{cm}^3$ brain. We describe the attributes of these segments in Table 8.4. We detail the distributions for radius in Figure 8.25, length in Figure 8.26, surface area in Figure 8.27, volume in Figure 8.28,

and distance in Figure 8.29. We also present the distributions for tortuosity, length-to-diameter ratio, and surface-area-to-volume ratio in Figures 8.30, 8.31, and 8.32 respectively.

```
MATCH path = (start:NODE)-[s:SEGMENT]-(end:NODE)
WHERE  s.avg_radius > 20
RETURN path, nodes(path) as nodes, rels(path) as rels
```

Figure 8.24: Cypher query used retrieve the relationship `SEGMENT`s with an average radius greater than 20.

Table 8.4: Mean, median, minimum, and maximum values for vessel segment attributes calculated across the large-sized vessel segments.

Vessel Attribute	***Mean***	***Median***	***Min***	***Max***
Radius (DIA/2)	28.21 µm	25.35 µm	20.00 µm	72.00 µm
Length (LEN)	58.98 µm	40.51 µm	1.17 µm	846.74 µm
Surface Area (SA)	9,469.45 μm^2	6,617.24 μm^2	220.29 μm^2	138,101.36 μm^2
Volume (V)	135,801.60 μm^3	86,564.09 μm^3	3,022.02 μm^3	2,430,766.41 μm^3
Distance	56.25 µm	39.70 µm	1.17 µm	839.69 µm
Tortuosity	1.03	1.00	1.00	5.97
LEN:DIA	1.15	0.76	0.02	18.44
SA:V	0.08	0.08	0.03	0.17

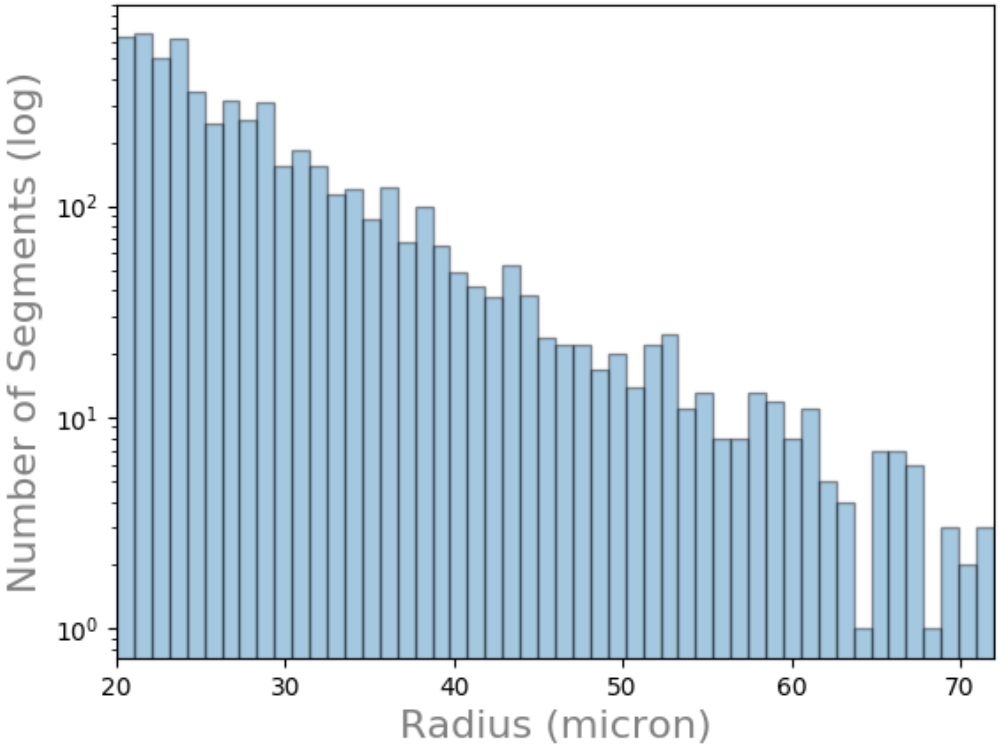


Figure 8.25: Large-sized vessels: number of vessel segments (log) by radius.

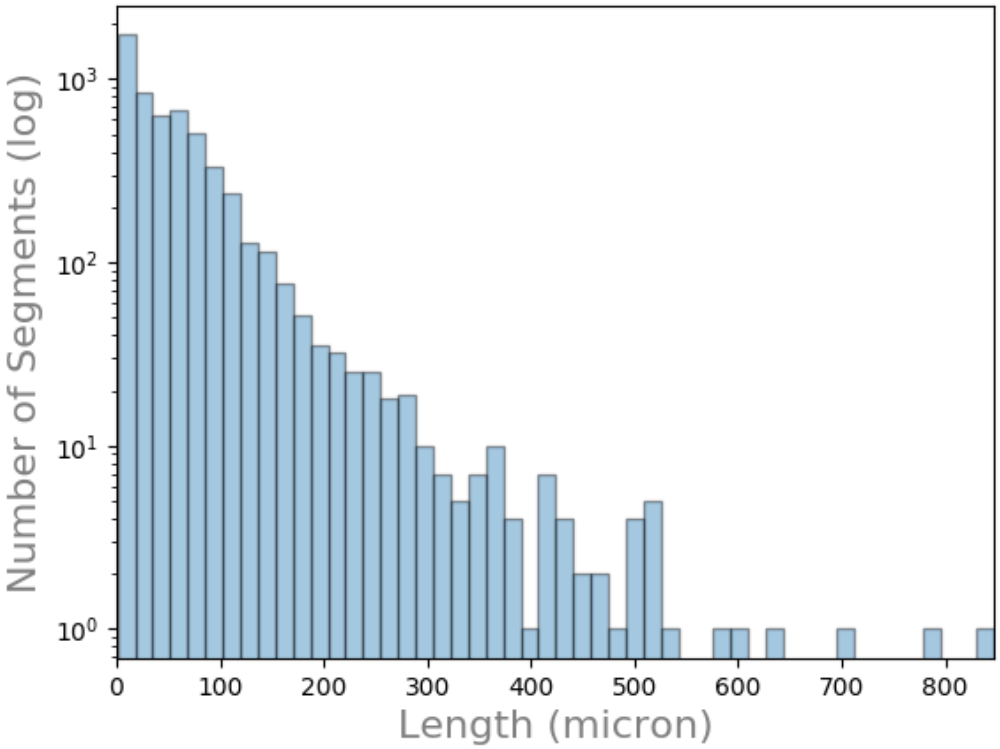


Figure 8.26: Large-sized vessels: number of vessel segments (log) by length.

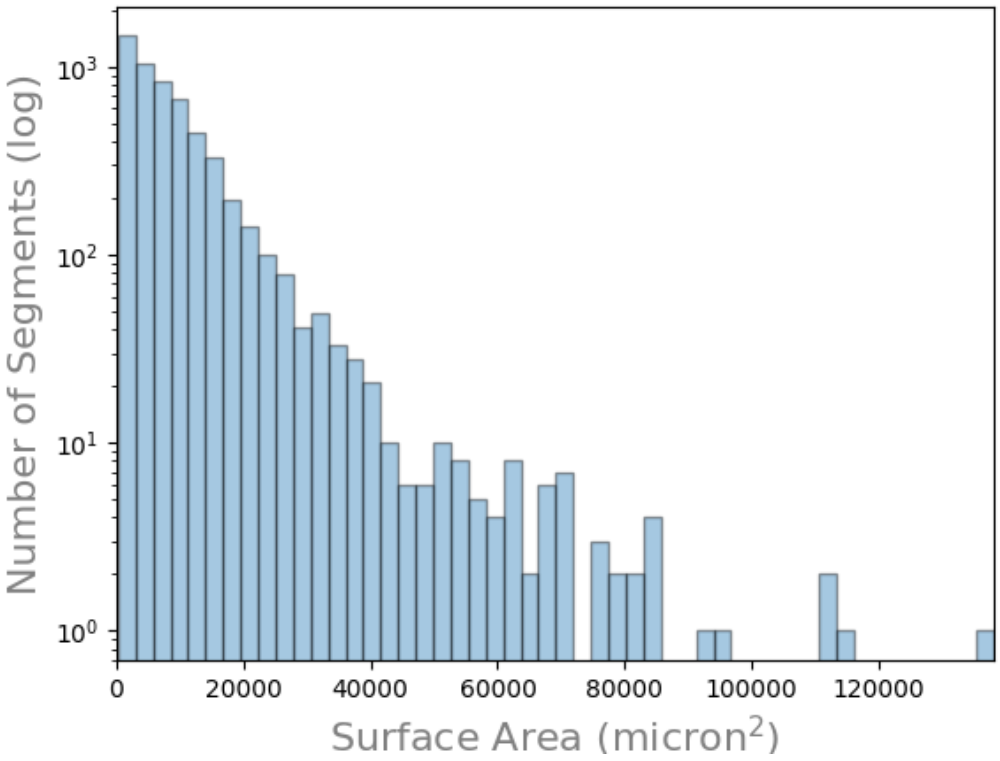


Figure 8.27: Large-sized vessels: number of vessel segments (log) by surface area.

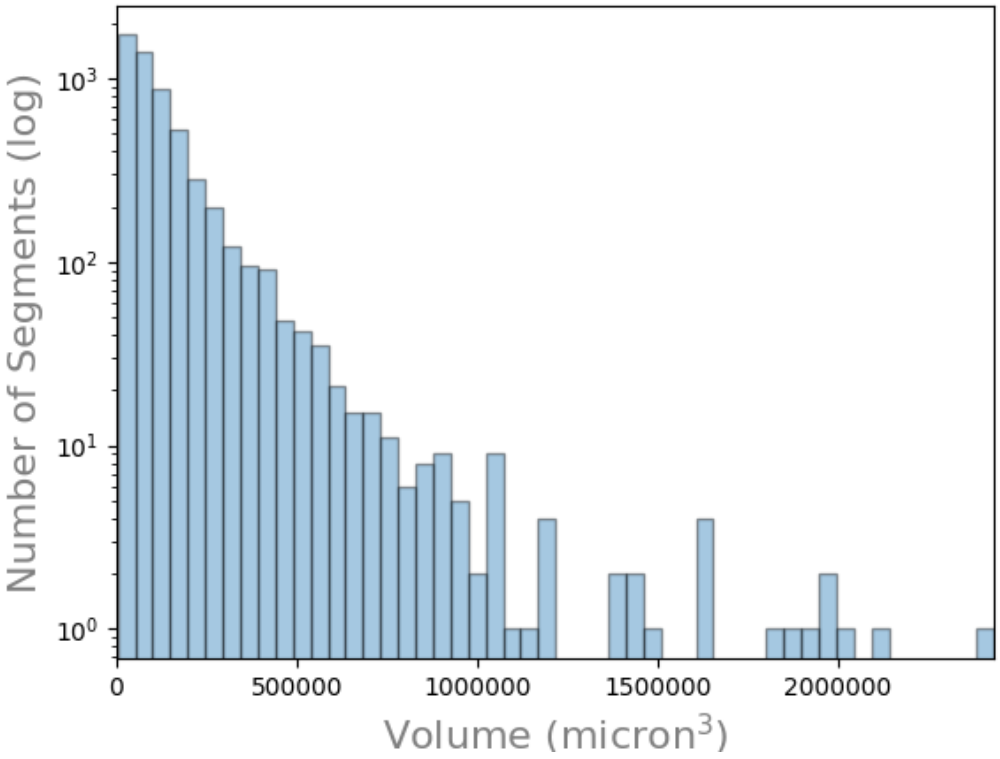


Figure 8.28: Large-sized vessels: number of vessel segments (log) by volume.

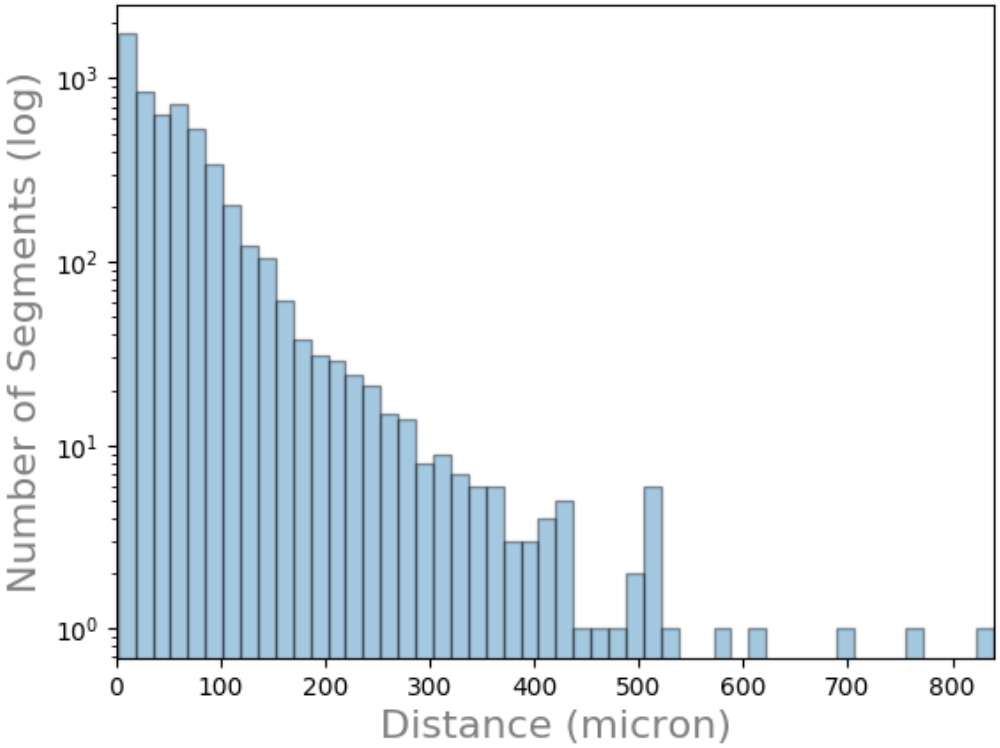


Figure 8.29: Large-sized vessels: number of vessel segments (log) by distance.

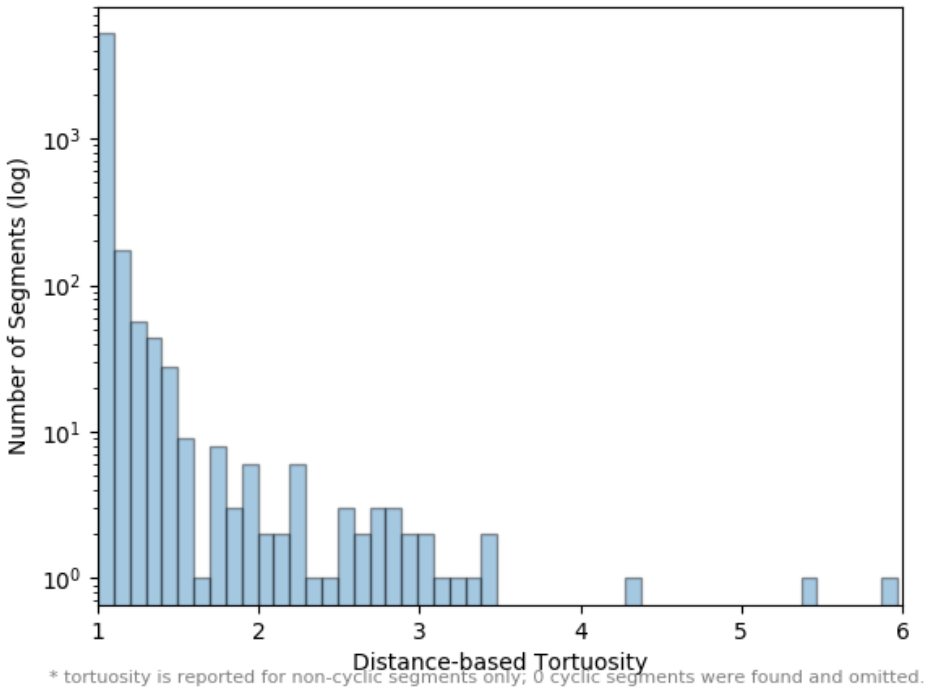


Figure 8.30: Large-sized vessels: number of vessel segments (log) by tortuosity.

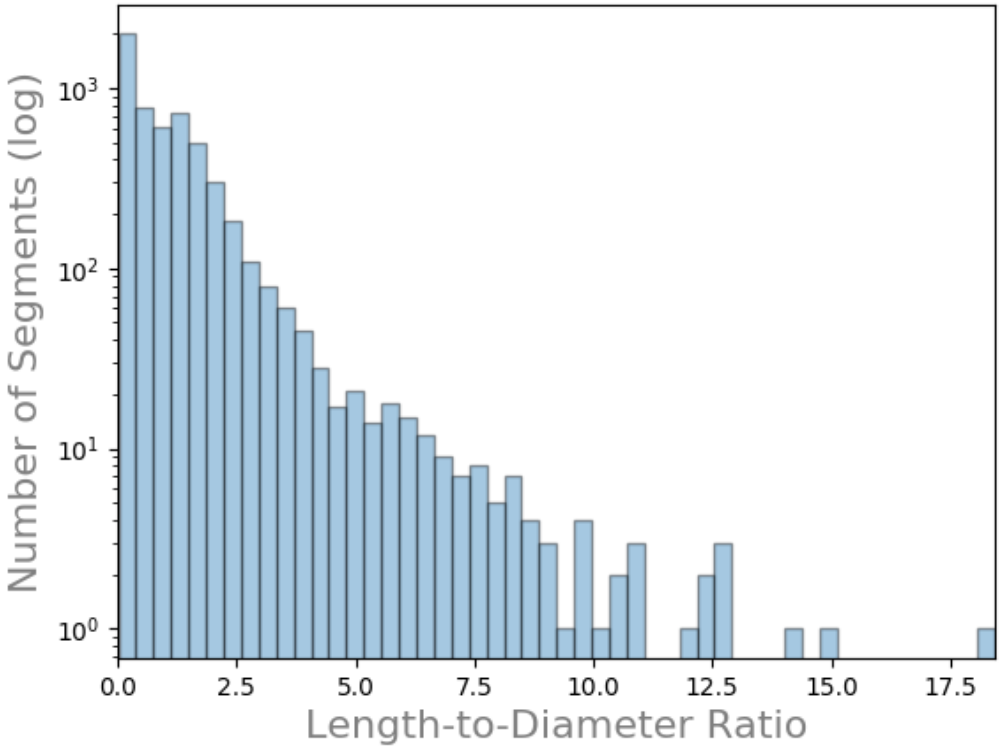


Figure 8.31: Large-sized vessels: number of vessel segments (log) by length-to-diameter ratio.

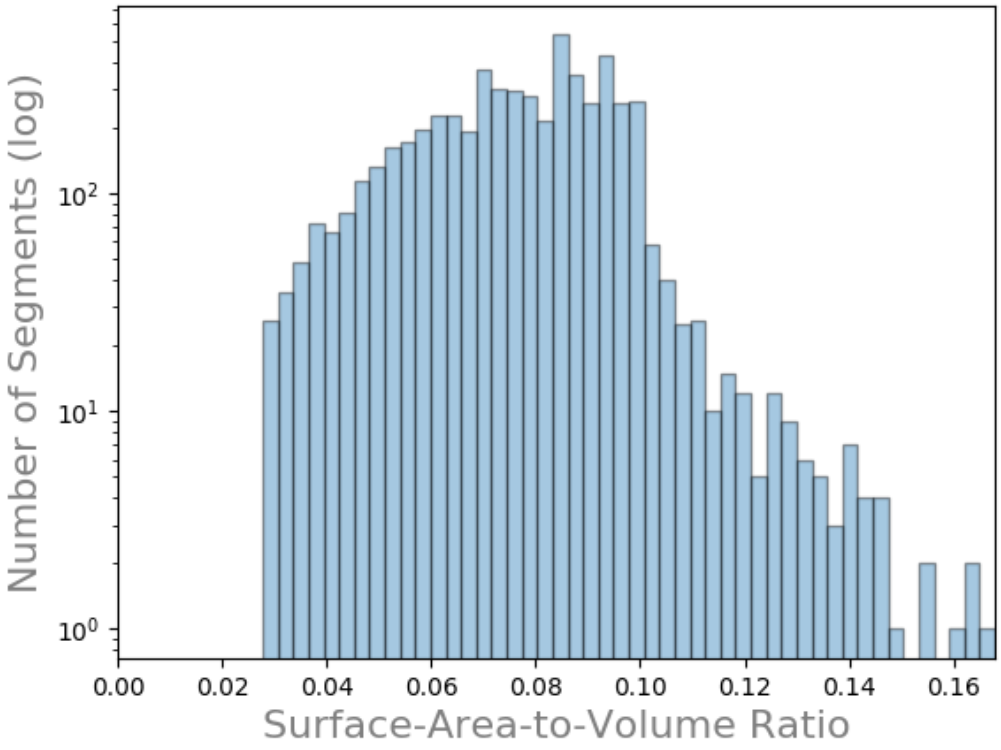


Figure 8.32: Large-sized vessels: number of vessel segments (log) by surface-area-to-volume ratio.

8.3.3.3.2 Analysis of the Medium-sized Vessels ($10\,\mu\mathrm{m} < \textit{radius} \leq 20\,\mu\mathrm{m}$*)*

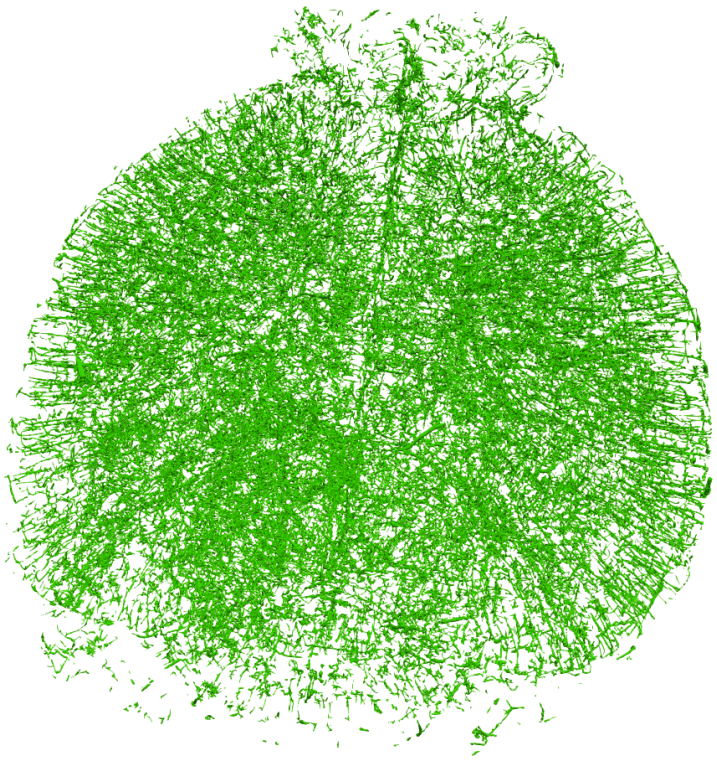


Figure 8.33: Medium-sized vessels presenting in the whole-brain dataset (transverse view: ↑, anterior; ↓, posterior).

We isolated the medium vascular segments by querying our database for segments with an average radius $\mathtt{r}$ such that $10\,\mu\mathrm{m} < \mathtt{r} \leq 20\,\mu\mathrm{m}$ (Figure 8.33). This query is presented in 8.34; it is all that is required to isolated the relevant segments for analysis. Using this data, we report the vascular characteristics for all medium-sized vessels as follows: total volume as $932{,}969{,}299.53$ $\mu\mathrm{m}^3$, total surface area as $131{,}645{,}344.71\,\mu\mathrm{m}^2$, and total length as $1{,}651{,}778.83\,\mu\mathrm{m}$. We estimate the fractional volume of the vasculature as 0.093% of the whole-brain. We describe the attributes of these segments in Table 8.5. We present the distributions for radius in Figure 8.35, length in Figure 8.36, surface area in Figure 8.37, volume in Figure 8.38, and distance in Figure 8.39. Furthermore,

we detail the distributions for tortuosity, length-to-diameter ratio, and surface-area-to-volume ratio in Figures 8.40, 8.41, and 8.42 respectively.

```
MATCH path = (start:NODE)-[s:SEGMENT]-(end:NODE)
WHERE  s.avg_radius <= 20 and s.avg_radius > 10
RETURN path, nodes(path) as nodes, rels(path) as rels
```

Figure 8.34: Cypher query used retrieve the relationship `SEGMENT`s with an average radius greater than 20.

Table 8.5: Mean, median, minimum, and maximum values for vessel segment attributes calculated across the medium-sized vessel segments.

Vessel Attribute	***Mean***	***Median***	***Min***	***Max***
Radius (DIA/2)	13.11 µm	12.20 µm	10.00 µm	20.00 µm
Length (LEN)	64.99 µm	45.89 µm	0.98 µm	1,167.92 µm
Surface Area (SA)	5,179.62 μm^2	3,542.14 μm^2	88.29 μm^2	165,663.09 μm^2
Volume (V)	36,707.95 μm^3	21,626.68 μm^3	567.40 μm^3	2,269,249.39 μm^3
Distance	61.10 µm	44.21 µm	0.66 µm	1,066.74 µm
Tortuosity	1.05	1.00	1.00	10.96
LEN:DIA	2.54	1.78	0.03	41.41
SA:V	0.16	0.16	0.05	0.36

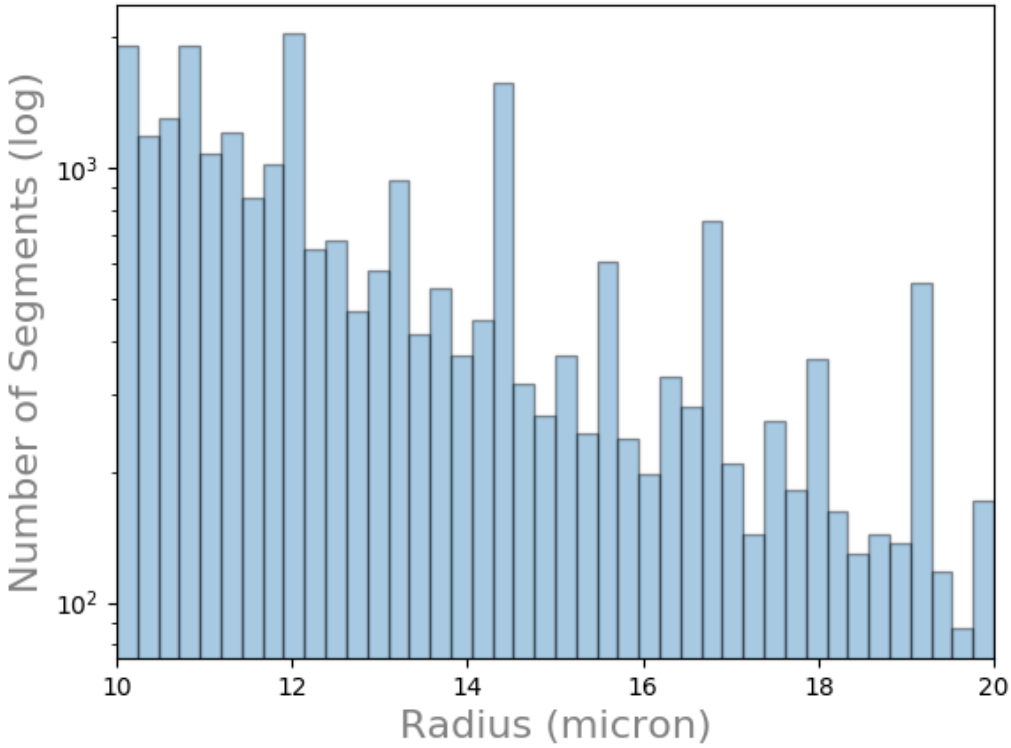


Figure 8.35: Medium-sized vessels: number of vessel segments (log) by radius.

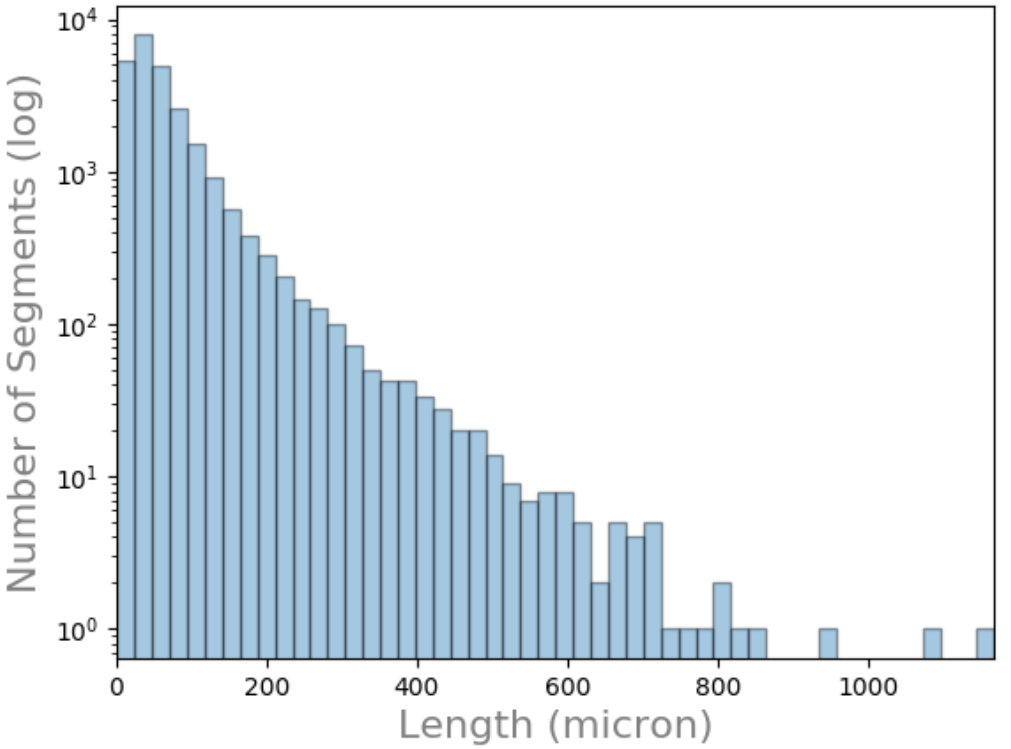


Figure 8.36: Medium-sized vessels: number of vessel segments (log) by length.

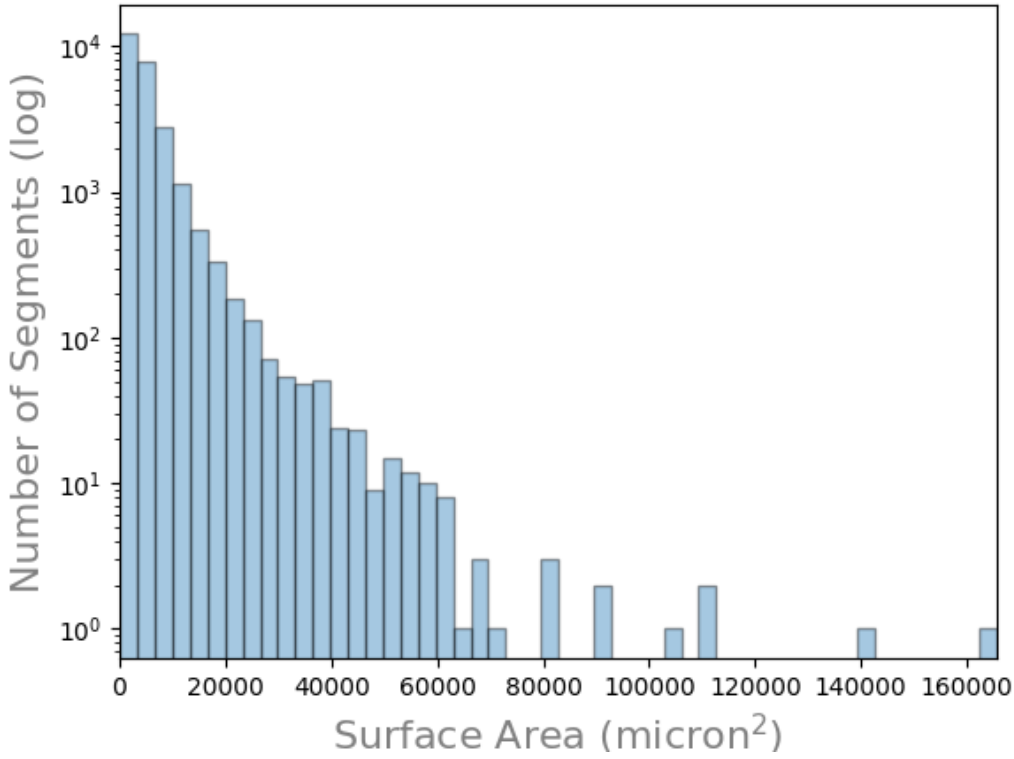


Figure 8.37: Medium-sized vessels: number of vessel segments (log) by surface area.

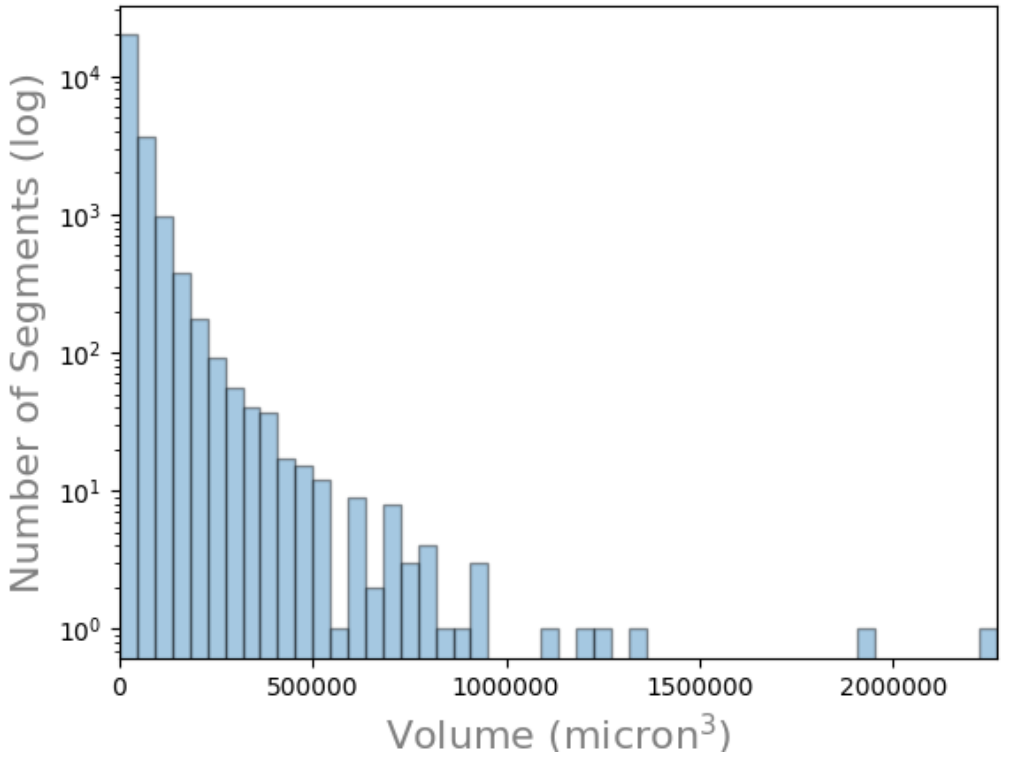


Figure 8.38: Medium-sized vessels: number of vessel segments (log) by volume.

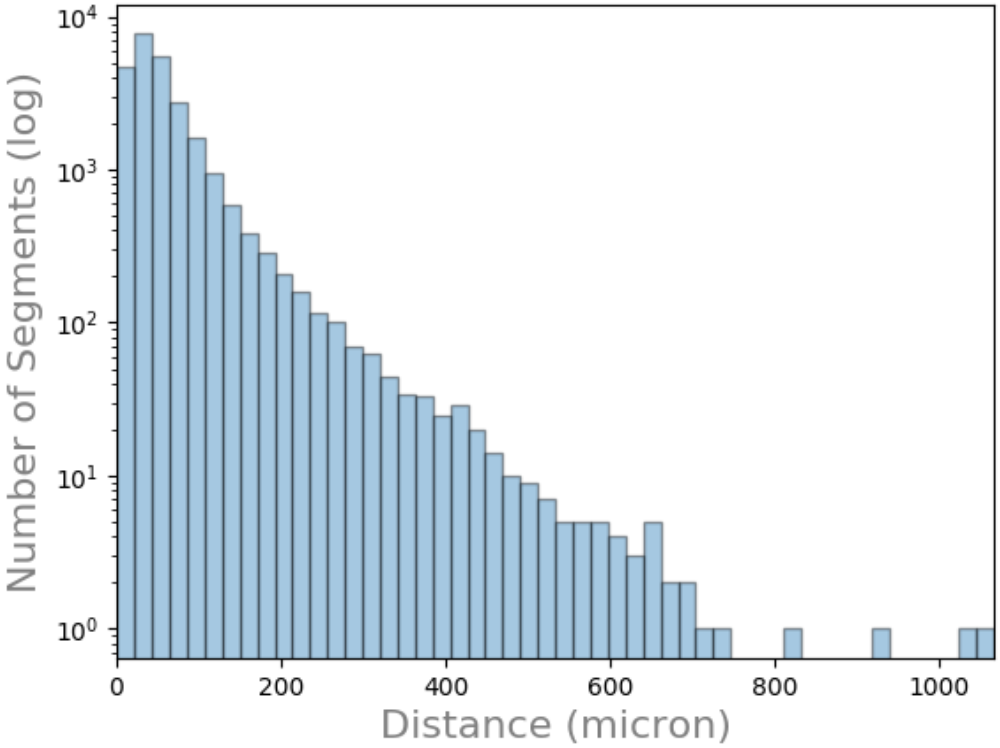


Figure 8.39: Medium-sized vessels: number of vessel segments (log) by distance.

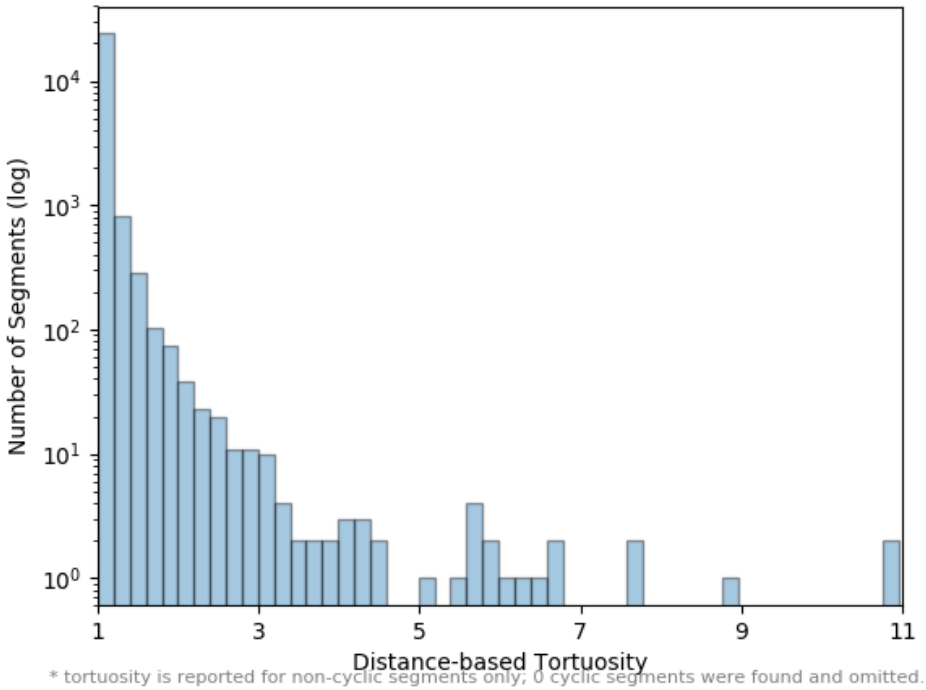


Figure 8.40: Medium-sized vessels: number of vessel segments (log) by tortuosity.

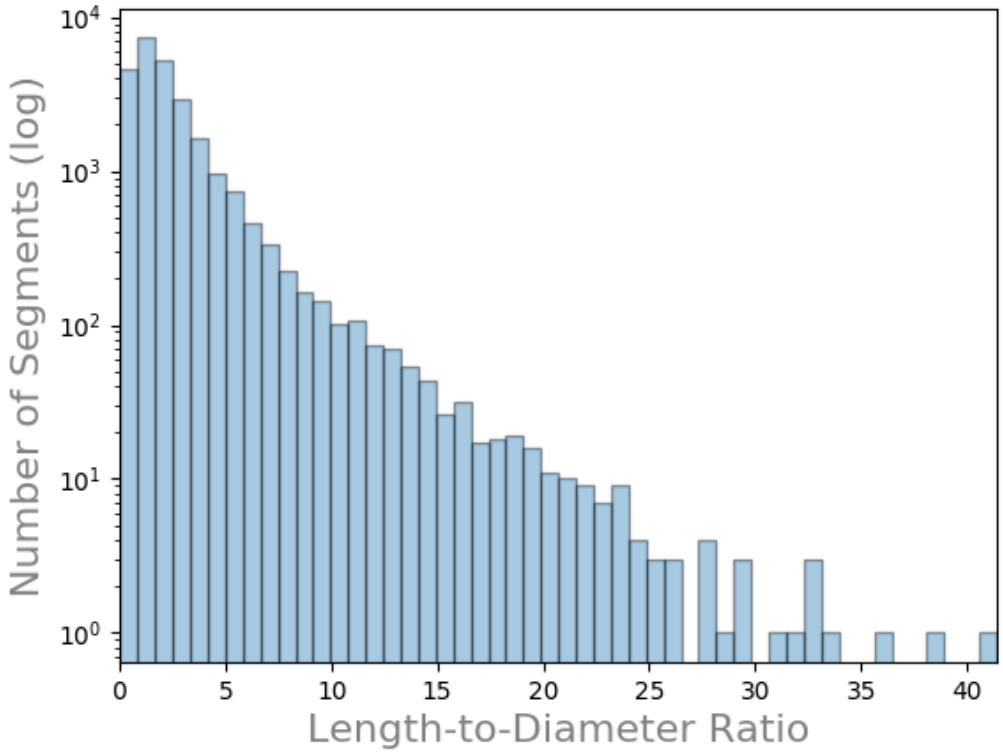


Figure 8.41: Medium-sized vessels: number of vessel segments (log) by length-to-diameter ratio.

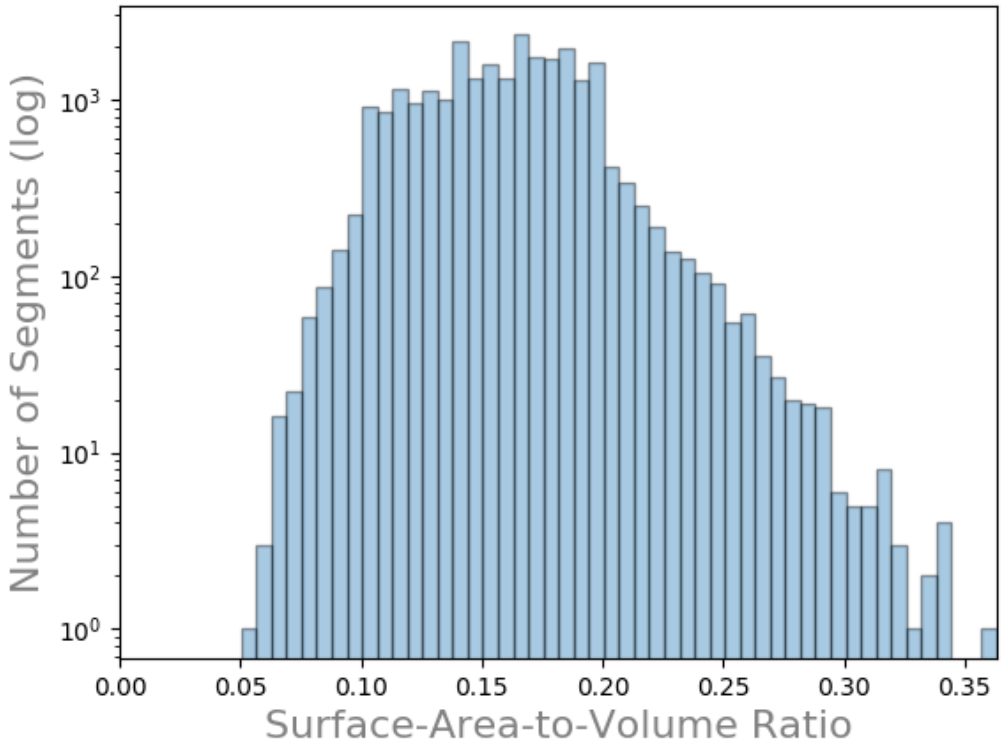


Figure 8.42: Medium-sized vessels: number of vessel segments (log) by surface-area-to-volume ratio.

8.3.3.3.3 *Analysis of the Small-sized Vessels (*$5\,\mu\mathrm{m} < $ *radius* $\leq 10\,\mu\mathrm{m}$*)*

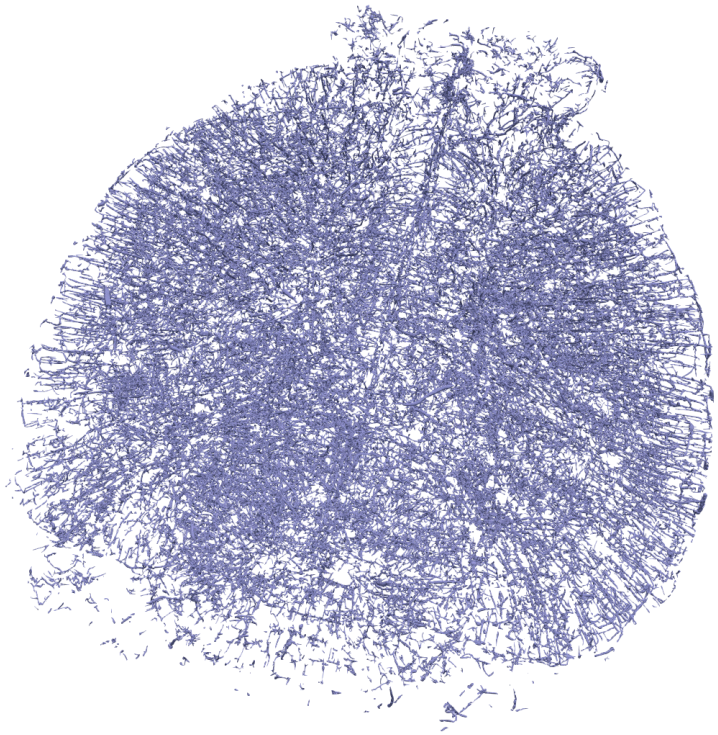


Figure 8.43: Small-sized vessels presenting in the whole-brain dataset (transverse view: ↑, anterior; ↓, posterior).

We isolated the small vascular segments by querying our database for segments with an average radius $\mathtt{r}$ such that $5\,\mu\mathrm{m} < \mathtt{r} \leq 10\,\mu\mathrm{m}$ (Figure 8.43). Using this data, we report the vascular characteristics for all small-sized vessels as follows: total volume as $1{,}007{,}300{,}014.21\,\mu\mathrm{m}^3$, total surface area as $280{,}200{,}401.66\,\mu\mathrm{m}^2$, and total length as $6{,}768{,}182.89\,\mu\mathrm{m}$. We estimate the fractional volume of the vasculature as 0.10% of the whole-brain. We describe the attributes of these segments in Table 8.6. We detail the distributions for radius in Figure 8.44, length in Figure 8.45, surface area in Figure 8.46, volume in Figure 8.47, and distance in Figure 8.48. We also present the distributions for tortuosity, length-to-diameter ratio, and surface-area-to-volume ratio in Figures 8.49, 8.50, and 8.51 respectively.

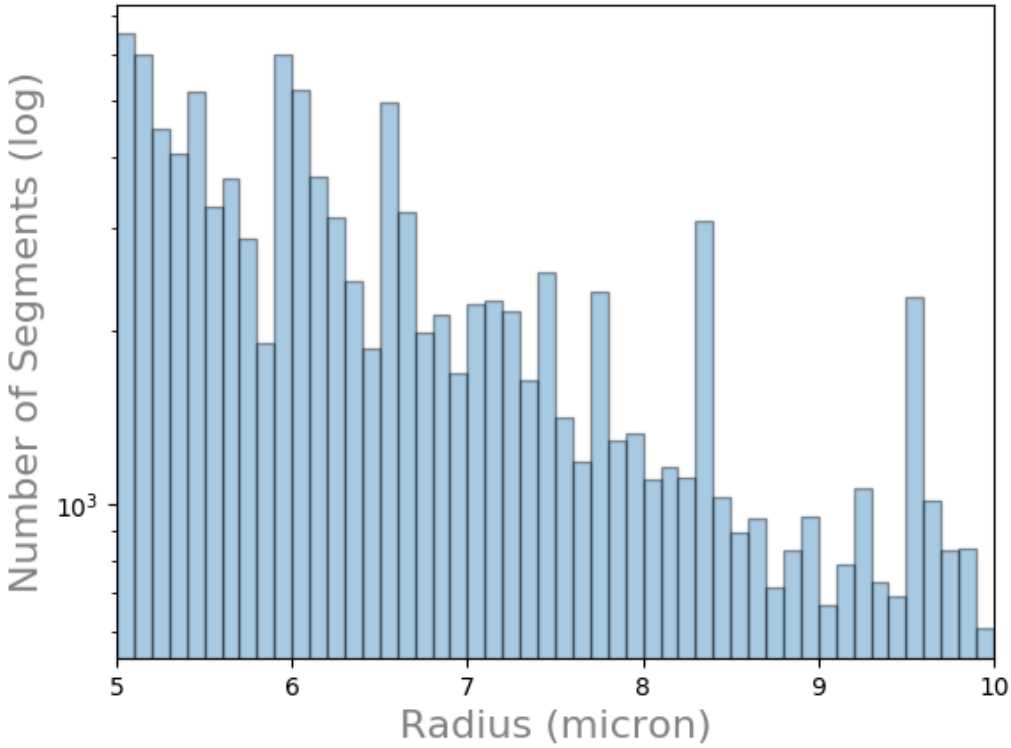


Figure 8.44: Small-sized vessels: number of vessel segments (log) by radius.

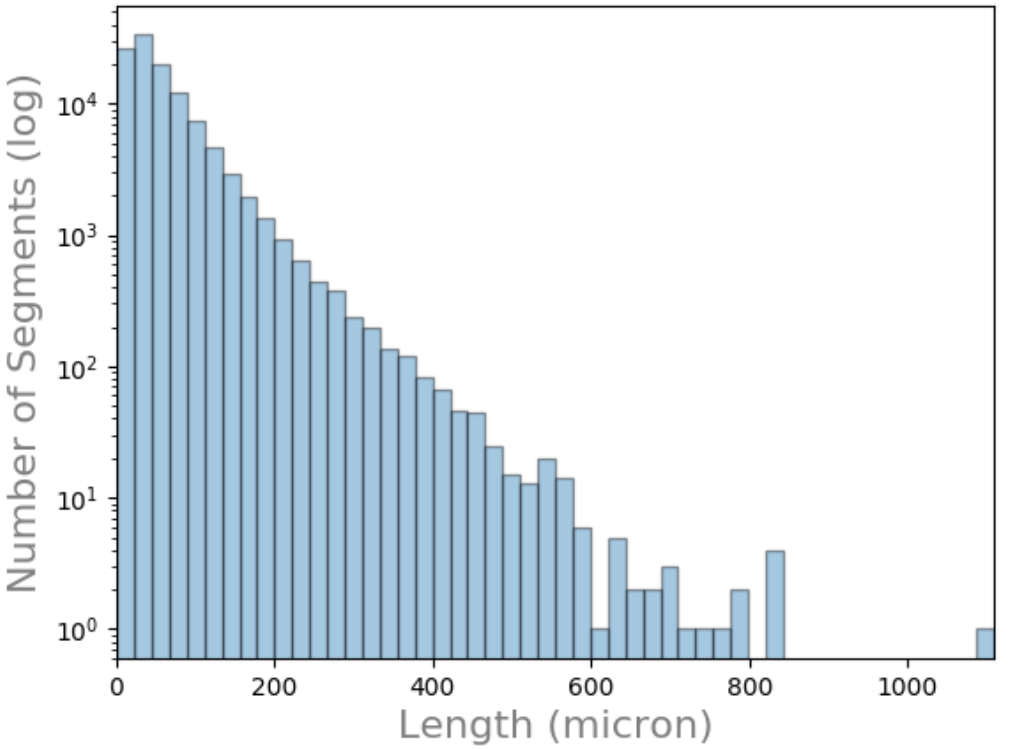


Figure 8.45: Small-sized vessels: number of vessel segments (log) by length.

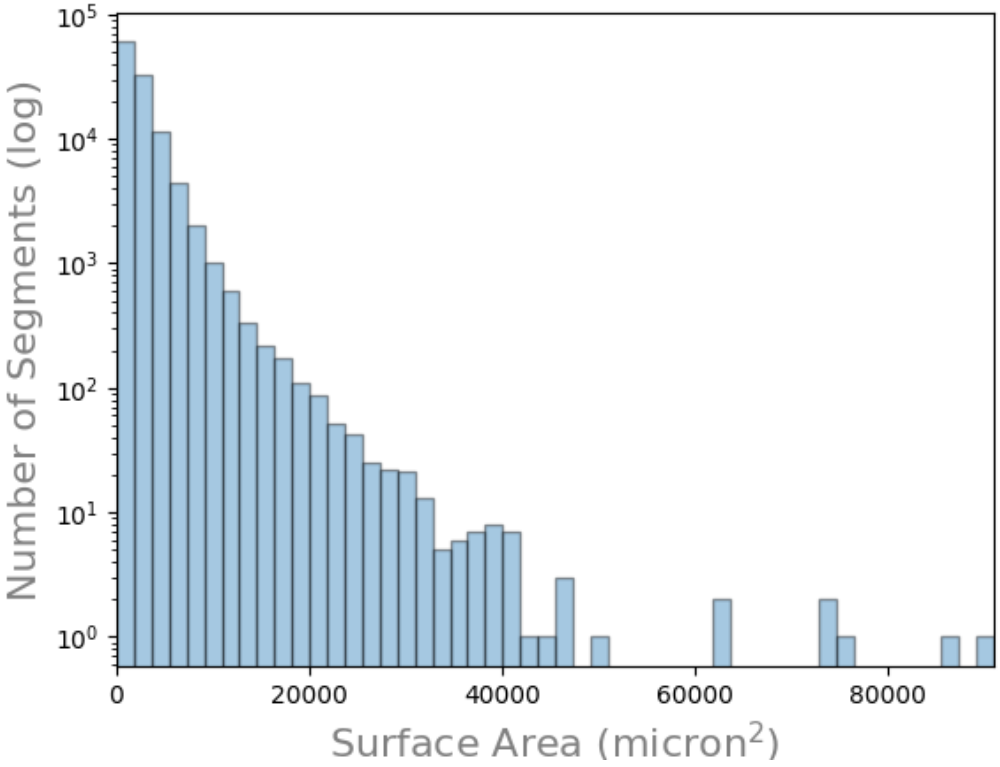


Figure 8.46: Small-sized vessels: number of vessel segments (log) by surface area.

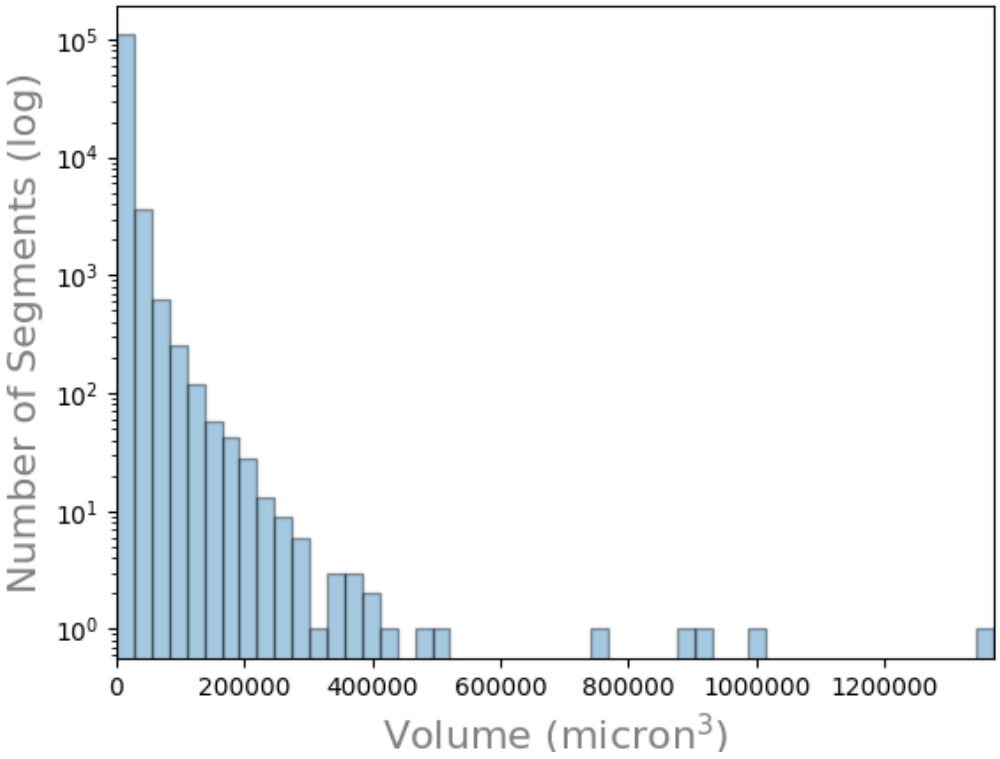


Figure 8.47: Small-sized vessels: number of vessel segments (log) by volume.

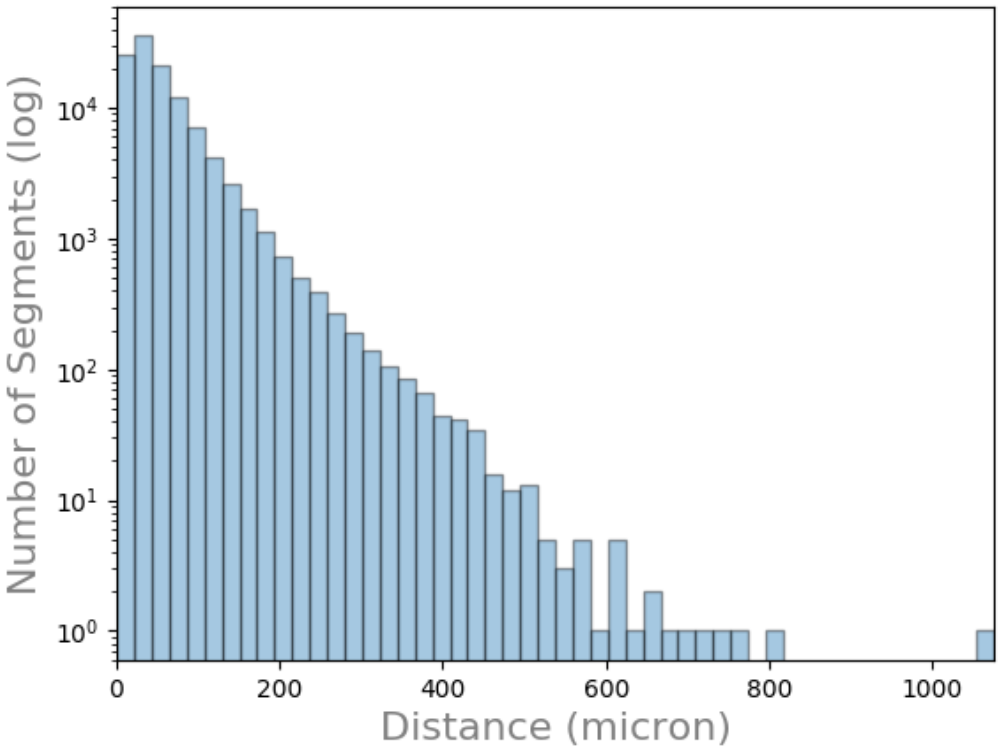


Figure 8.48: Small-sized vessels: number of vessel segments (log) by distance.

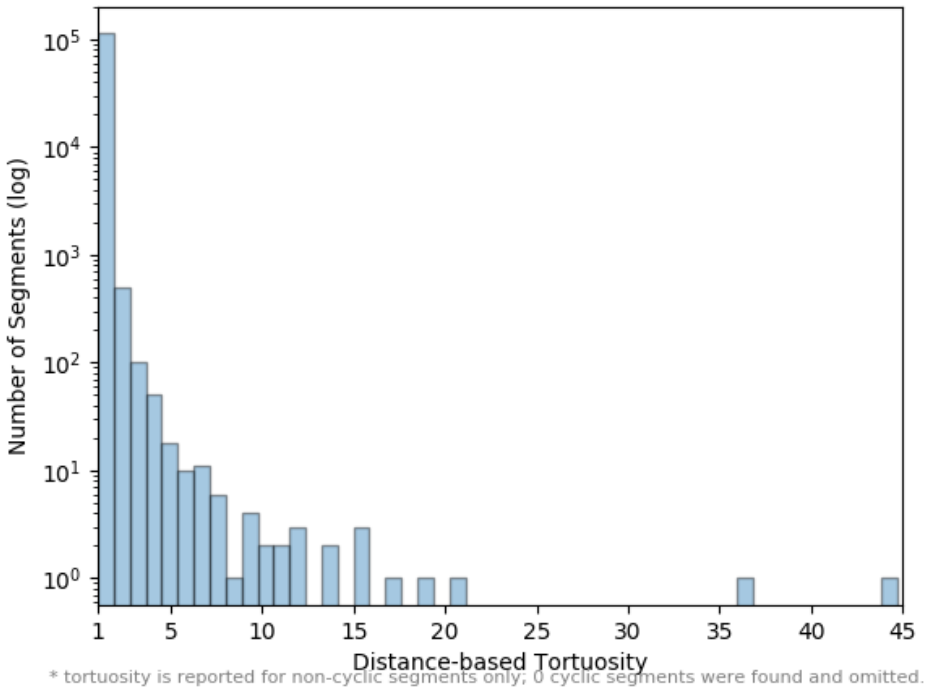


Figure 8.49: Small-sized vessels: number of vessel segments (log) by tortuosity.

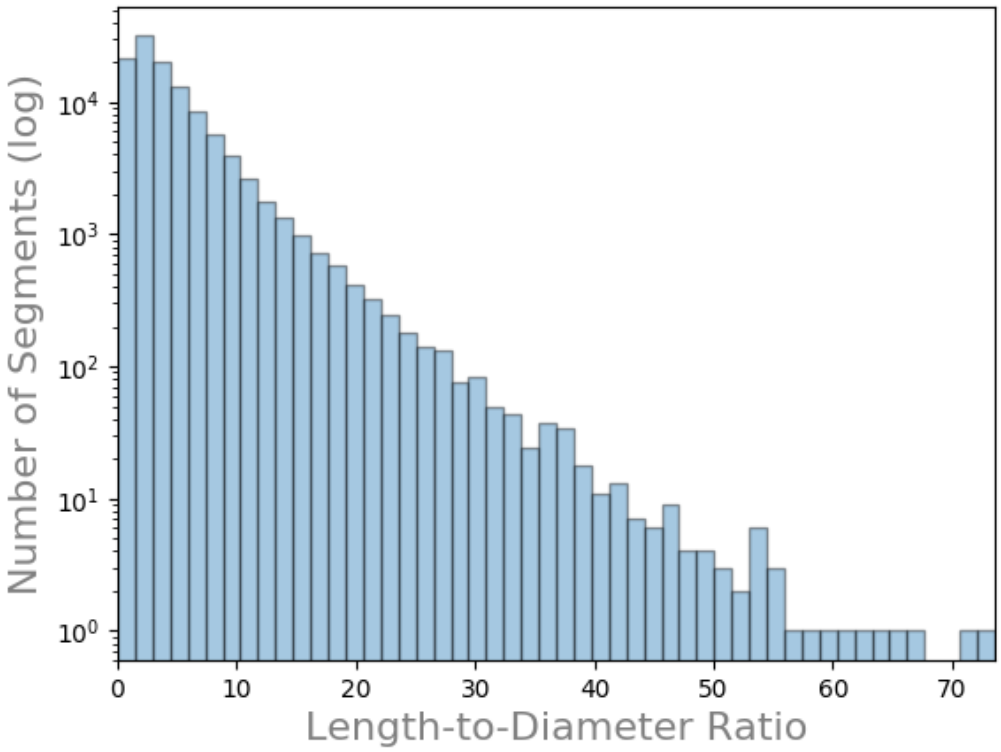


Figure 8.50: Small-sized vessels: number of vessel segments (log) by length-to-diameter ratio.

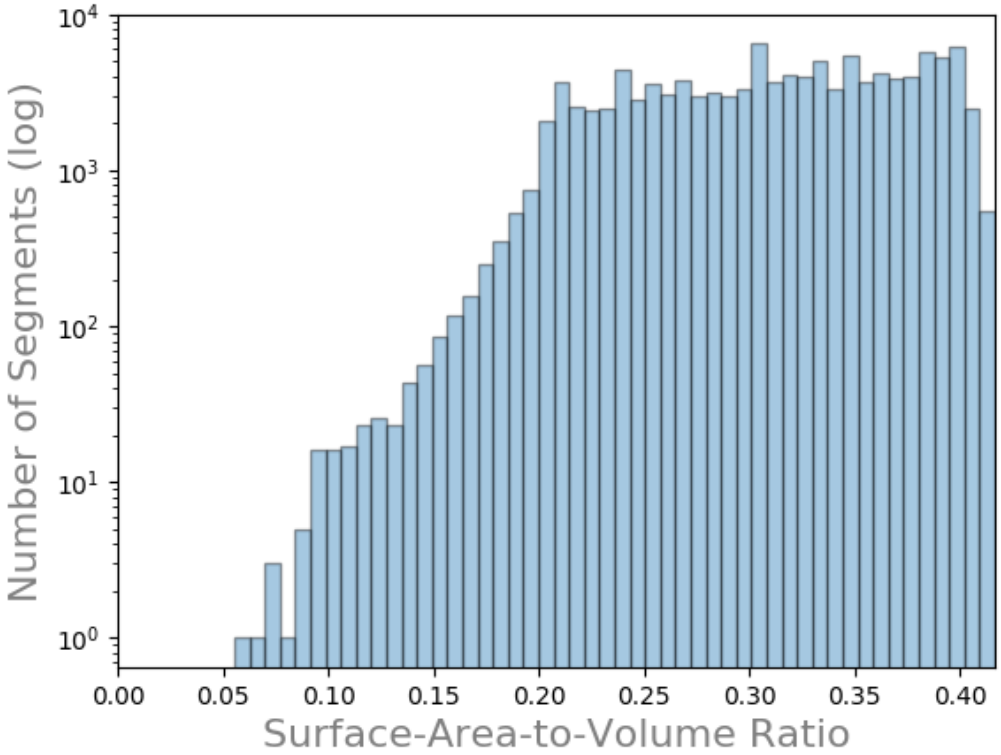


Figure 8.51: Small-sized vessels: number of vessel segments (log) by surface-area-to-volume ratio.

Table 8.6: Mean, median, minimum, and maximum values for vessel segment attributes calculated across the small-sized vessel segments.

Vessel Attribute	***Mean***	***Median***	***Min***	***Max***
Radius (DIA/2)	6.69 µm	6.34 µm	5.00 µm	10.00 µm
Length (LEN)	59.41 µm	42.25 µm	0.18 µm	1,109.14 µm
Surface Area (SA)	2,459.41 μm^2	1,731.67 μm^2	11.10 μm^2	91,064.73 μm^2
Volume (V)	8,841.39 μm^3	5,595.21 μm^3	53.29 μm^3	1,370,950.58 μm^3
Distance	54.89 µm	39.94 µm	0.18 µm	1,075.49 µm
Tortuosity	1.07	1.01	1.00	44.69
LEN:DIA	4.60	3.21	0.01	73.59
SA:V	0.31	0.32	0.06	0.42

8.3.3.3.4 Analysis of the Capillaries (radius $\leq 5\,\mu m$)

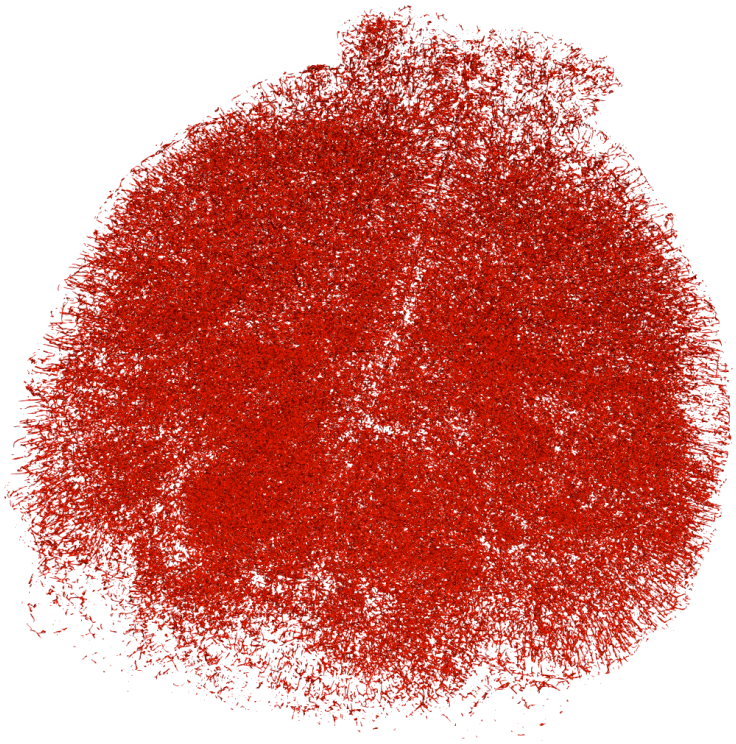


Figure 8.52: Capillaries presenting across the whole-brain dataset (transverse view: ↑, anterior; ↓, posterior).

We isolated the capillary segments by querying our database for segments with an average radius less than 5 (Figure 8.52). Using this data, we report the vascular characteristics for all capillaries as follows: total volume as $975{,}363{,}980.24\,\mu m^3$, total surface area as $402{,}058{,}481.73$ μm^2, and total length as $13{,}236{,}429.63\,\mu m$. We estimate the fractional volume of the vasculature as 0.10% of the whole-brain. We describe the attributes of these segments in Table 8.7. We present the distributions for radius in Figure 8.53, length in Figure 8.54, surface area in Figure 8.55, volume in Figure 8.56, and distance in Figure 8.57. We also detail the distributions for tortuosity, length-to-diameter ratio, and surface-area-to-volume ratio in Figures 8.58, 8.59, and 8.60 respectively.

Table 8.7: Mean, median, minimum, and maximum values for capillary segment attributes calculated across the whole-mouse brain.

Vessel Attribute	***Mean***	***Median***	***Min***	***Max***
Radius (DIA/2)	$4.81\,\mu m$	$4.80\,\mu m$	$4.80\,\mu m$	$5.00\,\mu m$
Length (LEN)	$37.21\,\mu m$	$24.20\,\mu m$	$0.25\,\mu m$	$692.48\,\mu m$
Surface Area (SA)	$1{,}130.38\,\mu m^2$	$731.03\,\mu m^2$	$7.54\,\mu m^2$	$40{,}423.75\,\mu m^2$
Volume (V)	$2{,}742.21\,\mu m^3$	$1{,}757.41\,\mu m^3$	$18.10\,\mu m^3$	$475{,}023.82\,\mu m^3$
Distance	$34.10\,\mu m$	$23.20\,\mu m$	$0.25\,\mu m$	$663.95\,\mu m$
Tortuosity	1.06	1.01	1.00	171.10
LEN:DIA	3.87	2.52	0.03	72.13
SA:V	0.42	0.42	0.09	0.42

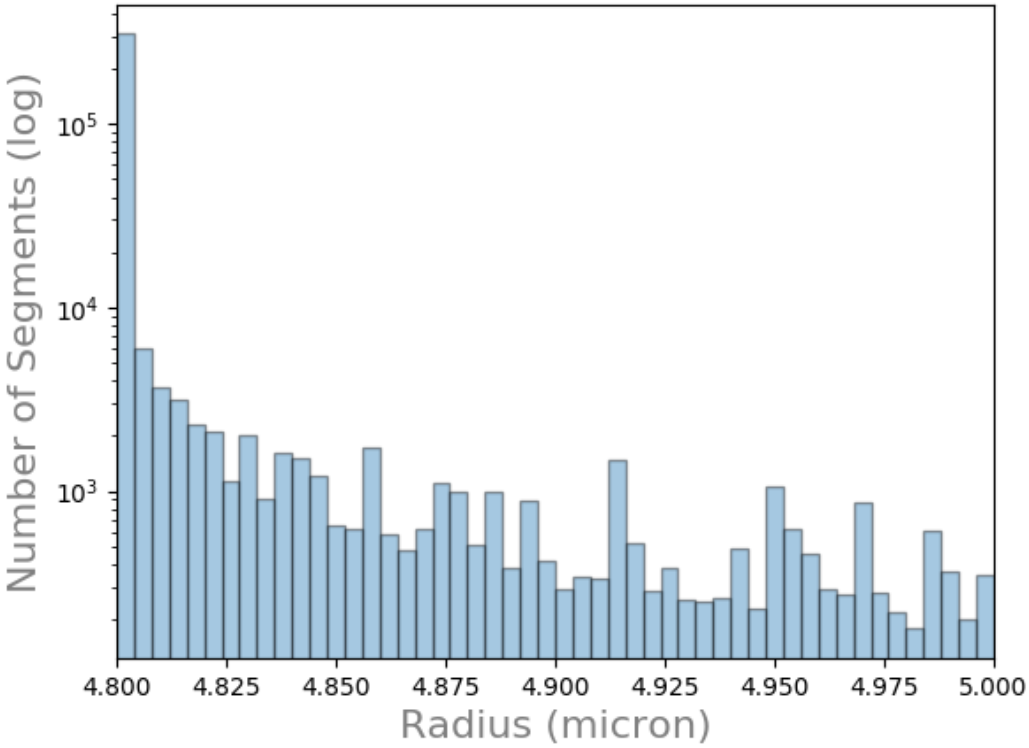


Figure 8.53: Capillaries: number of vessel segments (log) by radius.

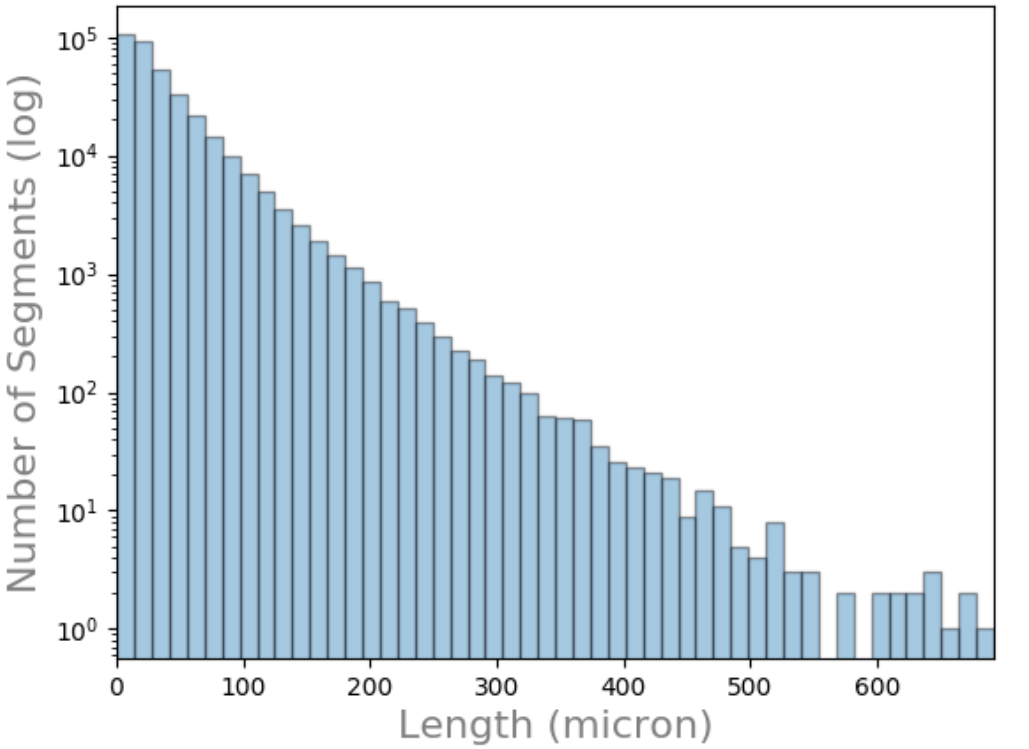


Figure 8.54: Capillaries: number of vessel segments (log) by length.

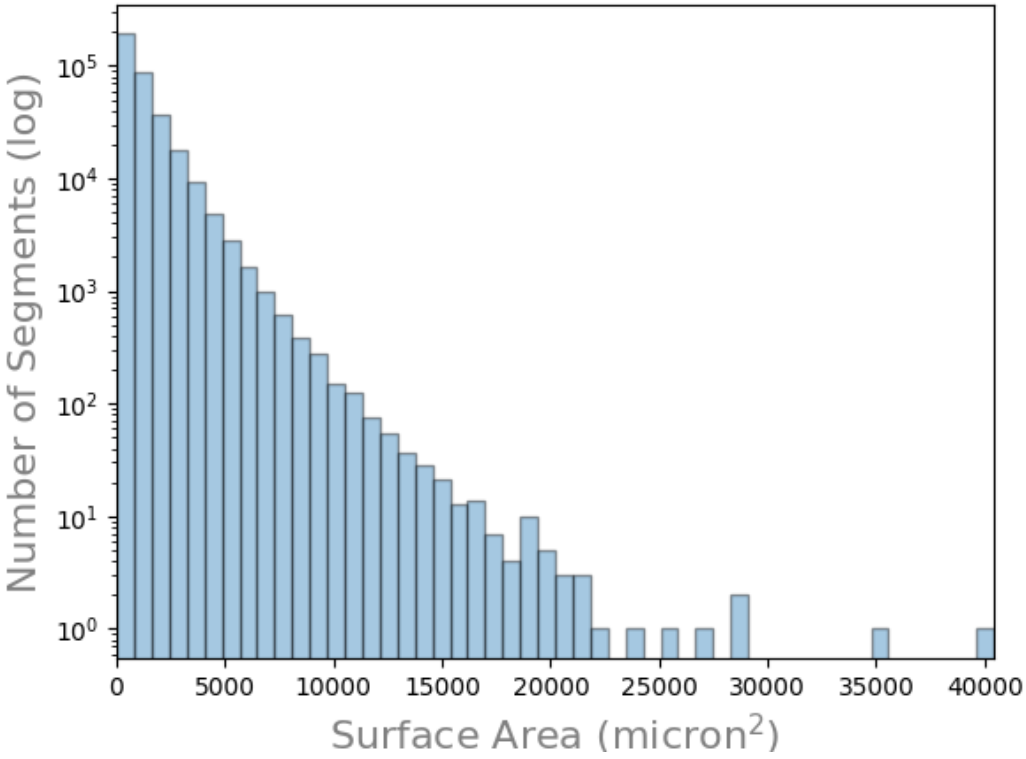


Figure 8.55: Capillaries: number of vessel segments (log) by surface area.

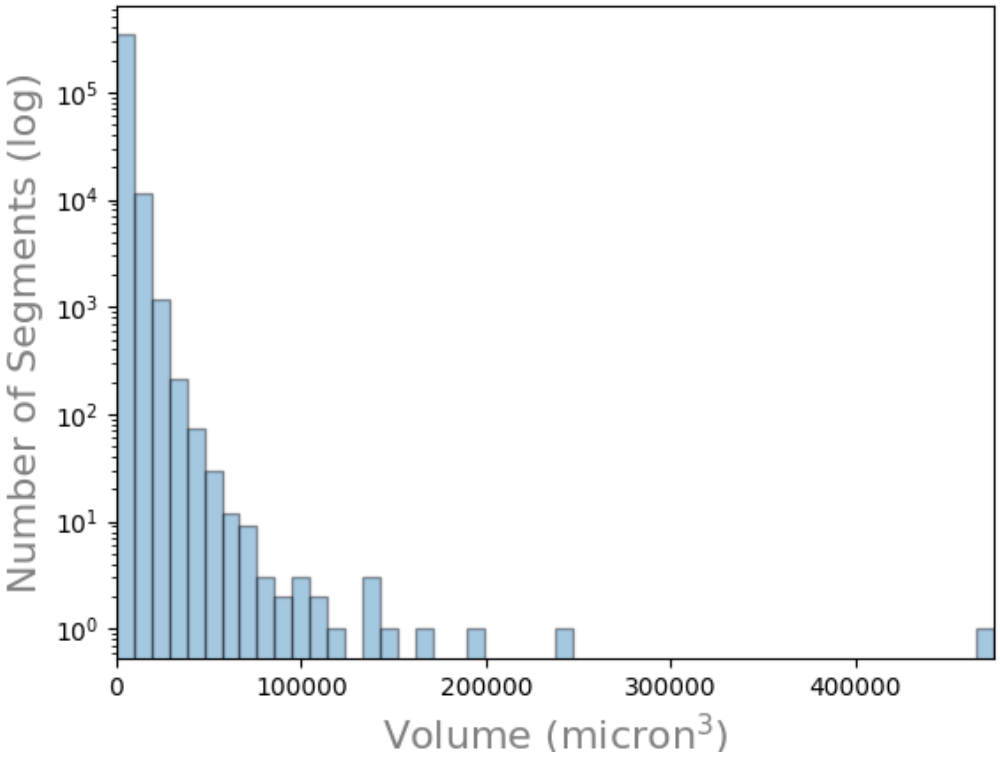


Figure 8.56: Capillaries: number of vessel segments (log) by volume.

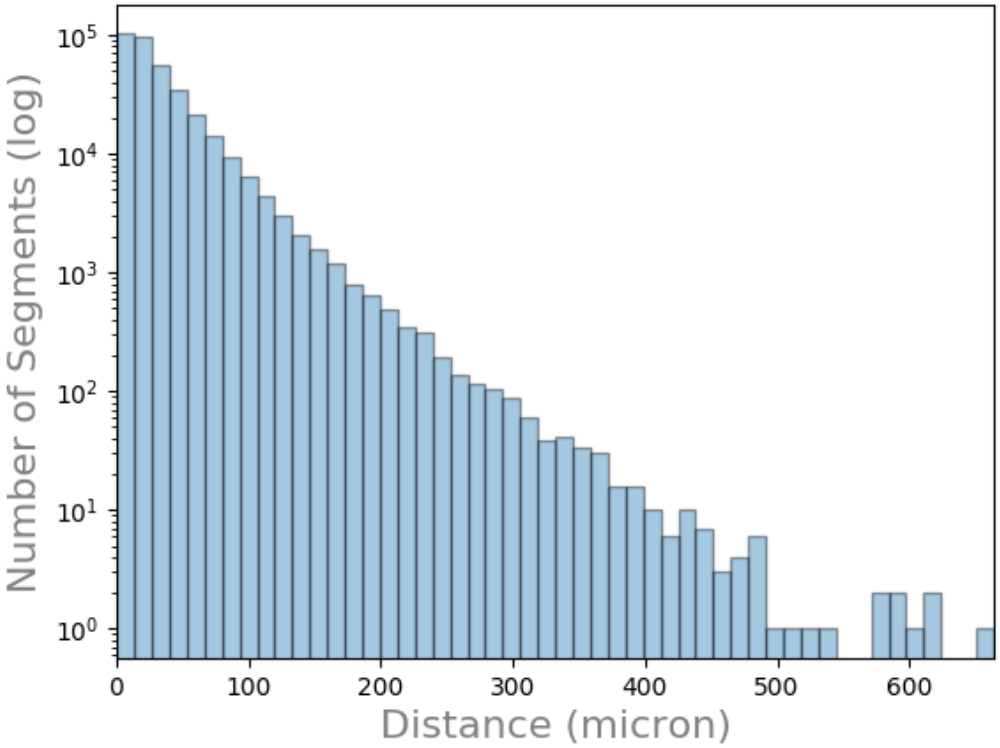


Figure 8.57: Capillaries: number of vessel segments (log) by distance.

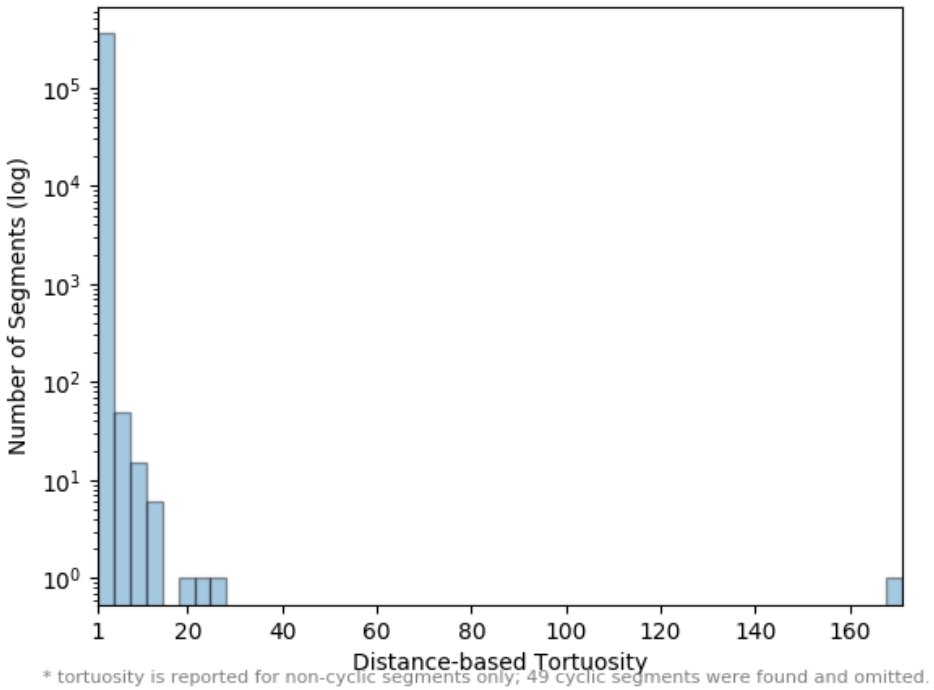


Figure 8.58: Capillaries: number of vessel segments (log) by tortuosity.

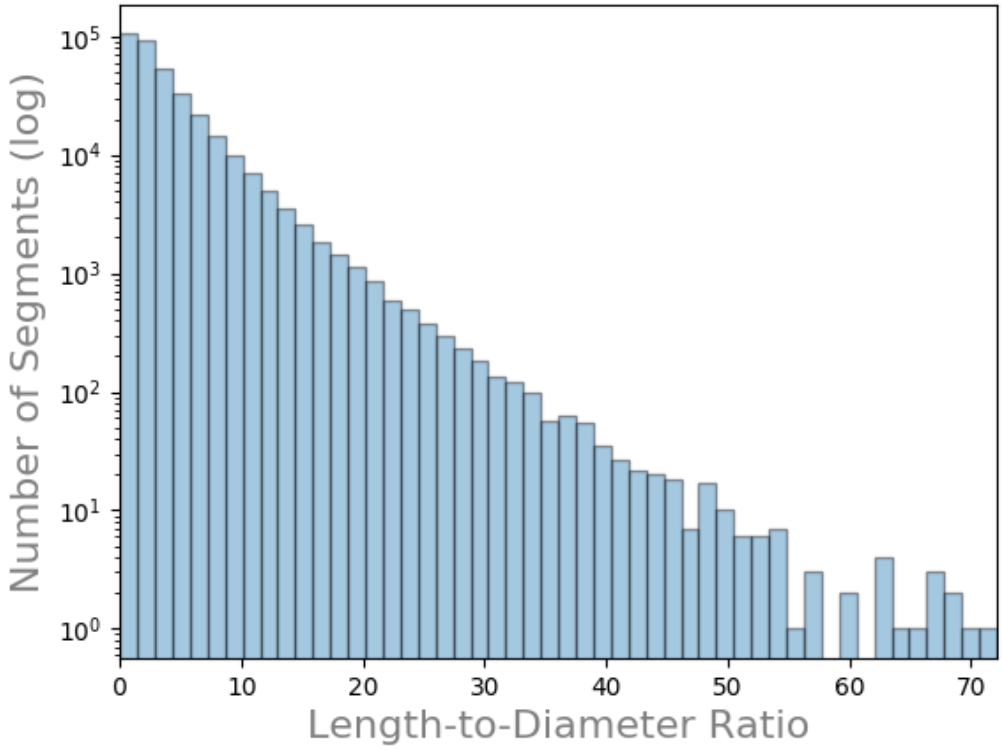


Figure 8.59: Capillaries: number of vessel segments (log) by length-to-diameter ratio.

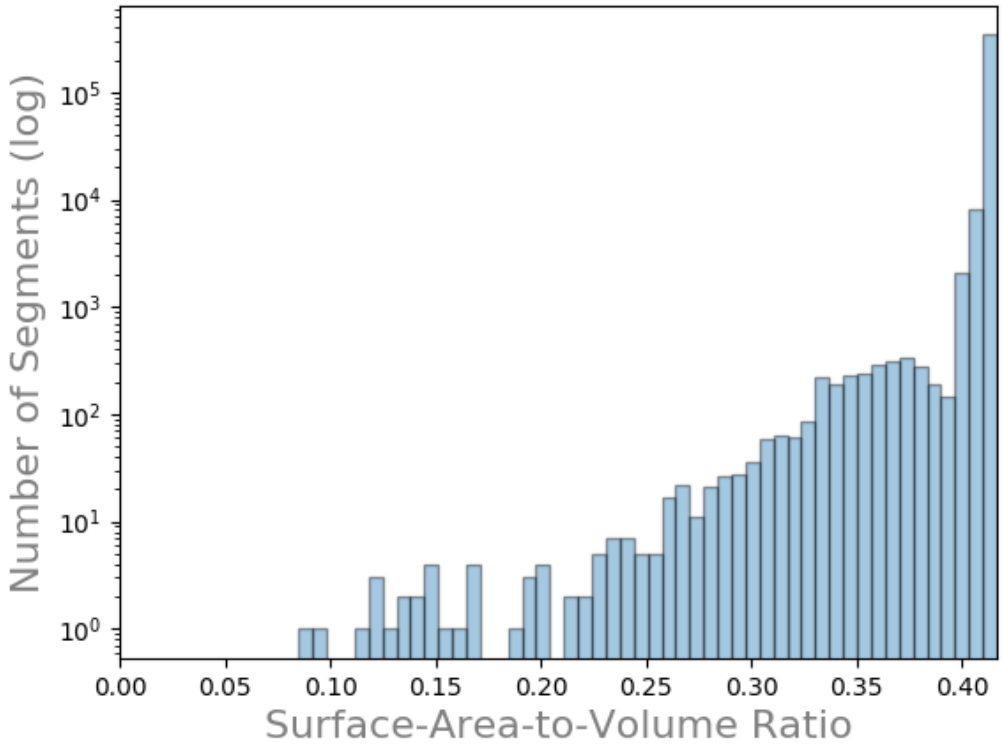


Figure 8.60: Capillaries: number of vessel segments (log) by surface-area-to-volume ratio.

Comparing and contrasting the different sized-vessels, including the capillaries

We present the distributions of vessel segment attributes side-by-side in Figure 8.61. We caution the reader to carefully analyze the histograms presented within that figure as the y-axis scales are different across the different vessel sizes. Half-page representations of these graphs have been presented earlier in this section, and we refer the reader to them if readability is difficult. Across the different classifications of vessels based on size (Table 8.2), we observe similar trends for number of segments (log) by segment attribute. We see the radius decrease in number within and across each class. For the large-sized vessels, the observed decrease in size can almost be modeled as a linear function for the log number of segments, albeit it does exhibit more characteristics of a power function. As we move towards smaller classifications of vessels, the linear characteristics diminish, and the power relationship begins to show more prevalence. Along the way, we see more spikes in radius for the medium- and small-sized vessels compared to the large-sized vessels and capillaries. While we are not sure as to the cause of these spikes (could be image processing artifacts or biological in nature), they do illustrate our method's potential to generate interesting questions. As for length, surface area, and volume, the shape of the number of segments (log) by respective attribute share similar trends to one another, which makes sense given their geometrical relationships to one another mathematically.

Tortuosity seems to be more variable within the larger classifications of vessels compared to the smaller-sized vessel groups. The values between groups are similar to one another, with each group exhibiting a small number of outliers and dominated by straight vessel segments, which was unexpected especially for the capillaries that should form a mesh-like network (and thus be tortuous). Tortuosity, like *all* the other distributions of attributes, is likely biased due to the preparation of the dataset by [6], primarily due to the damage to filament structure inevitably encountered during their extraction procedure. The length-to-diameter ratio shows similar trends across all vessel groups and exhibits the smoothest power function of all the vessel attributes. The surface-area-to-volume (SA:V) ratio distributions exhibit a seemingly normal distribution for the large- and medium-sized vessel groups that are positively skewed with some kurtosis. Therefore, the log segment count

decrease in number as SA:V increases. Meanwhile, the small-sized vessel and capillary groups exhibit a power function relationship of increasing log segment count with increasing SA:V value.

8.3.4 Validation against piecemeal data in the literature

Numerous studies in the literature have analyzed many aspects of the cerebrovasculature system, albeit with distinct focuses and across different cerebral regions. Such investigations have contributed vast amounts of piecemeal data about the brain's vascular system to the neurosciences. In this section, we intend to compare our results against the pre-existing data derived from the more authoritative of these studies as relevant to our whole-brain investigation.

First, we would like to determine the validity of the radius measurements that we have reported in Table 8.3. Our rationale for exploring this vessel attribute first is due to its use in the calculation of other attributes (e.x., surface area). For radius, we reported a means value of $5.92\,\mu\mathrm{m}$ and medium of $4.80\,\mu\mathrm{m}$. These results confirm those reported by Lee (2018) [6], who described an average radius of $5.31\,\mu\mathrm{m}$ and a medium of $4.80\,\mu\mathrm{m}$. Lee extracted the cerebrovasculature dataset used in our study and therefore it is not surprising that our results corroborate this. However, neither of our findings match exactly the piecemeal data reported in the literature.

Blinder et al. (2013) [29] reported a mean and median radius value approximately $2\,\mu\mathrm{m}$. This result was acquired after the analysis of a small dataset acquired using all-optical histology. Their focused result has been confirmed across the whole-brain by Zhang et al. (2018) [4], who applied the CLARITY protocol to the brain and subsequently imaged it using light-sheet microscopy. It would appear that our measurements of radius are off by a factor of two. We hypothesize that this difference is due to the preparation process, that the subsampling and morphological operations applied by Lee (2018) [6] lost the smallest vessels.

Meanwhile, the maximum radius size that we observed was $72.00\,\mu\mathrm{m}$. This observation is sensible. The major arteries and veins that were occupying the cerebrovasculature system of the mouse model used in our study measure to about this value [113]. In contrast, the minimum value that we recorded was $4.80\,\mu\mathrm{m}$. However, earlier investigations from our laboratory using small samples of the KESMBA India Ink dataset (the same dataset used in our investigation) measured

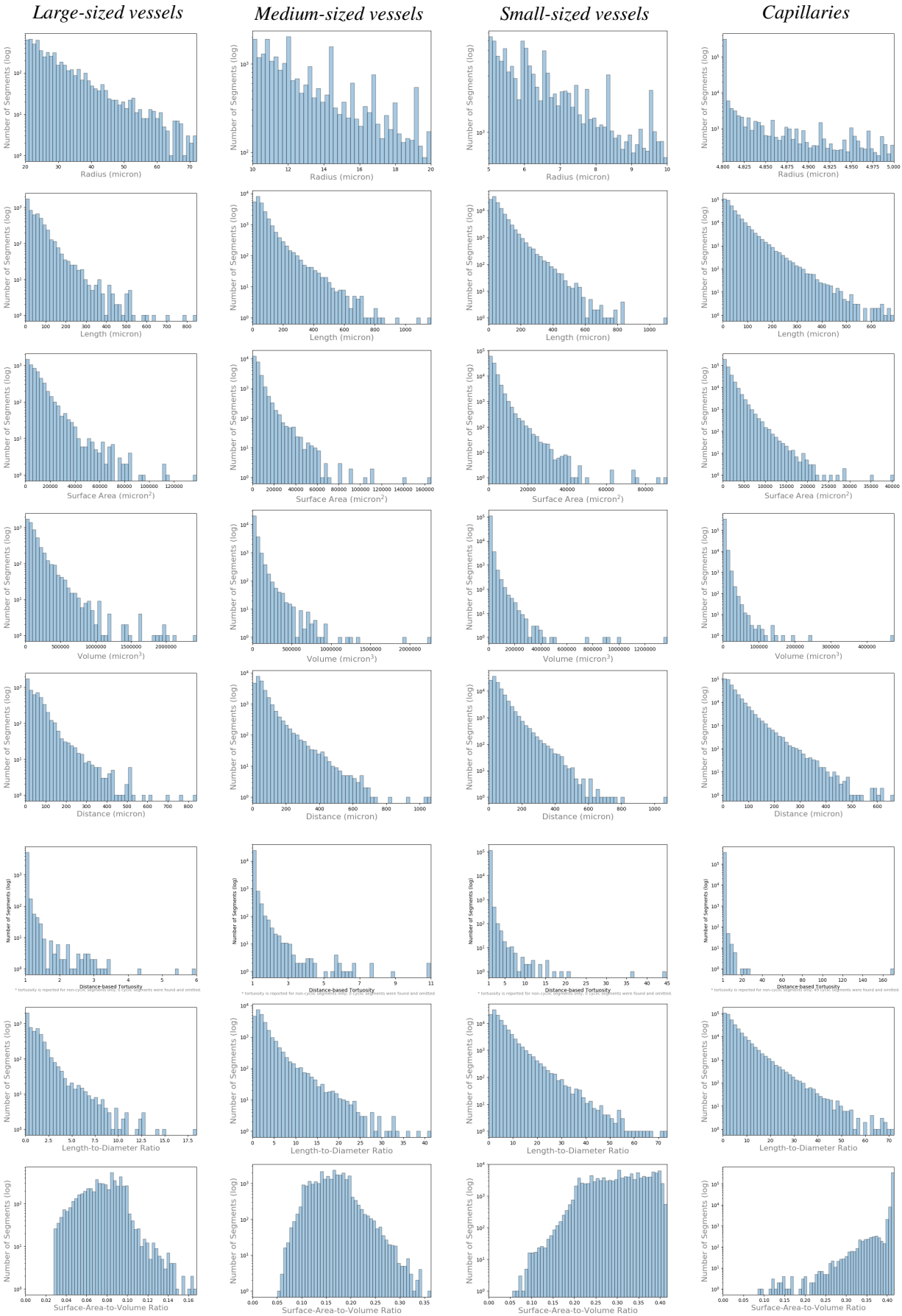


Figure 8.61: Side-by-side comparisons of vessel attributes across whole-brain.

radius values that were less than $4.80\,\mu\mathrm{m}$ [7, 13] at approximately $2.00\,\mu\mathrm{m}$. Therefore, we hypothesize that during the preparation process, the subsampling and morphological operations applied by Lee (2018) [6] inadvertently removed the smaller filaments. We would like the reader to keep in mind that the segmentation of filament data remains a challenge and, in most cases, the completeness and accuracy of the reconstruction of whole-brain datasets should always be questioned and challenged.

The range of vessel segment lengths that we reported across the whole-brain was broad, with a minimum segment length of $0.18\,\mu\mathrm{m}$ and maximum length of $1{,}167.92\,\mu\mathrm{m}$. The careful reader will notice that $0.18\,\mu\mathrm{m}$ is below the resolution of KESM. This is due to the image processing regiment applied by [6] to the raw data: small parts of a vessel can break off or become discontinuous with the parent vessel. We reported the mean and median lengths as $43.92\,\mu\mathrm{m}$ and $29.27\,\mu\mathrm{m}$ respectively. The most relevant data that we could find for comparison was a description of the median microvasculature length of $50\,\mu\mathrm{m}$ in the mouse by Blinder et al. (2010) [28]. Blinder et al.'s investigation were focused within the territory of the middle cerebral artery and near the cortical surface. We suspect that the length measures that we've reported are reasonable for our dataset. However, given the quality of the dataset, we believe that our measurements of length are also underestimated.

Our informatics platform uses the radius and length measurements of a segment to calculate its surface area and volume directly. The distributions of log segments by each of these attributes appear to follow the biological distributions, given that similar trends have been observed for rats [25], pigs [114, 115], and even humans [24, 10]. Therefore, the overall trends observed for radius, length, surface area, and volume (see Figure 8.61) in this study appear validated.

We used vessel segment length and distance to calculate tortuosity using the distance metric [87]. While we have not been able to find an adequate resource to compare our observations for distance, there are numerous investigations exploring tortuosity in the literature. The vasculature in the capillary network has been shown to be more torturous than the larger-sized vessels [84]. This is because the capillary network is mesh-like by nature [28]. The pial surface arteriole and

venule networks are also highly interconnected [31] and are likely to exhibit multiple twists and turns too. The penetrating vessels that connect the surface to the capillary networks, however, tend to dive straight into the cortex, at approximately 90-degrees [29]. The penetrating vessels have been characterized as being approximately $9\,\mu\mathrm{m}$ to $11\,\mu\mathrm{m}$ in diameter.

In our study, we have captured some of this variation. However, with our characterization of a vessel segment as simply the filaments presenting between branch points and/or endpoints, there is some information loss. In most studies, a vessel segment might have branches coming off of it. For instance, the middle cerebral artery would be considered a full vessel segment through some points, despite many branches coming off it. In our study, the middle cerebral artery is partitioned into many individual segments, which can affect the histogram distributions of the attributes. This means that the histograms for tortuosity in Figure 8.61 may be skewed as there may be many small segments hiding the true distribution. Therefore, we caution the reader against using the "absolute" values presented. Nonetheless, when looking at log vessel segments by tortuosity in Figure 8.61, the "relative" nature of our results does follow the biological predisposition that smaller vessels exhibit more tortuosity than the larger vessels.

The distributions of length-to-diameter ratios (LDR) and surface-area-to-volume ratios (SA:V) also follow the natural tendencies of nature when considered the vessel segments are grouped by size. This is evident in Figure 8.61, where we look at the distributions of log vessel segments by these attributes and compare them between our vessel size groupings (large-, medium, small-sized vessels and capillaries; Table 8.2). For LDR, the almost fractal nature of the tree-like tubes directing flow to the capillary mesh-like networks sees the steep decrease in LDR moving from large-sized vessels down to the capillaries [84]. This trend is reflected in our data. For SA:V, the geometrical implications of SA:V are reflected in biology. The large-sized vessels have a small SA:V as the pipes directing cerebral blood flow are interested in transport. As flow is directed to the capillary beds, SA:V will increase as to maximize the surface area available for fluid exchange to occur through diffusion and other means [15]. This trend is also reflected in our data due to the fluid mechanics of blood flow.

8.3.5 Dissemination of connectivity data

Currently, we disseminate our connectivity data as Neo4j data model files and through serialized *Networkx* graph binary files for the whole brain "updated original" and "reduced" graphs. The former allows researchers to download the data model and seamlessly import it into an instance of Neo4j. Meanwhile, the latter allows users to load the connectivity data directly into our informatics platform. The connectivity data can be queried and interacted with using either method; we provide the serialized files to increase accessibility to the data. All of these files are distributed in our Github repository, accessible via www.github.com/michaelrnowak.

8.4 Summary

The principal contribution of this work is the first (to our knowledge) queryable database of vascular connectivity across the whole brain. Our database contains an edge for every single vessel segment that was segmented successfully from the KESMBA India Ink dataset by Lee (2018) [6] along with that vessel's attributes. Each of these `SEGMENT` edges can be used to index into the `FILAMENT` edges encoding the geometry of that respective vessel segment. In this whole-brain study, we report the following volume characteristics across the whole brain:

- Total length
- Total surface area
- Total volume
- Total-surface-area-to-total-volume ratio

In addition to the volume characteristics, we also investigated the attributes of vessel segments and their distributions across the brain. Those attributes are:

- Radius
- Length
- Tortuosity, using the distance metric defined in [87]

- Distance (between the beginning and the terminal end of the segment)
- Surface area
- Volume
- Surface-area-to-volume ratio (i.e., SA:V)
- Length-to-diameter ratio (i.e., LEN:DIA)

We reported the distributions of log number of segments by each of these attributes across the entire mouse brain. Subsequently, we used our database to retrieve large-, medium-, small-sized vessels, and the capillaries to focus our analysis to blood vessels meeting the size criteria denoted in Table 8.2. We analyzed each group of vessels individually, describing their volume characteristics and exploring their distributions of attributes. Although not illustrated, we emphasize that different graph algorithms can be applied to our database to isolate subgraphs meeting some criteria, upon which additional analyses can be performed. Finally, we compared the trends that we observed across the whole-brain against the relevant piecemeal data available in the literature. We understand that there are some quality issues associated with the dataset used in this study; we affirm and acknowledge those issues here. However, it is important to recognize that the framework presented throughout this dissertation can be applied to cerebrovasculature datasets acquired with different modalities, resolution, and volume. Further, by incorporating our work, *all that is needed to explore the whole-brain are database queries*.

We expect that our methods will open the door for data-driven, quantitative investigation of the brain's vascular system. For instance, in order to understand the development and function of cerebral vasculature in the brain, it is important to map out its anatomical structure in the healthy brain and compare differences to the disordered brain. Accordingly, we could construct databases of many healthy brains and explore the variance of many vessel attributes within them. Interesting attributes of the network can be elucidated using our data-driven approach, such as the profound spikes in radius for the medium-sized and small-sized vessels evident in Figures 8.35 and 8.44.

While we are not sure as to the cause of these spikes (could be image processing artifacts or biological in nature), they do illustrate our method's potential to generate interesting questions. In general, our data models for healthy brains could be compared against those acquired from an animal model of human pathology. We believe that such whole-brain studies are the key to the discovery of novel biomarkers of processes underlying healthy and diseased brains. In this chapter, we have shown that such studies are indeed achievable with our methods.

9. DISCUSSION AND CONCLUSION

In this chapter, we will consider the future directions of our work, provide a summary of the specifics aims that we have accomplished, detail the contributions of our work, and close by discussing the expected impact of our dissertation.

9.1 Future Work

In future work, we should investigate the cerebrovasculature of healthy murine brains further and begin studying its presentation in murine models of human pathology. This would begin through the construction of a whole-brain connectivity database that is more complete and exhibits greater coherence than that prepared by [6] and used in Chapter 8 to illustrate our methodologies. Once that connectivity database is prepared, a comparable analysis to that presented in Chapter 8 should commence. The results of this analysis should be compared to both our own and the piecemeal data in the literature. If the conformance is acceptable, we believe that the graph database should be leveraged to its fullest potential for further analysis. This would include the application of graph algorithms similar to those applied [31] to determine the robust and fragile aspects of the cerebrovasculature network. Indeed, [31] has characterized this well within the territory of the middle cerebral artery. Therefore, future investigations would first confirm the findings of [31] before charting the less studied regions of the brain.

The study of a single, healthy brain at micrometer resolution yields significant information about the individual specimen. However, this provides little information about the variation of cerebrovasculature across healthy brains. Therefore, the next step would be to acquire many whole-brain datasets for healthy brains and construct a database for each. The variation in cerebrovasculature structure could then be reported globally and for localized regions of interest.

Once the cerebrovasculature of healthy brains has been characterized, the next step is to acquire databases of murine models of human pathology. The variation in cerebrovasculature structure could be determined globally and for localized regions of interest within the disease model.

These results would then be compared between the healthy and pathological brains. This would provide significant insight as to how the pathophysiology of some disease may affect the cerebrovasculature system. Furthermore, whether pathogenesis is biased to the disruption of a specific cortical, or subcortical, the region could be explored. These are questions that could be answered using our technique. More elaborate studies could be conducted by obtaining brain datasets acquired at different periods from disease onset. Such longitudinal data could help elucidate how the cerebrovasculature is disrupted as a function of time since the onset of the pathological condition.

Future work using the application of our methods for investigations into the cerebrovasculature and of cerebral blood flow will require significant collaboration with neuroscientists and others interested in those topics. In the interim, it would make sense to focus on expanding our informatics platform and graph database as they do have such dependencies on outside personnel. We believe that the most significant addition would be to label whole segments. For instance, in our method, the middle cerebral artery is encoded as several segments presenting between bifurcations. Together, these segments comprise this major artery; however, they are analyzed independently. We would like to represent and calculate attributes for the middle cerebral artery in its entirety. We would intend to do this for all vessels. There are many approaches to determining where one vessel starts, and another vessel ends. It all comes down to a new definition of a segment and an algorithm that can extract segments based upon that definition. For instance, in our work, we stated that a segment was the vasculature presenting between points of bifurcation. Using a diameter-defined Strahler model scheme, one can define a segment in a manner that allows the segment to be defined across points of bifurcation. The representation of such a segment in our graph database would need to be considered.

We also need to publish our whole-brain vascular connectivity data model online (in a Neo4j database) eventually. Doing so will allow users to interact with the data through the Neo4j Browser and/or query with the database from our informatics platform. For instance, by interacting with the database in this manner, relevant information can be pulled down from the Internet, alleviating the need for the user to store the whole-brain connectivity data locally. Therefore, future work should

incorporate a plan to publish the data model once funding is in place to support the necessary infrastructure.

9.2 Conclusion

In this dissertation I have accomplished the following *specific aims*:

1. Generation of synthetic cerebrovasculature models. Create large-scale, biologically-grounded, artificial cerebrovascular volumes for the validation of segmentation procedures. *Completed in Chapters 4 and 3*.

2. Model-based validation system. Design a validation method that evaluates the correctness of segmentation algorithms over large-scale, biologically-grounded, synthetic cerebrovasculature models. *Completed in Chapter 5*.

3. Informatics platform. Develop an open-source informatics platform that: (1) converts data to graph representation; (2), calculates descriptive statistics describing vascular filament attributes across every element in the model, while characterizing the encoded cerebrovascular volume as a whole; and, (3) convert graph representation to the data model for a graph database. *Completed in Chapter 6*.

4. Whole-brain analysis. Perform quantitative analysis across the Knife-Edge Scanning Microscope Brain Atlas India ink whole-brain dataset: (1) assess algorithmic performance of the segmentation algorithm used to extract the vascular filaments using the model-based validation system; (2) use the informatics platform to quantitatively the describe specific attributes and characteristics of the mouse cerebrovascular system; and, (3) disseminate raw connectivity data online. *Completed in Chapter 8*.

Our work has contributed a synthetic model-based validation system to assess segmentation algorithm performance, an informatics platform for vascular analysis, and queryable database of vascular connectivity. Together, these contributions provide an intuitive and systematic way to

characterize the cerebrovasculature system of murine animal models. We have illustrated the capability of the final product of this system in Chapter 8. Using our cerebrovasculature database, we were able to study the cerebrovasculature system by issuing text-based queries to extract the vessel segments that we were interested in. The outcome of our investigation was a wealth of information about the cerebrovasculature system as a whole, and about the different classifications of vessels comprising it. These simple queries even revealed profound spikes in radius for the medium-sized and small-sized vessels evident in Figures 8.35 and 8.44. While we are not sure as to the cause of these spikes – they could be image processing artifacts or biological – they do illustrate the potential for queries to generate interesting biological questions.

We expect that the methods described in this dissertation will open the door for data-driven, quantitative investigation across the whole-brain. At the time of writing – and to the best of our knowledge that prior to this work – there was not a systemic way to assess segmentation algorithm performance, calculate attributes for each segment of vasculature extracted across the whole brain, and store those results in a queryable database that also stores geometry and topology of the entire cerebrovasculature system. We believe that our method can and will set the standard for large-scale cerebrovasculature research. Therefore, in conclusion, we state that our methods contribute a standardized, quantitative approach to the study of cerebrovasculature datasets acquired using modern imaging techniques.

REFERENCES

[1] P. Osten and T. W. Margrie, "Mapping brain circuitry with a light microscope.," *Nature methods*, vol. 10, pp. 515–523, June 2013.

[2] M. J. Pesavento, C. Miller, K. Pelton, M. Maloof, C. E. Monteith, V. Vemuri, and M. Klimen, "Knife-edge scanning microscopy for bright-field multi-cubic centimeter analysis of microvasculature," *Microscopy Today*, vol. 25, no. 4, pp. 14–21, 2017.

[3] D. Mayerich, B. H. McCormick, and J. Keyser, "Noise and artifact removal in knife-edge scanning microscopy," in *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 556–559, IEEE, 2007.

[4] L.-Y. Zhang, P. Lin, J. Pan, Y. Ma, Z. Wei, L. Jiang, L. Wang, Y. Song, Y. Wang, Z. Zhang, K. Jin, Q. Wang, and G.-Y. Yang, "CLARITY for High-resolution Imaging and Quantification of Vasculature in the Whole Mouse Brain," *Aging and Disease*, vol. 9, no. 2, p. 262, 2018.

[5] W. Ding, A. Li, J. Wu, Z. Yang, Y. Meng, S. Wang, and H. Gong, "Automatic macroscopic density artefact removal in a nissl-stained microscopic atlas of whole mouse brain," *Journal of microscopy*, vol. 251, no. 2, pp. 168–177, 2013.

[6] J. Lee, *Mapping and Analyzing the Full Vascular Network in the Mouse Brain at Submicrometer Resolution*. PhD dissertation, Texas A&M University, Aug. 2018.

[7] D. Mayerich, J. Kwon, Y. Choe, L. Abbott, and J. Keyser, "Constructing high resolution microvascular models," in *Third Workshop on Microscopic Image Analysis with Applications in Biology*, 2008.

[8] J. A. Tyrrell, B. Roysam, E. di Tomaso, R. Tong, E. B. Brown, and R. K. Jain, "Robust 3-D modeling of tumor microvasculature using superellipsoids.," *ISBI*, 2006.

[9] A. Rodriguez, D. B. Ehlenberger, P. R. Hof, and S. L. Wearne, "Three-dimensional neuron tracing by voxel scooping.," *Journal of neuroscience methods*, vol. 184, pp. 169–175, Oct. 2009.

[10] F. Cassot, F. Lauwers, C. Fouard, S. Prohaska, and V. Lauwers-cances, "A novel three-dimensional computer-assisted method for a quantitative study of microvascular networks of the human cerebral cortex.," *Microcirculation*, vol. 13, pp. 1–18, Jan. 2006.

[11] M. R. Nowak, D. Han, and Y. Choe, "Biologically grounded synthetic cerebrovasculature models for validation of segmentation algorithms," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pp. 714–717, IEEE, 2017.

[12] D. H. Han, *Rapid 3D Tracing Of The Mouse Brain Neurovasculature With Local Maximum Intensity Projection And Moving Windows*. PhD dissertation, Texas A&M University, Aug. 2009.

[13] D. Mayerich, J. Kwon, C. Sung, L. Abbott, J. Keyser, and Y. Choe, "Fast macro-scale transmission imaging of microvascular networks using KESM.," *Biomedical optics express*, vol. 2, pp. 2888–2896, Oct. 2011.

[14] B. J. Andreone, B. Lacoste, and C. Gu, "Neuronal and vascular interactions," *Annual review of neuroscience*, vol. 38, pp. 25–46, 2015.

[15] D. Rubenstein, W. Yin, and M. D. Frame, *Biofluid mechanics: an introduction to fluid mechanics, macrocirculation, and microcirculation*. Academic Press, 2015.

[16] J. G. Betts, P. Desaix, E. Johnson, J. E. Johnson, O. Korol, D. Kruse, B. Poe, J. A. Wise, M. Womble, and K. A. Young, *Anatomy and Physiology*. OpenStax, 2013.

[17] F. Cassot, F. Lauwers, S. Lorthois, P. Puwanarajah, V. Cances-Lauwers, and H. Duvernoy, "Branching patterns for arterioles and venules of the human cerebral cortex," *Brain research*, vol. 1313, pp. 62–78, 2010.

[18] M. R. Nowak, A. Lozovskiy, D. Dobroskok, and Y. Choe, "Knife-edge scanning microscopy for in silico study of cerebral blood flow: From biological imaging data to flow simulations,"

in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pp. 5957–5960, IEEE, 2016.

[19] D. Purves, G. J. Augustine, D. Fitzpatrick, L. C. Katz, A.-S. LaMantia, J. O. McNamara, and S. M. Williams, *Neuroscience*. Sunderland, MA: Sinauer Associates, 2008.

[20] P. Carmeliet, "Angiogenesis in life, disease and medicine," *Nature*, vol. 438, no. 7070, p. 932, 2005.

[21] A. R. Pries, B. Reglin, and T. W. Secomb, "Structural response of microcirculatory networks to changes in demand: information transfer by shear stress," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 284, no. 6, pp. H2204–H2212, 2003.

[22] P. T. So, "Two-photon fluorescence light microscopy," *eLS*, 2001.

[23] G. S. Kassab, C. A. Rider, N. J. Tang, and Y.-C. Fung, "Morphometry of pig coronary arterial trees," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 265, no. 1, pp. H350–H365, 1993.

[24] F. Lauwers, F. Cassot, V. Lauwers-Cances, P. Puwanarajah, and H. Duvernoy, "Morphometry of the human cerebral cortex microcirculation: general characteristics and space-related profiles," *Neuroimage*, vol. 39, no. 3, pp. 936–948, 2008.

[25] D. Lapi, P. Marchiafava, and A. Colantuoni, "Geometric characteristics of arterial network of rat pial microcirculation," *Journal of vascular research*, vol. 45, no. 1, pp. 69–77, 2008.

[26] R. E. Horton, "Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology," *Geological society of America bulletin*, vol. 56, no. 3, pp. 275–370, 1945.

[27] G. Silasi and T. H. Murphy, "Stroke and the Connectome: How Connectivity Guides Therapeutic Intervention," *NEURON*, vol. 83, pp. 1354–1368, Sept. 2014.

[28] P. Blinder, A. Y. Shih, C. Rafie, and D. Kleinfeld, "Topological basis for the robust distribution of blood to rodent neocortex," *Proceedings of the National Academy of Sciences*, vol. 107, no. 28, pp. 12670–12675, 2010.

[29] P. Blinder, P. S. Tsai, J. P. Kaufhold, P. M. Knutsen, H. Suhl, and D. Kleinfeld, "The cortical angiome: an interconnected vascular network with noncolumnar patterns of blood flow," *Nature neuroscience*, vol. 16, no. 7, p. 889, 2013.

[30] N. Nishimura, C. B. Schaffer, B. Friedman, P. D. Lyden, and D. Kleinfeld, "Penetrating arterioles are a bottleneck in the perfusion of neocortex," *Proceedings of the National Academy of Sciences*, vol. 104, no. 1, pp. 365–370, 2007.

[31] A. Y. Shih, C. Rühlmann, P. Blinder, A. Devor, P. J. Drew, B. Friedman, P. M. Knutsen, P. D. Lyden, C. Mateo, L. Mellander, *et al.*, "Robust and fragile aspects of cortical blood flow in relation to the underlying angioarchitecture," *Microcirculation*, vol. 22, no. 3, pp. 204–218, 2015.

[32] A. Y. Shih, P. Blinder, P. S. Tsai, B. Friedman, G. Stanley, P. D. Lyden, and D. Kleinfeld, "The smallest stroke: occlusion of one penetrating vessel leads to infarction and a cognitive deficit," *Nature neuroscience*, vol. 16, no. 1, p. 55, 2013.

[33] C. B. Schaffer, B. Friedman, N. Nishimura, L. F. Schroeder, P. S. Tsai, F. F. Ebner, P. D. Lyden, and D. Kleinfeld, "Two-photon imaging of cortical surface microvessels reveals a robust redistribution in blood flow after vascular occlusion," *PLoS biology*, vol. 4, no. 2, p. e22, 2006.

[34] P. Osten and T. W. Margrie, "Mapping brain circuitry with a light microscope.," *Nature methods*, vol. 10, pp. 515–523, June 2013.

[35] A. J. Ewald, H. Mcbride, M. Reddington, S. E. Fraser, and R. Kerschmann, "Surface imaging microscopy, an automated method for visualizing whole embryo samples in three dimensions at high resolution.," *Developmental dynamics : an official publication of the American Association of Anatomists*, vol. 225, pp. 369–375, Nov. 2002.

[36] P. S. Tsai, B. Friedman, A. I. Ifarraguerri, B. D. Thompson, V. Lev-Ram, C. B. Schaffer, Q. Xiong, R. Y. Tsien, J. A. Squier, and D. Kleinfeld, "All-optical histology using ultrashort laser pulses.," *NEURON*, vol. 39, pp. 27–41, July 2003.

[37] G. B. Sands, D. A. Gerneke, D. A. Hooks, C. R. Green, B. H. Smaill, and I. J. Legrice, "Automated imaging of extended tissue volumes using confocal microscopy.," *Microscopy research and technique*, vol. 67, pp. 227–239, Aug. 2005.

[38] T. Ragan, J. D. Sylvan, K. H. Kim, H. Huang, K. Bahlmann, R. T. Lee, and P. T. C. So, "High-resolution whole organ imaging using two-photon tissue cytometry.," *Journal of biomedical optics*, vol. 12, p. 014015, Jan. 2007.

[39] D. Mayerich, L. Abbott, and B. McCormick, "Knife-edge scanning microscopy for imaging and reconstruction of three-dimensional anatomical structures of the mouse brain.," *Journal of microscopy*, vol. 231, pp. 134–143, July 2008.

[40] A. Li, H. Gong, B. Zhang, Q. Wang, C. Yan, J. Wu, Q. Liu, S. Zeng, and Q. Luo, "Micro-Optical Sectioning Tomography to Obtain a High-Resolution Atlas of the Mouse Brain," *Science*, vol. 330, pp. 1404–1408, Dec. 2010.

[41] T. Ragan, L. R. Kadiri, K. U. Venkataraju, K. Bahlmann, J. Sutin, J. Taranda, I. Arganda-Carreras, Y. Kim, H. S. Seung, and P. Osten, "Serial two-photon tomography for automated ex vivo mouse brain imaging.," *Nature methods*, vol. 9, pp. 255–258, Jan. 2012.

[42] H. Gong, S. Zeng, C. Yan, X. Lv, Z. Yang, T. Xu, Z. Feng, W. Ding, X. Qi, A. Li, J. Wu, and Q. Luo, "Continuously tracing brain-wide long-distance axonal projections in mice at a one-micron voxel resolution.," *NeuroImage*, vol. 74, pp. 87–98, July 2013.

[43] J. Yuan, H. Gong, A. Li, X. Li, S. Chen, S. Zeng, and Q. Luo, "Visible rodent brain-wide networks at single-neuron resolution.," *Frontiers in neuroanatomy*, vol. 9, p. 70, 2015.

[44] W. Denk, J. H. Strickler, and W. W. Webb, "Two-photon laser scanning fluorescence microscopy.," *Science*, vol. 248, pp. 73–76, Apr. 1990.

[45] H. Siedentopf and R. Zsigmondy, "Uber sichtbarmachung und größenbestimmung ultramikoskopischer teilchen, mit besonderer anwendung auf goldrubingläser," *Annalen der Physik*, vol. 315, no. 1, pp. 1–39, 1902.

[46] W. Spalteholz *et al.*, "Uber das durchsichtigmachen von menschlichen und tierischen praparaten und seine theoretischen bedingungen, nebst anhang," 1914.

[47] H.-U. Dodt, U. Leischner, A. Schierloh, N. Jährling, C. P. Mauch, K. Deininger, J. M. Deussing, M. Eder, W. Zieglgänsberger, and K. Becker, "Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain," *Nature methods*, vol. 4, no. 4, p. 331, 2007.

[48] K. Chung and K. Deisseroth, "Clarity for mapping the nervous system," *Nature methods*, vol. 10, no. 6, p. 508, 2013.

[49] H. Hama, H. Kurokawa, H. Kawano, R. Ando, T. Shimogori, H. Noda, K. Fukami, A. Sakaue-Sawano, and A. Miyawaki, "Scale: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain," *Nature neuroscience*, vol. 14, no. 11, p. 1481, 2011.

[50] E. A. Susaki, K. Tainaka, D. Perrin, F. Kishino, T. Tawara, T. M. Watanabe, C. Yokoyama, H. Onoe, M. Eguchi, S. Yamaguchi, *et al.*, "Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis," *Cell*, vol. 157, no. 3, pp. 726–739, 2014.

[51] A. Ertürk, K. Becker, N. Jährling, C. P. Mauch, C. D. Hojer, J. G. Egen, F. Hellal, F. Bradke, M. Sheng, and H.-U. Dodt, "Three-dimensional imaging of solvent-cleared organs using 3disco," *Nature protocols*, vol. 7, no. 11, p. 1983, 2012.

[52] N. Renier, Z. Wu, D. J. Simon, J. Yang, P. Ariel, and M. Tessier-Lavigne, "idisco: a simple, rapid method to immunolabel large tissue samples for volume imaging," *Cell*, vol. 159, no. 4, pp. 896–910, 2014.

[53] H. Gong, S. Zeng, C. Yan, X. Lv, Z. Yang, T. Xu, Z. Feng, W. Ding, X. Qi, A. Li, *et al.*, "Continuously tracing brain-wide long-distance axonal projections in mice at a one-micron voxel resolution," *Neuroimage*, vol. 74, pp. 87–98, 2013.

[54] Y. Choe, L. C. Abbott, D. Han, P.-S. Huang, J. Keyser, J. Kwon, D. Mayerich, Z. Melek, and B. H. McCormick, "Knife-edge scanning microscopy: high-throughput imaging and analysis of massive volumes of biological microstructures," *High-Throughput Image Reconstruction and Analysis: Intelligent Microscopy Applications*, pp. 11–37, 2008.

[55] A. P. Di Giovanna, A. Tibo, L. Silvestri, M. C. Müllenbroich, I. Costantini, A. L. A. Mascaro, L. Sacconi, P. Frasconi, and F. S. Pavone, "Whole-brain vasculature reconstruction at the single capillary level," *Scientific reports*, vol. 8, no. 1, p. 12573, 2018.

[56] N. Yayon, A. Dudai, N. Vrieler, O. Amsalem, M. London, and H. Soreq, "Intensify3D: Normalizing signal intensity in large heterogenic image stacks.," *Scientific reports*, vol. 8, p. 4311, Mar. 2018.

[57] J. Lee, W. An, and Y. Choe, "Mapping the full vascular network in the mouse brain at submicrometer resolution.," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2017, pp. 3309–3312, July 2017.

[58] B. Xiong, A. Li, Y. Lou, S. Chen, B. Long, J. Peng, Z. Yang, T. Xu, X. Yang, X. Li, T. Jiang, Q. Luo, and H. Gong, "Precise Cerebral Vascular Atlas in Stereotaxic Coordinates of Whole Mouse Brain.," *Frontiers in neuroanatomy*, vol. 11, p. 128, 2017.

[59] S. Lim, M. R. Nowak, and Y. Choe, "Automated neurovascular tracing and analysis of the knife-edge scanning microscope Rat Nissl data set using a computing cluster," *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6445–6448, May 2016.

[60] R. C. Gonzalez, R. E. Woods, *et al.*, "Digital image processing," 2002.

[61] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[62] M. Piccardi, "Background subtraction techniques: a review," in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, vol. 4, pp. 3099–3104, IEEE, 2004.

[63] V. N. P. Vemuri, *Feature-based Analysis of Microvasculature in High Resolution Microscopy Images of Mice Brains*. Masters thesis, University of Houston, May 2016.

[64] D. Mayerich and J. Keyser, "Hardware accelerated segmentation of complex volumetric filament networks.," *IEEE transactions on visualization and computer graphics*, vol. 15, pp. 670–681, July 2009.

[65] C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," *ACM Computing Surveys (CSUR)*, vol. 36, no. 2, pp. 81–121, 2004.

[66] A. Sarwal and A. Dhawan, "3-d reconstruction of coronary arteries," in *Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 504–505, IEEE, 1994.

[67] A. Dima, M. Scholz, and K. Obermayer, "Automatic segmentation and skeletonization of neurons from confocal microscopy images based on the 3-d wavelet transform," *IEEE Transactions on image processing*, vol. 11, no. 7, pp. 790–801, 2002.

[68] N. Niki, Y. Kawata, H. Satoh, and T. Kumazaki, "3d imaging of blood vessels using x-ray rotational angiographic system," in *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, pp. 1873–1877, IEEE, 1993.

[69] T. Tozaki, Y. Kawata, N. Niki, H. Ohmatsu, K. Eguchi, and N. Moriyama, "Three-dimensional analysis of lung areas using thin-slice ct images," in *Proceedings of 13th International Conference on Pattern Recognition*, vol. 3, pp. 548–552, IEEE, 1996.

[70] X. He, E. Kischell, M. Rioult, and T. J. Holmes, "Three-dimensional thinning algorithm that peels the outmost layer with application to neuron tracing," *Journal of Computer-Assisted Microscopy*, vol. 10, no. 3, pp. 123–135, 1998.

[71] H. Schmitt, M. Grass, V. Rasche, O. Schramm, S. Haehnel, and K. Sartor, "An x-ray-based method for the determination of the contrast agent propagation in 3-d vessel structures," *IEEE Transactions on medical imaging*, vol. 21, no. 3, pp. 251–262, 2002.

[72] J. F. O'Brien and N. F. Ezquerra, "Automated segmentation of coronary vessels in angiographic image sequences utilizing temporal, spatial, and structural constraints," in *Visualization in Biomedical Computing 1994*, vol. 2359, pp. 25–38, International Society for Optics and Photonics, 1994.

[73] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis, "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," *Medical image analysis*, vol. 2, no. 2, pp. 143–168, 1998.

[74] K. A. Al-Kofahi, S. Lasek, D. H. Szarowski, C. J. Pace, G. Nagy, J. N. Turner, and B. Roysam, "Rapid automated three-dimensional tracing of neurons from confocal image stacks," *IEEE Transactions on information technology in biomedicine*, vol. 6, no. 2, pp. 171–187, 2002.

[75] D. Mayerich and J. Keyser, "Filament tracking and encoding for complex biological networks.," *Symposium on Solid and Physical Modeling*, 2008.

[76] J. A. Tyrrell, E. di Tomaso, *et al.*, "Robust 3-d modeling of vasculature imagery using superellipsoids.," *IEEE transactions on medical imaging*, vol. 26, pp. 223–237, Feb. 2007.

[77] P. Teikari, M. Santos, C. Poon, and K. Hynynen, "Deep learning convolutional networks for multiphoton microscopy vasculature segmentation," *arXiv preprint arXiv:1606.02382*, 2016.

[78] M. Haft-Javaherian, L. Fang, V. Muse, C. B. Schaffer, N. Nishimuraa, and M. R. Sabuncu, "Deep convolutional neural networks for segmenting 3d in vivo multiphoton images of vasculature in alzheimer disease mouse models," *arXiv preprint arXiv:1801.00880*, 2018.

[79] J. Wu, Y. He, Z. Yang, C. Guo, Q. Luo, W. Zhou, S. Chen, A. Li, B. Xiong, T. Jiang, and H. Gong, "3D BrainCV: Simultaneous visualization and analysis of cells and capillaries in a whole mouse brain with one-micron voxel resolution," *NeuroImage*, vol. 87, pp. 199–208, 2014.

[80] S. Almasi, X. Xu, A. Ben-Zvi, B. Lacoste, C. Gu, and E. L. Miller, "A novel method for identifying a graph-based representation of 3-D microvascular networks from fluorescence microscopy image stacks.," *Medical image analysis*, vol. 20, pp. 208–223, Feb. 2015.

[81] R. Bise, I. Sato, K. Kajiya, and T. Yamashita, "3D Structure Modeling of Dense Capillaries by Multi-objects Tracking," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1333–1341, IEEE, 2016.

[82] J. A. Tyrrell, E. di Tomaso, D. Fuja, R. Tong, K. Kozak, R. K. Jain, and B. Roysam, "Robust 3-D modeling of vasculature imagery using superellipsoids.," *IEEE transactions on medical imaging*, vol. 26, pp. 223–237, Feb. 2007.

[83] H.-C. Han, "Twisted blood vessels: symptoms, etiology and biomechanical mechanisms," *Journal of vascular research*, vol. 49, no. 3, pp. 185–197, 2012.

[84] S. Lorthois, F. Lauwers, and F. Cassot, "Tortuosity and other vessel attributes for arterioles and venules of the human cerebral cortex," *Microvascular research*, vol. 91, pp. 99–109, 2014.

[85] H. J. Kagan, V. D. Belekdanian, J. Chen, P. Backeris, N. Hammoudi, I. C. Turnbull, K. D. Costa, and R. J. Hajjar, "Coronary capillary blood flow in a rat model of congestive heart failure," *Journal of Applied Physiology*, pp. jap–00741, 2017.

[86] W. R. Brown, D. M. Moody, C. R. Thore, J. A. Anstrom, and V. R. Challa, "Microvascular changes in the white mater in dementia," *Journal of the neurological sciences*, vol. 283, no. 1, pp. 28–31, 2009.

[87] E. Bullitt, K. E. Muller, I. Jung, W. Lin, and S. Aylward, "Analyzing attributes of vessel populations," *Medical image analysis*, vol. 9, no. 1, pp. 39–49, 2005.

[88] E. Bullitt, G. Gerig, S. M. Pizer, W. Lin, and S. R. Aylward, "Measuring tortuosity of the intracerebral vasculature from mra images," *IEEE transactions on medical imaging*, vol. 22, no. 9, pp. 1163–1171, 2003.

[89] E. Bullitt, G. Gerig, S. Aylward, S. Joshi, K. Smith, M. Ewend, and W. Lin, "Vascular attributes and malignant brain tumors," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 671–679, Springer, 2003.

[90] E. Bullitt, I. Jung, K. Muller, G. Gerig, S. Aylward, S. Joshi, K. Smith, W. Lin, and M. Ewend, "Determining malignancy of brain tumors by analysis of vessel shape," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 645–653, Springer, 2004.

[91] S. Hirsch, J. Reichold, M. Schneider, G. Székely, and B. Weber, "Topology and hemodynamics of the cortical cerebrovascular system," *Journal of Cerebral Blood Flow & Metabolism*, vol. 32, no. 6, pp. 952–967, 2012.

[92] P. S. Tsai, J. P. Kaufhold, P. Blinder, B. Friedman, P. J. Drew, H. J. Karten, P. D. Lyden, and D. Kleinfeld, "Correlations of Neuronal and Microvascular Densities in Murine Cortex Revealed by Direct Counting and Colocalization of Nuclei and Vessels," *Journal of Neuroscience*, vol. 29, pp. 14553–14570, Nov. 2009.

[93] I. Robinson, J. Webber, and E. Eifrem, *Graph databases*. " O'Reilly Media, Inc.", 2013.

[94] J. R. Chung, C. Sung, D. Mayerich, J. Kwon, D. E. Miller, T. Huffman, J. Keyser, L. C. Abbott, and Y. Choe, "Multiscale Exploration of Mouse Brain Microstructures Using the Knife-Edge Scanning Microscope Brain Atlas.," *Front. Neuroinform.*, vol. 5, p. 29, 2011.

[95] D. C.-Y. Eng and Y. Choe, "Stereo pseudo 3d rendering for web-based display of scientific volumetric data," in *Proceedings of the Fifth Eurographics/IEEE VGTC conference on Point-Based Graphics*, pp. 73–80, Eurographics Association, 2008.

[96] M. R. Nowak and Y. Choe, "Towards an open-source framework for the analysis of cerebrovasculature structure," in *Engineering in Medicine and Biology Society (EMBC), 2018 IEEE 40th Annual International Conference of the*, IEEE, 2018.

[97] L. Kuan, Y. Li, C. Lau, D. Feng, A. Bernard, S. M. Sunkin, H. Zeng, C. Dang, M. Hawrylycz, and L. Ng, "Neuroinformatics of the allen mouse brain connectivity atlas," *Methods*, vol. 73, pp. 4–17, 2015.

[98] S. M. Sunkin, L. Ng, C. Lau, T. Dolbeare, T. L. Gilbert, C. L. Thompson, M. Hawrylycz, and C. Dang, "Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system.," *Nucleic acids research*, vol. 41, pp. D996–D1008, Jan. 2013.

[99] M. R. Nowak, J. Lee, and Y. Choe, "A queryable graph representation of vascular connectivity in the whole mouse brain," 2019.

[100] A. Hodler, "Optimized graph algorithms in neo4j," tech. rep., Neo4j, Inc., San Mateo, CA, 2018.

[101] M. R. Nowak and Y. Choe, "Data-driven synthetic cerebrovascular models for validation of segmentation algorithms," in *Engineering in Medicine and Biology Society (EMBC), 2018 IEEE 40th Annual International Conference of the*, IEEE, 2018.

[102] E. Stockley, H. Cole, A. Brown, and H. Wheal, "A system for quantitative morphological measurement and electrotonic modelling of neurons: three-dimensional reconstruction," *Journal of neuroscience methods*, vol. 47, no. 1-2, pp. 39–51, 1993.

[103] W. J. Schroeder, B. Lorensen, and K. Martin, *The visualization toolkit: an object-oriented approach to 3D graphics*. Kitware, 2004.

[104] P. Min, "binvox." `http://www.patrickmin.com/binvox` or `https://www.google.com/search?q=binvox`, 2004 - 2019.

[105] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," in *ACM siggraph computer graphics*, vol. 21, pp. 163–169, ACM, 1987.

[106] M. R. Nowak and Y. Choe, "Towards a more complete validation framework for cerebrovascular segmentation algorithms," Manuscript forthcoming.

[107] L. M. Lorigo, O. D. Faugeras, W. E. L. Grimson, R. Keriven, R. Kikinis, A. Nabavi, and C.-F. Westin, "Curves: Curve evolution for vessel segmentation," *Medical image analysis*, vol. 5, no. 3, pp. 195–206, 2001.

[108] A. Hagberg, P. Swart, and D. S Chult, "Exploring network structure, dynamics, and function using networkx," tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[109] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[110] Daisuke Miyamoto, "swc2vtk." `https://github.com/DaisukeMiyamoto/swc2vtk`.

[111] Kitware Inc., "Paraview." `https://www.paraview.org/`.

[112] E. Lugo-Hernandez, A. Squire, N. Hagemann, A. Brenzel, M. Sardari, J. Schlechter, E. H. Sanchez-Mendoza, M. Gunzer, A. Faissner, and D. M. Hermann, "3D visualization and quantification of microvessels in the whole ischemic mouse brain using solvent-based clearing and light sheet microscopy," *Journal of Cerebral Blood Flow & Metabolism*, vol. 37, no. 10, pp. 3355–3367, 2017.

[113] B. Qian, R. F. Rudy, T. Cai, and R. Du, "Cerebral artery diameter in inbred mice varies as a function of strain," *Frontiers in neuroanatomy*, vol. 12, p. 10, 2018.

[114] G. S. Kassab, D. H. Lin, and Y. C. Fung, "Morphometry of pig coronary venous system.," *The American journal of physiology*, vol. 267, pp. H2100–13, Dec. 1994.

[115] G. S. Kassab, C. A. Rider, N. J. Tang, and Y. C. Fung, "Morphometry of pig coronary arterial trees.," *The American journal of physiology*, vol. 265, pp. H350–65, July 1993.