# Deep Learning-Based Analysis of Mouse Retinal Vasculature Development: A Reproducible Pipeline for Vessel Segmentation and Morphological Feature Extraction

Ishita Pethkar
University of North Carolina, Chapel Hill
ishpet@unc.edu

Yashas Majmudar
University of North Carolina, Chapel Hill
yashashm@unc.edu

November 2025

## Abstract

The postnatal development of the mouse retinal vasculature in C57 mice (P1-P7) serves as a critical and highly reproducible model for studying physiological angiogenesis. Quantifying the morphology of this complex network is essential to understand vascular disease and the efficacy of preclinical drugs. This paper introduces a robust and fully automated computational pipeline for the extraction and comprehensive quantitative analysis of vascular features from immunofluorescently stained (IB4) retinal images. The pipeline utilizes a mixed-precision mixed-precision U-Net neural network (CNN) for high-fidelity pixel-level vessel segmentation (Dice Score: 0.936). The resulting binary masks are then subjected to graph-based network analysis using skeletonization to accurately calculate key metrics, including total vessel length, density, branch point density, and tortuosity. Designed with stringent version control and dependency pinning, this solution adheres to the platinum standard for computational reproducibility. The methodology provides a robust and high-throughput platform for statistically analyzing angiogenic processes and generating meaningful biological insights throughout the developmental stages.

**GitHub Repository: https://github.com/yashas-hm/AngioNet**

## 1 Introduction

Small blood vessels are essential for maintaining vascular health and are intimately involved in the progression of major diseases such as myocardial infarction, stroke, and dementia. Although genetic and environmental perturbations are known to influence microvascular function, the biological mechanisms underlying these changes are incompletely understood. Because vascular morphology often reflects underlying physiological and pathological processes, quantitative analysis of vascular architecture has become an essential approach for studying vascular development and disease.

The mouse retina provides a particularly powerful and widely used model for investigating angiogenesis and microvascular remodeling. Retinal vascular development occurs entirely postnatally from birth (P0) through approximately postnatal day 25 (P25) and proceeds through two major phases: formation of the primary superficial plexus and subsequent development of the deeper vascular layers. Between P0 and P7, endothelial cells sprout from vessels emerging at the optic nerve head and extend radially across the retina, guided by a pre-existing astrocytic scaffold. After P7, vessels begin to invade deeper retinal layers to form the deep plexus.

Two key biological processes drive these early morphological changes: angiogenesis, the sprouting of new vessels from existing ones, and microvascular remodeling, which encompasses vessel pruning, changes in vessel diameter and area, and artery–vein specification without new vessel formation. These processes not only shape normal organ development but also contribute to pathological neovascularization in diseases and tumor progression. Understanding how vascular structure evolves during these stages is therefore critical for identifying signatures of healthy versus diseased microvascular growth.

Despite its importance, automated extraction and quantitative analysis of retinal vascular morphology remains an active area of research. Several software tools have been developed such as Angioquant, AngioTool, RAVE, and Autotube but most rely heavily on threshold-based segmentation strategies. With rapid progress in computer vision, particularly convolutional neural networks (CNNs) and other deep learning approaches, more accurate and robust segmentation methods have become feasible. Image segmentation can be framed as a pixel-wise classification problem, and architectures such as U-Net have proven particularly effective for biomedical segmentation tasks due to their ability to capture both local and global context.

In this project, we analyze confocal microscopy images of mouse retinas collected from C57 mice aged P1 to P7. Retinas were isolated, immunofluorescently stained using biotinylated isolectin B4 and streptavidin-conjugated fluorophores, and imaged as wholemounts using laser scanning confocal microscopy. These images provide a high-resolution view of early retinal vascular development, allowing computational extraction of vessel structure and morphology.

The primary goal of this project is to establish a fully reproducible analysis pipeline for the segmentation and quantitative analysis of mouse retinal vasculature. This includes guaranteeing that every step from raw image preprocessing, patch extraction, and deep learning-based vessel segmentation, to graph-based feature extraction and statistical analysis can be executed by independent researchers with identical results. Achieving full reproducibility involves controlling all sources of variability, including software dependencies, random initialization in model training, data preprocessing parameters, and analysis scripts, while providing comprehensive documentation and version-controlled code. By ensuring reproducibility, the pipeline enables transparent, verifiable, and extensible research that can be confidently applied to future studies in vascular biology and related biomedical fields.

Objectives of this work are fourfold:

- Develop a vessel segmentation algorithm.
- Extract retinal vascular networks from the segmented images.
- Quantify morphological and topological features, such as vessel density, branching complexity, diameter distribution, and network connectivity.
- Perform exploratory data analysis and statistical interpretation of vascular features across developmental stages (P2–P7), relating observed patterns to established biological knowledge of angiogenesis and microvascular remodeling.

Through this pipeline, we aim to characterize early postnatal vascular development in the mouse retina and provide biologically meaningful interpretations that align with previous findings in angiogenesis research. This framework also lays the groundwork for future quantitative studies of vascular abnormalities in disease models.

# 2 Methodology

The analysis of mouse retinal vasculature was conducted using a multi-stage computational pipeline designed to integrate high-resolution imaging data with deep learning-based segmentation and graph-theoretical feature extraction. The methodology combines rigorous data preprocessing, advanced convolutional neural network architectures, and network analysis techniques to quantitatively characterize vascular morphology

and topology. Each stage of the pipeline, including raw image acquisition, normalization, patch extraction, U-Net segmentation, skeletonization, and feature computation, was developed with reproducibility and biological interpretability as guiding principles. By structuring the workflow in modular and well-documented stages, we ensured that the pipeline could be executed consistently across datasets and by independent researchers. This methodology allows for detailed assessment of developmental trends in retinal vasculature, capturing both microstructural features such as vessel width and tortuosity, as well as macrostructural network properties including branching complexity and radial expansion.

## 2.1 Data Acquisition

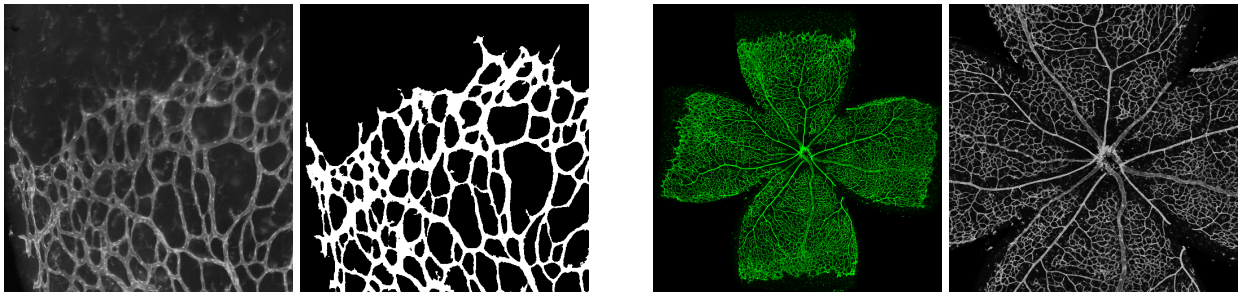
The retinal images analyzed in this study were obtained from C57 mice at postnatal days P1 through P7. Following euthanasia, the eyes were enucleated and fixed in 4% paraformaldehyde (PFA) for 100 minutes on ice. Retinas were carefully isolated under a dissecting microscope and prepared for whole-mount immunofluorescent staining using biotinylated isolectin B4 (IB4) in combination with Alexa Fluor-conjugated streptavidin. Confocal imaging was performed on an Olympus FluoView system at 10X magnification to capture high-resolution Z-stack images of the developing retinal vasculature. These images provided detailed structural information on vascular networks across early postnatal stages, enabling precise computational analysis of both microvascular morphology and network topology.

The dataset was organized into training, validation, and test sets. The training and validation sets each consisted of a single high-resolution retinal image paired with a ground truth segmentation mask, while the test set comprised multiple retinal images spanning P2 to P7. No ground truth was provided for test images, allowing for unbiased assessment of the trained segmentation model.

## 2.2 Data Preprocessing Pipeline

Raw confocal images underwent a preprocessing pipeline to prepare them for deep learning-based vessel segmentation. Initially, images were converted from RGB to grayscale and normalized to the range [0,1] to ensure consistent input intensity scaling. Large images were then divided into overlapping 512×512-pixel patches with a 50% overlap. This patch-based approach allowed the U-Net architecture to process high-resolution images within GPU memory constraints while preserving local vessel continuity.

To prevent the network from learning uninformative background regions, patches with less than 5% vessel pixels were excluded from training. Additionally, training masks were automatically corrected for inverted foreground/background conventions relative to validation masks, ensuring consistency across datasets. This preprocessing framework produced standardized, high-quality inputs that facilitated robust model training and evaluation.



(a) Train Image and Train Mask

(b) Raw Test Image and Pre-processed Test Image

Figure 1: Pre-processed Data Used for Model Training and Evaluation

## 2.3 U-Net Architecture

For vessel segmentation, a U-Net convolutional neural network was implemented. The architecture consists of an encoder-decoder structure with skip connections that preserve fine-grained spatial information lost during downsampling. The encoder performs successive downsampling operations using max pooling and double convolutional layers, progressively capturing contextual information. A dropout layer with a probability of 0.5 was incorporated at the bottleneck to reduce overfitting.

The decoder reconstructs the spatial resolution using bilinear upsampling followed by concatenation with corresponding encoder features. Double convolutional layers refine the combined feature maps, ultimately producing a single-channel probability map representing vessel likelihood. This design allows the network to capture both local details and global vascular patterns, crucial for accurately segmenting vessels of varying widths and branching complexity.

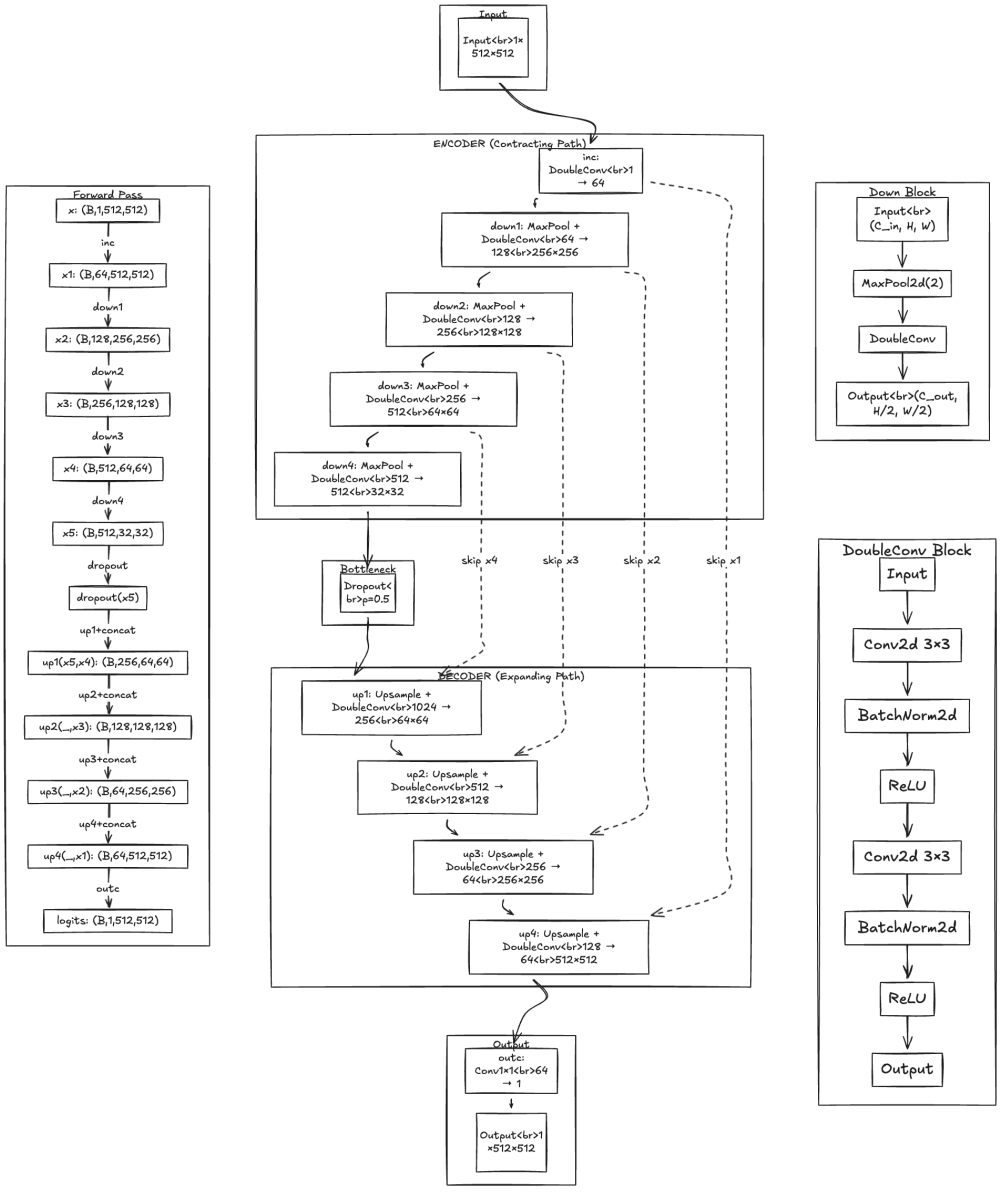



Figure 2: U-Net Architecture Overview

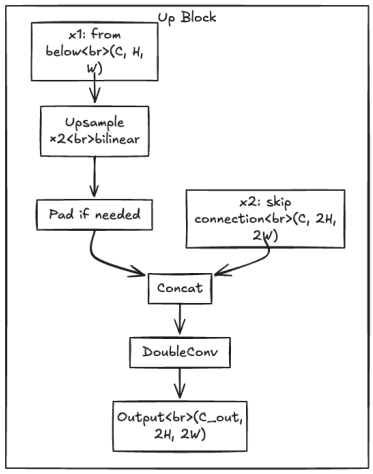



Figure 3: Detailed Representation of the U-Net Encoder-Decoder Path

## 2.4 Training Configuration

The model was trained using the Adam optimizer with a learning rate of $1 \times 10^{-5}$ and a batch size of eight patches. Binary cross-entropy with logits loss was employed, combining sigmoid activation with cross-entropy for numerical stability. Training was performed for seven epochs using mixed precision (FP16) to improve GPU memory efficiency.

Model performance was evaluated using the Dice coefficient and pixel-level accuracy. The Dice score, which quantifies overlap between predicted and ground truth masks, reached 0.936 on the validation set, while pixel accuracy achieved 91.85%. These metrics demonstrate the network's capacity to generalize to unseen retinal images and provide reliable segmentation for downstream feature extraction.

## 2.5 Prediction

The trained U-Net model was applied to the test images using the same preprocessing and patch extraction strategy. Each patch was processed independently, followed by recombination to reconstruct the full-size segmentation mask. Probability maps were thresholded at 0.5 to generate binary vessel masks, which were then saved for feature extraction. This pipeline ensures consistent predictions across images and preserves vascular topology for graph-based analysis.

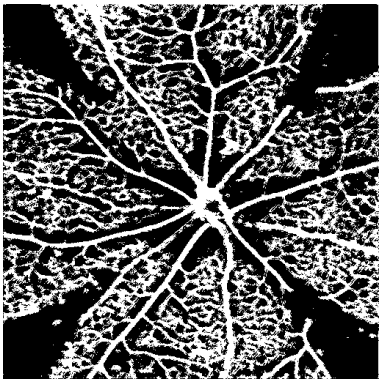


Figure 4: Predicted Image for Test Image (fig. 1b)

## 2.6 Vessel Network Feature Extraction

Binary segmentation masks were processed to extract quantitative vascular features. Skeletonization using morphological thinning reduced vessels to one-pixel-wide centerlines while maintaining topological integrity. Skeletons were converted into graph representations using NetworkX, with nodes representing branch points and endpoints, and edges representing vessel segments.

Graph refinement steps removed spurious edges and small disconnected components, ensuring clean network structures. For each vessel segment, features including Euclidean length, path length, width, width variance, tortuosity, and pixel coordinates were computed. Node-level features, such as coordinates, distance from the image center, and node degree, were also extracted. Node degree provides biologically meaningful information: endpoints indicate sprouting tips, degree-3 nodes represent bifurcations, and higher-degree nodes correspond to complex junctions. These extracted features form the basis for quantitative analysis of vascular morphology and network topology.

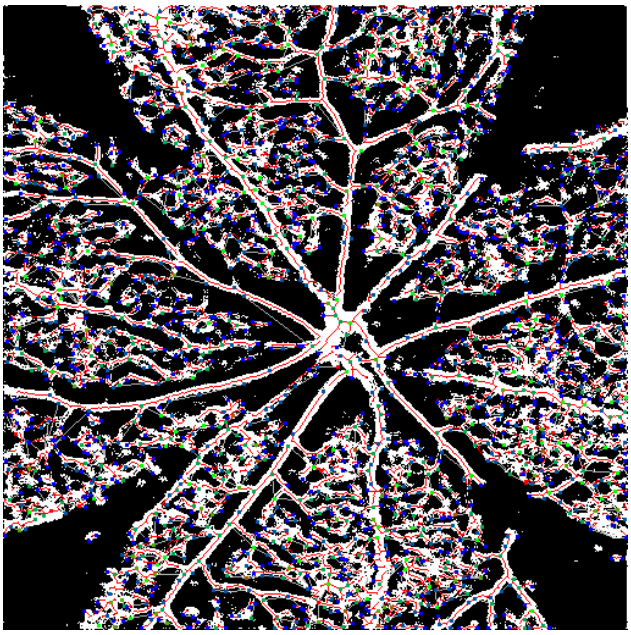


Figure 5: Features Extracted for Test Image (fig. 1b)

## 2.7 Exploratory Data Analysis (EDA)

Following segmentation and feature extraction, an extensive exploratory data analysis (EDA) workflow was conducted to characterize the morphological and topological properties of the mouse retinal vasculature. The goal of this stage was to quantify how vessel structure evolves during early postnatal development, with a focus on changes occurring between postnatal days P2 and P7. This developmental window is defined by two major biological processes. The first is angiogenesis, where new vessels emerge through endothelial sprouting from existing vasculature. The second is vascular remodeling, which involves the refinement of the network through pruning, diameter regulation, and early artery–vein specification. Because these processes together shape both microstructural and macrostructural vessel properties, the EDA framework was designed to analyze features at the level of individual vessel segments as well as at the level of global network structure.

The EDA operates on the outputs of the feature extraction pipeline, which provides three primary files per image patch: a vessel segment feature table, a node degree table, and a network visualization. The segment table contains detailed geometrical measurements, including Euclidean length, curved path length, vessel width, and tortuosity. These features describe the physical morphology of the vasculature and

enable comparisons of vessel maturation, stability, and curvature. The node degree table captures graph-derived properties such as the number of connections at each node and the distance of each node from the retinal center. Node degrees provide insight into branching complexity and sprouting behavior. Endpoints correspond to active growth fronts, while bifurcation nodes indicate branching events. Together, these tables enable a comprehensive characterization of both localized and global vascular traits.

To facilitate analysis, the pipeline includes a set of modular functions that support both individual patch-level evaluation and comparisons across developmental stages. It begins by visualizing the extracted vessel network, overlays the skeleton on the original image, and computes descriptive statistics for all vessel and node features. Morphological distributions are plotted for core metrics such as vessel length, diameter, and tortuosity. These distributions help identify heterogeneity within a patch and serve as internal quality controls. For example, extremely long or excessively wide segments may indicate segmentation artifacts, while an abnormally skewed tortuosity distribution could suggest localized abnormalities in vascular patterning.

Branching behavior is examined using node degree analysis, which produces bar charts and pie charts describing the proportions of endpoints, continuation points, bifurcations, and higher-order junctions. Because angiogenesis is characterized by frequent sprouting, early developmental stages typically show a high proportion of degree-1 nodes. As the network matures and remodeling progresses, continuation points and bifurcations increase, while the proportion of endpoints stabilizes. These trends offer a quantitative view of how vascular complexity evolves over time.

Feature correlations are computed using both correlation matrices and pairwise scatter plots. Morphological features often co-vary: for example, longer vessels may also exhibit greater width due to arterial expansion, and tortuosity may negatively correlate with vessel length in more mature networks. Understanding these relationships provides insight into coordinated developmental behavior and supports the identification of distinct vessel classes or growth patterns. Outlier detection based on the interquartile range is then applied to highlight vessel segments that deviate substantially from expected morphology. Outliers may represent rare biological events such as excessively curved angiogenic sprouts or technical artifacts such as segmentation irregularities. All outlier candidates are plotted for manual inspection.

Beyond patch-level analysis, the EDA includes a developmental comparison function, which aggregates features across all images corresponding to postnatal days P2 through P7. This stage of analysis is critical for identifying global developmental trajectories rather than local variations within individual patches. Summary statistics for vessel length, width, tortuosity, and branching properties are computed for each postnatal day. These statistics are visualized using trend lines, box plots, and violin plots that capture both mean-level changes and distributional shifts. Across normal retinal development, vessel segments are expected to elongate as the network expands, while tortuosity generally decreases as immature sprouts straighten during remodeling. Vessel width undergoes more variable changes reflecting ongoing artery–vein differentiation.

Branching patterns show similarly informative trends. The bifurcation ratio, defined as the proportion of degree-3 nodes relative to total nodes, increases with network maturation as vessel branches become more hierarchically organized. Conversely, the endpoint ratio tends to decrease over time because the number of active sprouting fronts diminishes once the primary vascular plexus is established. The radial expansion of the vascular network is quantified by measuring each node's distance from the retinal center; these distances increase progressively from P2 to P7 and reflect outward growth from the optic nerve head.

The EDA results collectively provide a biologically interpretable picture of retinal angiogenesis and remodeling. Lengthening of vessels corresponds to network expansion, decreasing tortuosity reflects structural stabilization, increasing bifurcation ratios indicate growing topological complexity, and widening distances from the center confirm radial vascular spread. These findings align with established developmental biology literature and validate both the segmentation and feature extraction processes. In addition to supporting biological interpretation, the EDA framework establishes a foundation for downstream statistical analysis, phenotype comparison, or disease modeling.

# 3 Reproducibility Measures

Ensuring reproducibility is essential for reliable vessel segmentation and feature extraction. All pipeline stages, environment setup, and analysis have been designed to facilitate exact replication of results.

## 3.1 Environment Management

### 3.1.1 Virtual Environment Isolation

All dependencies are installed in a dedicated Python virtual environment to avoid conflicts with system packages. The main pipeline script `run_pipeline.py` automatically creates and manages this environment.

### 3.1.2 Pinned Dependencies

All Python packages are pinned to specific versions to guarantee consistent behavior:

| Package | Version |
|---|---|
| torch | 2.9.0 |
| torchvision | 0.24.0 |
| numpy | 2.2.6 |
| scikit-image | 0.25.2 |
| opencv-python | 4.12.0.88 |
| networkx | 3.5 |
| pandas | 2.3.3 |
| matplotlib | 3.10.7 |
| xlsxwriter | 3.2.9 |
| seaborn | 0.13.2 |
| openpyxl | 3.1.2 |

Table 1: Pinned Python dependencies for reproducibility.

### 3.1.3 Environment Setup Module

A dedicated module `setup_environment.py` handles:

- Python version validation (3.10+)
- Virtual environment creation if it does not exist
- Dependency installation from `requirements.txt`
- Automatic relaunch of scripts within the virtual environment

```
from setup_environment import setup_environment, relaunch_in_venv

python_path, in_venv = setup_environment()
if not in_venv:
    relaunch_in_venv(python_path)
```

## 3.2 Version Control

All code is version controlled using Git with the following best practices:

- Meaningful commit messages
- Logical, single-unit commits
- Feature branches merged into `main` after verification

Large or generated files are excluded via `.gitignore`:

`venv/`, `__pycache__/`, `.ipynb_checkpoints/`

## 3.3 Pipeline Automation

### 3.3.1 Automation Scripts

The pipeline is modular and includes the following scripts under **Runners** directory:

| Script | Purpose |
|---|---|
| `run_pipeline.py` | Main pipeline runner; executes all stages and sets up environment |
| `oneclick.py` | One-click reproduction; downloads outputs, executes pipeline and EDA notebook |
| `setup_environment.py` | Handles virtual environment setup and dependency installation |

Table 2: Pipeline automation scripts.

### 3.3.2 Idempotent Execution

Before running each stage, the pipeline checks for existing outputs to prevent redundant computation:

```
if not os.path.exists('Data/preprocessed/train/image'):
    process_data(project_root)
if not os.path.exists('Model/unet_checkpoint.pth.tar'):
    train_model(project_root)
if not os.path.exists('Data/outputs/predictions'):
    run_prediction(project_root)
run_feature_extraction(project_root)
```

### 3.3.3 Standalone Execution

Each stage can be run independently:

```
python Scripts/data_preprocessing.py
python Scripts/train.py
python Scripts/predict.py
python Scripts/feature_extraction/feature_extraction.py
```

## 3.4 Code Organization

### 3.4.1 Modular Architecture

The codebase is structured as follows:

```
Scripts/
|---- constants.py
|---- data_preprocessing.py
|---- train.py
|---- predict.py
|---- unet_model.py
|---- dataset/
|   |---- vessel_dataset.py
|   |---- test_vessel_dataset.py
|---- feature_extraction/
    |---- feature_extraction.py
    |---- connectivity_matrix_test.py
    |---- utility.py
```

### 3.4.2 Centralized Configuration

All hyperparameters are stored in `constants.py` to avoid hardcoding:

```
PATCH_SIZE = 512
STRIDE = 256
LEARNING_RATE = 1e-5
BATCH_SIZE = 8
NUM_EPOCHS = 7
NUM_WORKERS = 2
IMAGE_HEIGHT = 512
IMAGE_WIDTH = 512
```

## 3.5 Documentation and EDA

Comprehensive documentation ensures reproducibility:

- **README.md:** Project overview, dependencies, pipeline description
- **EDA.md:** Function usage, biological interpretation, and examples
- **Reproducibility.md:** Detailed reproducibility measures and guide
- **Results.md:** Results and Key Findings
- **Docstrings:** All functions are documented with input/output descriptions

EDA is fully reproducible via `eda_angio_net.ipynb` with reusable functions such as `run_full_eda()` and `run_developmental_analysis()`.

### 3.6 Pre-computed Outputs

Pre-computed outputs are provided in a downloadable archive for users without GPU access:

- Trained model: `Model/unet_checkpoint.pth.tar` (Dice score 0.936)
- Predicted masks: `Data/outputs/predictions/`
- Extracted features: `Data/outputs/features/` including segment and node features and network visualizations

### 3.7 Data Organization

**Raw Data:** Stored separately in `Data/raw_data/` and never modified.
**Generated Outputs:** Stored in `Data/preprocessed/` and `Data/outputs/`.

### 3.8 Summary of Reproducibility Features

| **Feature** | **Implementation** |
|---|---|
| Environment Isolation | Python venv with pinned dependencies |
| Version Control | Git with meaningful commits |
| Single Entry Point | run_pipeline.py |
| One-Click Reproduction | oneclick.py (downloads outputs, runs pipeline, executes EDA) |
| Idempotent Execution | Checkpoint detection before each stage |
| Modular Code | Standalone, parameterized scripts |
| Centralized Configuration | constants.py |
| Documentation | README.md, EDA.md, Reproducibility.md, docstrings |
| Pre-computed Outputs | Model, predictions, features downloadable |
| Reproducible EDA | Jupyter notebook with reusable functions |
| Data Organization | Immutable raw data, separate generated outputs |
| Platform Independence | Compatible with Windows, macOS, Linux |

Table 3: Summary of reproducibility measures.

## 4 Results and Findings

This section presents the results obtained from the vessel segmentation pipeline and the exploratory data analysis (EDA) of mouse retinal vasculature development from postnatal day P2 to P7.

### 4.1 Model Performance

The U-Net model achieved high segmentation performance on the validation set. The Dice score of 0.936 indicates strong overlap between predicted vessel masks and ground truth. Pixel-level accuracy was measured at 91.85%. The model was trained for 7 epochs, taking approximately 2-5 minutes per epoch on a GPU. These results demonstrate that the network effectively captures vascular structures across varying vessel widths and branching patterns.

| Metric | Value |
|---|---|
| Dice Score | 0.936 |
| Pixel Accuracy | 91.85% |
| Training Epochs | 7 |
| Training Time | 2-5 minutes per epoch |

Table 4: U-Net model performance metrics.

## 4.2 Exploratory Data Analysis Results

We perform EDA for each patch and try to extract various trand and sttistical properties. For "p7" patch we get the followinfg statstics:

### 4.2.1 Vessel Segment Statistics

This table provides descriptive statistics for the vessel segments (the individual vessel pieces between two junctions or endpoints). It summarizes key morphological features like:

- line: The straight-line distance from the start to the end of a segment.
- length: The actual path length of the segment.
- width: The average thickness of the segment.
- width_var: How much the vessel's width varies along its length.
- tortuosity: A measure of how much the vessel deviates from a straight path (higher values mean more twists and turns).

| | line | length | width | width_var | tortuosity |
|---|---|---|---|---|---|
| **mean** | 11.791889 | 8.758900 | 3.359420 | 0.749206 | 0.826143 |
| **std** | 6.189816 | 5.014702 | 1.844694 | 1.301031 | 0.379954 |
| **min** | 5.099020 | 3.000000 | 2.000000 | 0.000000 | 0.121218 |
| **25%** | 7.615773 | 5.656856 | 2.000000 | 0.000000 | 0.539835 |
| **50%** | 10.198039 | 7.828428 | 2.000000 | 0.443750 | 0.777232 |
| **75%** | 14.115519 | 10.414214 | 4.000000 | 0.960000 | 1.077310 |
| **max** | 96.400207 | 76.911704 | 16.000000 | 20.690000 | 2.922032 |

Table 5: Vessel Segment Statistics for p7 patch

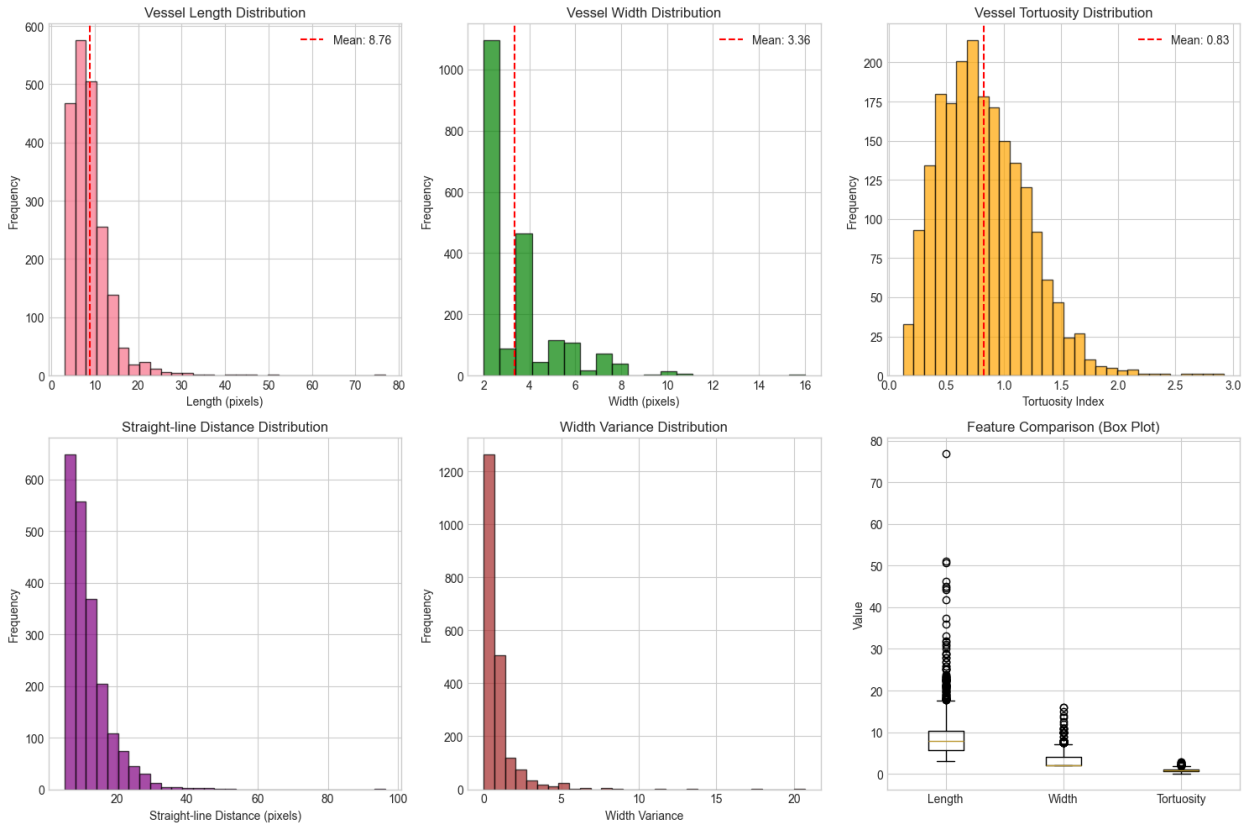



Figure 6: Plots showing distributions of Vessel Segment Statistics for p7 patch (fig. 1b)

Analysis of vessel segment features revealed developmental trends consistent with angiogenesis and vascular remodeling. Vessel length exhibited a right-skewed distribution, with most vessels being short to medium in length. Mean vessel length increased progressively from P2 to P7, reflecting network maturation. Vessel width distributions were often bimodal, indicating the presence of both capillaries and larger vessels. Width variability decreased slightly over development as vessel types differentiated. Tortuosity, defined as the ratio of actual vessel length to straight-line distance, showed a predominance of vessels in the normal range (1.0–1.2). Tortuosity decreased over time, consistent with optimization of vascular paths during maturation.

### 4.2.2 Network Node Statistics

Network-level metrics provided insights into the complexity and branching patterns of the vascular network. Node degree analysis identified endpoints (degree=1), continuation nodes (degree=2), bifurcations (degree=3), and higher-order junctions (degree $\geq 4$). Across development, the bifurcation ratio increased while the endpoint ratio decreased, indicating reduced sprouting and enhanced network complexity. Spatial distribution, measured as distance from the optic nerve head, expanded radially with development, confirming expected patterns of vascular growth.

| Node Type | Degree | Percentage | Biological Interpretation |
|---|---|---|---|
| Endpoint | 1 | 15-25% | Vessel tips, sprouting fronts |
| Continuation | 2 | 40-60% | Vessels pass through |
| Bifurcation | 3 | 20-35% | Vessel splits into two |
| Higher-order | 4+ | 5-15% | Complex junctions |

Table 6: Node degree distribution across the vascular network

| | distance | degree |
|---|---|---|
| **mean** | 191.773612 | 2.160752 |
| **std** | 61.906919 | 1.403927 |
| **min** | 7.280110 | 1.000000 |
| **25%** | 147.838593 | 1.000000 |
| **50%** | 200.158682 | 2.000000 |
| **75%** | 241.457018 | 3.000000 |
| **max** | 319.864034 | 12.000000 |

Table 7: Network Node Statistics for p7 patch

This table summarizes the properties of the network nodes: the points where vessels either terminate or branch. The distance column denotes the distance of each node from the center of the image patch and the degree column denotes the number of segments connected to a node.

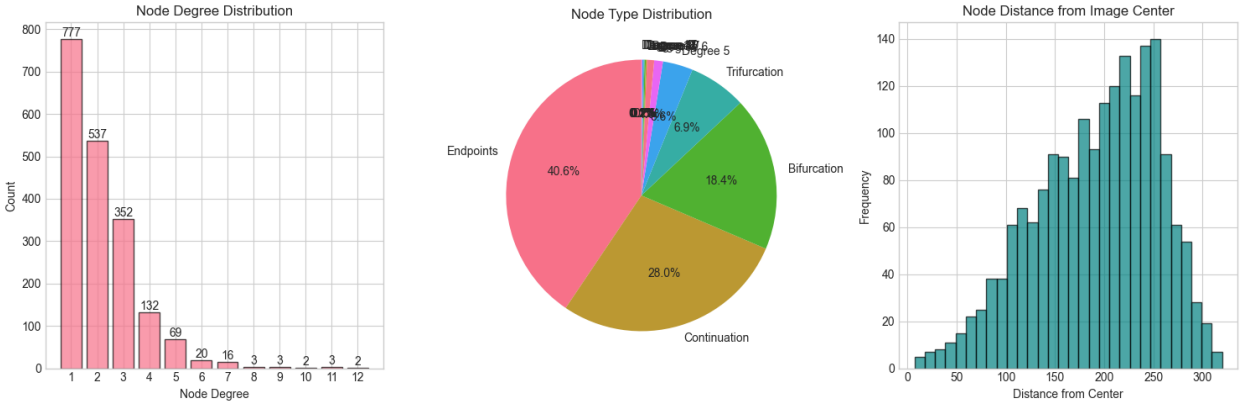



Figure 7: Plots showing distributions of Network Node Statistics Statistics for p7 patch(fig. 1b)

### 4.2.3 Feature Correlations

Correlation analysis among extracted features revealed several noteworthy trends. Vessel length strongly correlated with straight-line distance between endpoints, confirming geometric consistency. Longer vessels were moderately associated with higher tortuosity, indicating that longer segments naturally exhibit more curvature. Width correlated positively with width variability, suggesting that larger vessels show greater variation along their length. Other correlations, such as between tortuosity and width, were weak, indicating relative independence between these morphological properties.

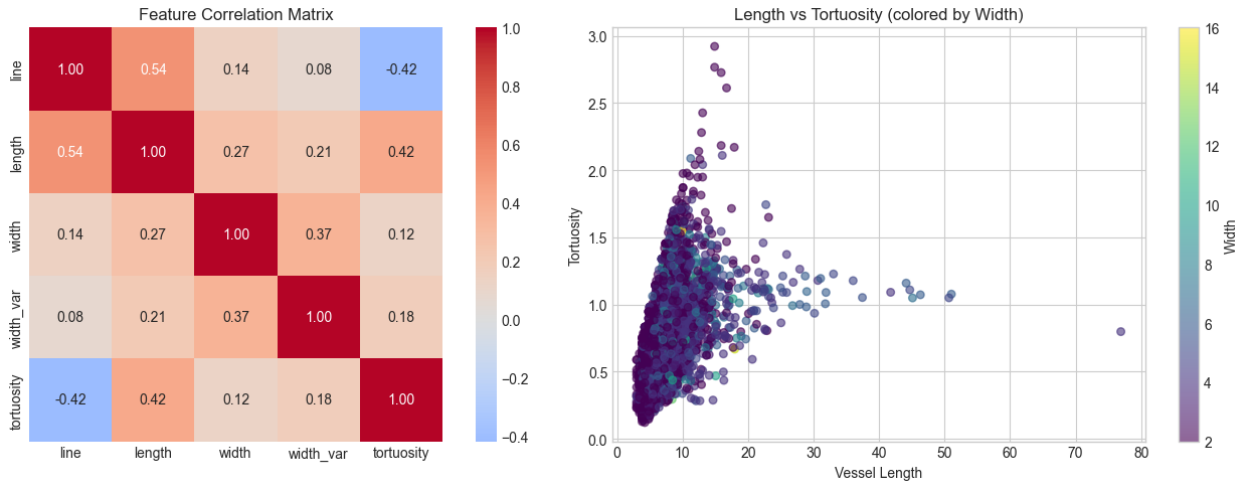



Figure 8: Feature Correlations for p7 patch (fig. 1b)

### 4.2.4 Outlier Analysis

We try to identify any vessel segments with unusually extreme values for length, width, or tortuosity. These outliers might represent unique biological structures or potential measurement artifacts.

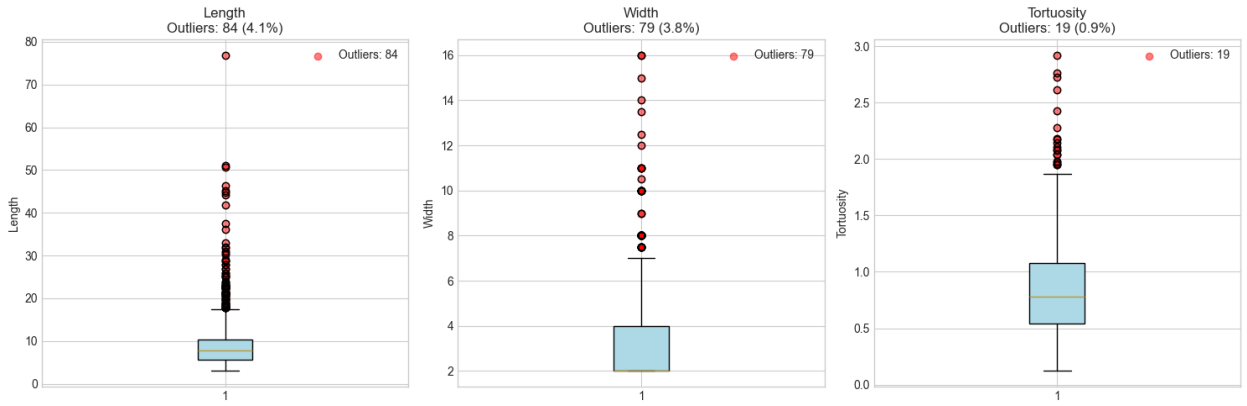



Figure 9: Outlier Detection for p7 patch (fig. 1b)

## 4.3 Developmental Analysis (P2–P7)

Comparative analysis of features across P2–P7 revealed consistent developmental trends. Mean vessel length increased steadily, reflecting network extension. Vessel width showed variability due to differentiation of arteries and veins. Tortuosity decreased as the vascular network optimized its paths. The total number of network nodes increased over time, indicating growing network complexity, while the endpoint ratio decreased and bifurcation ratio increased, signaling the transition from active sprouting to maturation. Radial expansion from the optic nerve head was observed in later stages, with vessels reaching further across the retina.

| Day | Vessel Segments | Network Nodes | Mean Length | Mean Width |
|---|---|---|---|---|
| P2 | 635 | 545 | 8.61 ± 3.52 | 4.36 ± 3.14 |
| P3 | 1180 | 947 | 8.74 ± 3.38 | 4.64 ± 2.88 |
| P4 | 1736 | 1530 | 8.48 ± 3.81 | 3.96 ± 2.59 |
| P5 | 2638 | 2132 | 9.04 ± 4.16 | 4.39 ± 2.83 |
| P6 | 2425 | 2009 | 8.89 ± 4.38 | 4.13 ± 2.41 |
| P7 | 4303 | 3705 | 8.88 ± 5.13 | 3.57 ± 1.88 |

| Day | Mean Tortuosity | Bifurcation % | Endpoint % | Mean Distance |
|---|---|---|---|---|
| P2 | 0.823 ± 0.365 | 19.4% | 39.4% | 100.3 |
| P3 | 0.846 ± 0.362 | 21.1% | 32.6% | 120.2 |
| P4 | 0.818 ± 0.385 | 16.6% | 40.7% | 157.3 |
| P5 | 0.858 ± 0.369 | 23.0% | 32.1% | 190.1 |
| P6 | 0.836 ± 0.374 | 20.4% | 36.2% | 193.0 |
| P7 | 0.833 ± 0.373 | 21.5% | 34.8% | 191.6 |

Table 8: Vascular Network Metrics by Postnatal Day

### 4.4 Summary of Findings

Overall, the segmentation pipeline successfully quantified both vessel morphology and network structure. The results demonstrated progressive vascular growth, increased branching complexity, and reduced tortuosity over development. These findings are consistent with known biological processes of angiogenesis and vascular remodeling in the mouse retina.

## 5 Limitations

While the vessel segmentation pipeline and subsequent EDA provide detailed insights into mouse retinal vascular development, several limitations should be noted. First, the analysis was conducted on 512×512 image patches rather than full retina images, which may omit global network features or peripheral structures. Second, the sample size per postnatal day was limited, potentially reducing statistical power and generalizability. Third, the analysis is restricted to 2D image data, and therefore does not capture the three-dimensional architecture of the vasculature. Fourth, the feature extraction relies on the quality of segmentation; any segmentation errors can propagate into the vascular metrics. Finally, although outliers were identified, some may represent biologically relevant rare events, and further validation with complementary experimental data is recommended.

## 6 Conclusion

In this study, a U-Net based segmentation pipeline was implemented to extract quantitative features of mouse retinal vasculature from postnatal days P2 to P7. The model achieved high performance, with a Dice score of 0.936, and reliably segmented vessels of varying widths and branching complexities. Graph-based analysis of vessel networks revealed developmental trends including increased vessel length, decreased tortuosity, enhanced bifurcation ratios, and radial expansion from the optic nerve head. These results are consistent with known biological processes of angiogenesis and vascular remodeling. The extracted metrics provide a foundation for quantitative comparison across developmental stages, identification of abnormal vascular patterns, and future longitudinal studies. The pipeline and EDA framework are fully reproducible, enabling robust and transparent analysis for further vascular biology research.

# 7 Contribution

This project was collaboratively developed by **Yashas Majmudar** and **Ishita Pethkar**, with each contributor leading key components of the pipeline. Yashas Majmudar was responsible for data preprocessing, dataset preparation, feature extraction, and building the automated one-click execution pipeline, as well as authoring the reproducibility documentation. Ishita Pethkar implemented the U-Net model architecture, training script, and prediction pipeline, and contributed extensively to exploratory data analysis, interpretation of results, and preparation of the environment and packaging for idempotent execution. She also authored the project README and startup documentation. Together, both contributors ensured the robustness, reproducibility, and scientific quality of the complete vessel-segmentation and analysis workflow.

# 8 Use of Artificaial Intelligence

Artificial intelligence played a key role in accelerating development and improving the overall quality of the vessel-segmentation pipeline. AI assistance was used to break down and interpret the problem statement, enabling a clearer understanding of the biological and computational objectives. It also helped analyze the legacy feature-extraction codebase, identify structural inefficiencies, and propose targeted fixes to improve reliability. Documentation throughout the project was expanded and refined with AI support, transforming technical explanations into clear, readable segments suitable for both developers and researchers. AI-driven optimization strategies were applied to improve model performance, streamline training behavior, and enhance prediction speed. Additionally, AI contributed to the generation and evaluation of edge-case scenarios, enabling more rigorous testing of the pipeline's robustness under unusual or challenging input conditions. Through these contributions, AI served as an effective augmentation tool, improving clarity, efficiency, and system reliability across the project.