# Evaluating Neuronal Structure through Functional Summaries of Persistent Homology

Master Project of Applied Mathematics

June 2025


**Author**
Alexia Paratte

**Supervisors**
Lida Kanari
Kathryn Hess Bellwald

**Oral defense Expert**
Martina Scolamiero

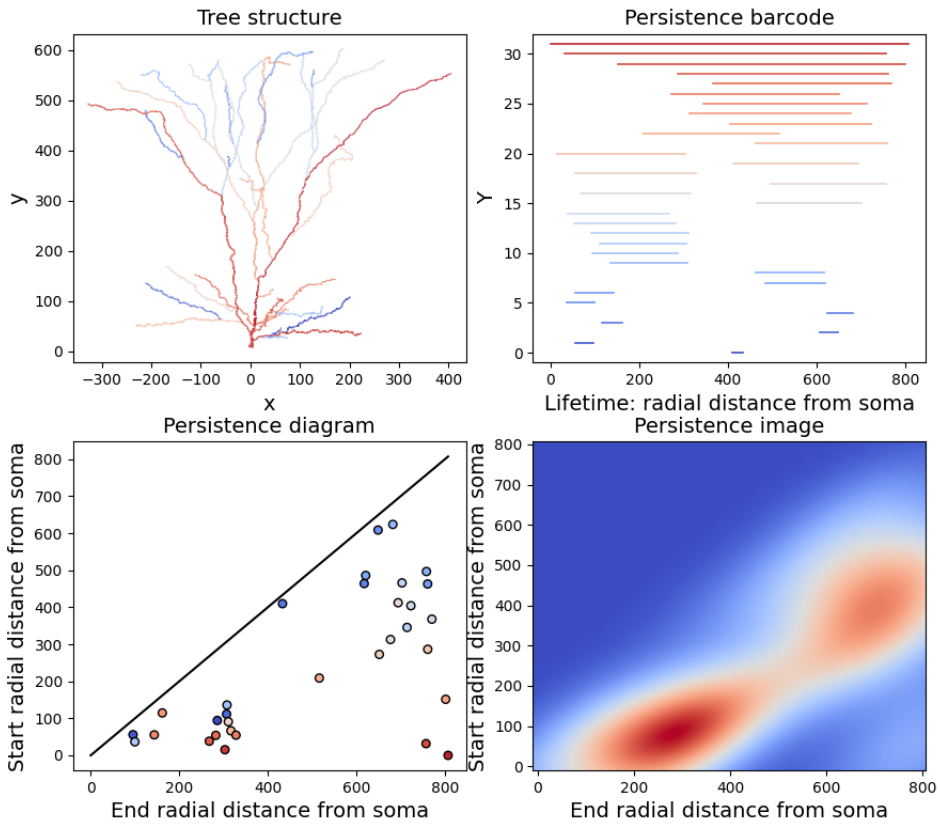




*Figure adapted from Kanari et al., 2025.*

# Abstract

The shapes of biological objects, such as trees, proteins, and brain networks, can be described with various mathematical tools. Numerous different cell types with complex branching structures are present in the brain (neurons, microglia, astrocytes, etc.) and together form complex networks. The relation between the shapes of single cells and the structure of the networks they form is not yet well understood, and the difficulty of studying this link in vivo further hampers our ability to investigate it. In Kanari et al. (2018), the authors developed methods rooted in algebraic topology to analyze the structure of individual neurons and computationally generate synthetic neurons that preserve their topological characteristics (Kanari et al. (2019)).

In this work, we build on these methods to study whether topological features can reveal meaningful structural differences between neuron populations. By applying functional summaries of persistence diagrams to morphological data from human and mouse cortex (Kanari et al. (2025)), we explore the potential of topological data analysis for statistical comparison in neuroscience.

# Acknowledgments

I am deeply grateful to my supervisor, Dr. Lida Kanari, who gave me the opportunity to work on this project within her research group. Her thoughtful guidance, patience, and constructive feedback have been invaluable throughout this journey. Thank you for your scientific insight and encouragement! I would also like to sincerely thank Prof. Kathryn Hess Bellwald for offering me the opportunity to carry out this master's project and for her inspiring work at the intersection of mathematics and neuroscience. It has been a real pleasure and privilege to work with such inspiring women in science. In a field where women remain underrepresented, their leadership, brilliance, and generosity have made this experience especially meaningful.

Although my academic background is primarily in mathematics, I have always been fascinated by biology. This project gave me the unique chance to explore how abstract mathematical concepts can be applied to the study of neuronal structures, an experience that has been both intellectually enriching and personally fulfilling. It also offered a beautiful and meaningful way to conclude my studies, by connecting my mathematical training with a subject that has always inspired my curiosity.

I would like to thank my family and friends for their constant support, encouragement, and belief in me throughout my studies. I am especially grateful to my parents and my sister for putting up with me and my chaotic tendencies, I truly would not have made it this far without them. A special thanks also goes to Shayan, I could not have asked for a funnier and more thoughtful best friend to share these years of Master's studies with! A thought goes to my roommates, whose generosity and culinary skills brought both comfort and joy during the final stretch of this project. Thank you to Noa, Aude, Garance, Romane, Nathan, Jacques, and all the friends who were by my side throughout this journey! I cannot wait for the adventures that lie ahead.

# Contents

# Chapter 1: Introduction

In this project, we investigate structural differences in neuronal morphology using topological and statistical methods, with a particular focus on comparing pyramidal neurons from humans and mice. Rather than aiming to classify neuron types, our objective is to extract shape descriptors that are both mathematically rigorous and biologically meaningful, and to determine whether the differences observed between species are statistically robust.

To this end, we employ tools from Topological Data Analysis (TDA), and in particular, persistent homology, which captures the evolution of topological features – such as connected components and loops – as a function of scale. The primary output of persistent homology is the persistence diagram, a multiset of birth-death pairs that encodes the lifespan of topological features. However, due to their non-Euclidean structure, persistence diagrams are not directly compatible with statistical procedures such as averaging or hypothesis testing. To overcome this limitation, we turn to *functional summaries*, which transform persistence diagrams into continuous, vectorizable representations. These include persistence entropy (a global scalar measure), persistence landscape, and persistence silhouette (both functional descriptors suitable for group comparison and inference).

This project is driven by four central research questions. First, we aim to determine whether human and mouse Layer 2/3 Pyramidal Cells exhibit statistically significant differences in their topological structure. This question stems from recent Kanari et al. (2025), which suggest consistent morphological distinctions across species. Second, we investigate the extent to which persistent homology and its associated functional summaries can capture these structural differences in a meaningful and interpretable way. Third, we assess the robustness of the observed differences by quantifying intra-group variability and applying statistical tools such as confidence bands and permutation tests. This is particularly important given the disparity in sample sizes between the two species. Finally, we evaluate and compare the effectiveness of the different functional summaries in supporting statistical inference.

We apply these methods to a dataset introduced in Kanari et al. (2025), which provides a carefully curated and large-scale collection of digitally reconstructed Layers 2/3 pyramidal neurons from both species. The dataset's biological relevance, consistent anatomical labeling, and cross-species design make it an ideal testbed for our analysis.

In the following sections, we first introduce the theoretical foundations of TDA, with a focus on persistent homology. We then describe the functional summaries used in this study and motivate their role in topological inference. Finally, we present our application to the neuronal dataset, detailing the computation of confidence bands and permutation-based statistical testing to evaluate the robustness of the observed morphological distinctions.

This work was conducted as part of a Master's project in Applied Mathematics at the École Polytechnique Fédérale de Lausanne (EPFL), under the supervision of Kathryn Hess Bellwald and Lida Kanari. All scripts and documentation are available in the accompanying GitHub repository.

# Chapter 2: Theoretical Foundations

A solid understanding of *Topological Data Analysis* (TDA) is essential for interpreting the results of the following analysis. In particular, the method of persistent homology (Edelsbrunner & Morozov (2010)) plays a central role in our study. Persistent homology is a technique from computational topology that allows us to extract and quantify the topological features of data – such as connected components, loops, or voids – and to track their persistence across multiple scales.

The key insight of TDA is that data should not be analyzed merely as a cloud of discrete points, but as a structure with global shape. By focusing on how data points connect and how features appear and disappear as we "zoom out," persistent homology provides a robust, multiscale summary of the underlying geometry. This makes it especially suitable for comparing complex biological structures such as neurons.

## 2.1 Topological and Algebraic Preliminaries

### 2.1.1 Chains, Boundaries, and Homology

A fundamental notion in algebraic topology is the *simplicial complex*, built from simple geometric building blocks called simplices:

- A 0-simplex is a point,
- A 1-simplex is a line segment,
- A 2-simplex is a filled triangle,
- A 3-simplex is a tetrahedron, and so on.

We define a *p-chain* as a formal sum of $p$-simplices with coefficients in a field $\mathbb{F}$ (commonly $\mathbb{Z}_2$). The set of all $p$-chains forms a vector space $C_p$.

The **boundary operator** $\partial_p : C_p \to C_{p-1}$ maps a $p$-simplex to the alternating sum of its $(p-1)$-dimensional faces:

$$\partial_p([v_0, \ldots, v_p]) = \sum_{i=0}^{p} (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_p]$$

where $\hat{v}_i$ indicates that $v_i$ is omitted. Importantly, $\partial_p \circ \partial_{p+1} = 0$, which implies that the boundary of a boundary is always zero.

From this, we define:

- The group of **cycles** $Z_p = \ker(\partial_p)$: chains with no boundary,
- The group of **boundaries** $B_p = \text{im}(\partial_{p+1})$: chains that are boundaries of higher-dimensional chains.

The $p$-th **homology group** is defined as:

$$H_p = Z_p / B_p$$

where $Z_p$ denotes the group of $p$-cycles (closed $p$-dimensional chains), and $B_p$ the group of $p$-boundaries (those that bound $(p+1)$-dimensional chains). Intuitively, $H_p$ captures the $p$-dimensional "holes" in a space: features that are closed but not boundaries. The dimension of each homology group is called a **Betti number**:

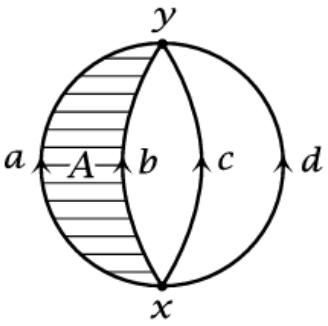


Figure 1

- $\beta_0 = \dim H_0$: counts connected components,
- $\beta_1 = \dim H_1$: counts one-dimensional loops (like rings),
- $\beta_2 = \dim H_2$: counts enclosed voids (like cavities).

Figure 1, adapted from Hatcher (2002), provides a visual example of a 1-dimensional homology class. The loop formed by the paths $b$, $c$, and $d$ is a 1-cycle that is not the boundary of any 2-dimensional cell, and thus represents a nontrivial element of $H_1$.

## 2.2 Persistent Homology

### 2.2.1 Filtrations

Most real-world data is inherently multi-scale. Persistent homology captures how topological features evolve across a range of scales. To do so, we consider a **filtration**, a nested sequence of simplicial complexes:

$$\emptyset = X_0 \subseteq X_1 \subseteq \cdots \subseteq X_n = X$$

Each inclusion adds simplices, gradually enriching the topology. It is typically constructed by growing the space according to a scalar function (e.g., a distance from a root point). Applying homology at each stage yields a sequence of vector spaces and linear maps:

$$H_p(X_0) \to H_p(X_1) \to \cdots \to H_p(X_n)$$

This forms a **persistence module**, from which we track the *birth* and *death* of homological features: when they appear and when they are filled in or merged.

### 2.2.2 Persistence Diagrams

Persistent homology analyzes how the topology of a space changes across a sequence of nested subspaces, which are the filtration defined earlier. In the case of tree-like neuronal structures, for instance, we may use a path-distance or radial-distance function from the soma to generate such a filtration. As the parameter $t$ increases, new topological features – such as connected components or loops – can appear (birth) and later merge or disappear (death). The output of persistent homology is recorded in a **persistence diagram** (see Figure 2), a multiset of points $(b, d)$ in the extended plane $\mathbb{R}^2$, where $b$ and $d$ represent the birth and death times of a given topological feature across the filtration. Each point encodes the lifetime of a feature: points near the diagonal $b = d$ correspond to short-lived features typically considered noise, whereas points farther from the diagonal indicate persistent features likely to represent meaningful structure. This distinction is central to the interpretability of persistent homology, as emphasized in Carlsson (2009).

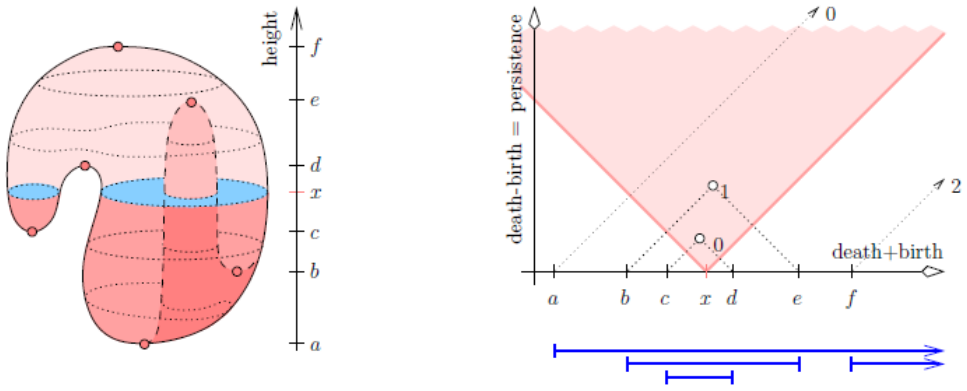



Figure 2: Illustration of a filtration induced by a height function on a topological surface and its associated persistence diagram. **Left**: The sublevel sets grow as the filtration parameter (height) increases, with new components appearing and merging over time. **Right**: The corresponding persistence diagram encodes the birth and death of topological features, with dots representing connected components ($\beta_0$) and loops ($\beta_1$). The lower diagram shows the equivalent barcode representation. Adapted from Edelsbrunner & Morozov (2010).

## 2.3 The Topological Morphology Descriptor: Algorithm and Mathematical Background

The Topological Morphology Descriptor (TMD) is a method for representing tree-like structures such as neurons using persistent homology Kanari et al. (2018). In the TMD framework, a neuron is modeled as a rooted tree embedded in a metric space (typically $\mathbb{R}^3$), where the root corresponds to the soma. A scalar function – most commonly the radial distance (Euclidean distance from the soma) or the path distance (accumulated length along the tree) – is defined on the tree.

This function induces a filtration, where connected components appear and merge as the filtration parameter increases. By applying persistent homology to this filtration, we obtain a persistence barcode, which records the birth and death of connected components across the tree (see Figure 3 from Kanari et al. (2018)). Each bar in the barcode corresponds to a branch or subtree: it is born when it first appears and dies when it merges with an older component.

The key advantage of TMD is that it captures the hierarchical structure of the neuron while remaining invariant to isometric transformations and robust to small perturbations. It reduces a complex 3D structure to a compact topological signature that can be used for comparison, clustering, or statistical inference. Because it relies only on the tree structure and a filtration function, the algorithm has linear complexity in the number of nodes, making it scalable for large datasets.

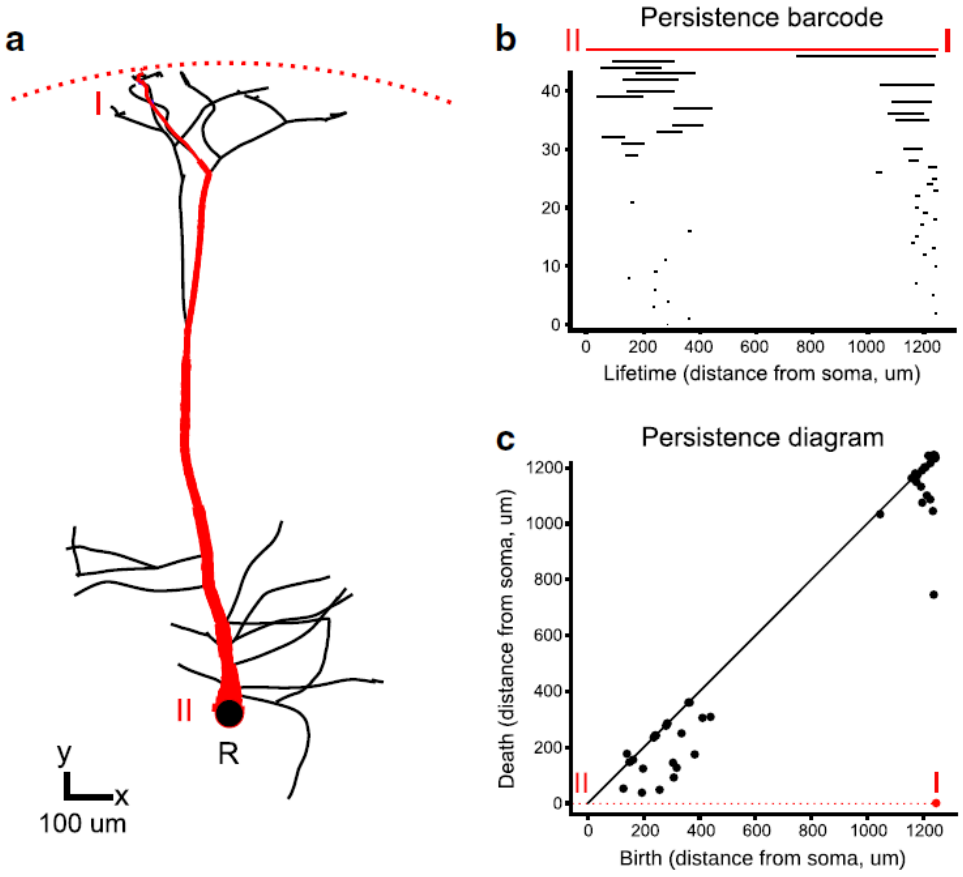



Figure 3: Topological analysis of a neuronal tree using TMD. **(a)** The tree is rooted at the soma (R), with the largest persistent component shown in red. **(b)** The persistence barcode encodes birth and death times (I, II) of connected components based on a scalar function. **(c)** The same information is shown as a persistence diagram. Adapted from Kanari et al. (2018).

## 2.4 Stability of Persistent Homology and the TMD

To compare persistence diagrams quantitatively, we rely on the following metrics:

- **Wasserstein distance** $W_q$ for $q \geq 1$: aggregates distances between matched points, providing a balance between global and local feature differences. Notably, persistence images are stable with respect to $W_1$.
- **Bottleneck distance** $W_\infty$: measures the largest displacement among matched features, and is especially relevant in contexts requiring worst-case guarantees, such as TMD-based comparisons.

A key strength of persistent homology is its stability: small perturbations in the input lead to small changes in the resulting persistence diagram. This property is particularly valuable when working with biological data, which is often subject to noise, reconstruction artifacts, or natural variability. The **Stability Theorem** (Cohen-Steiner et al. (2007)) provides a formal guarantee that the Bottleneck distance between persistence diagrams is bounded by the maximum deviation between the input functions:

$$W_\infty(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq \|f - g\|_\infty$$

This theoretical result ensures that persistent homology is robust under bounded perturbations and supports its application to real-world data.

In the context of neuronal morphology, this stability is also observed in practice. The TMD framework (Kanari et al. (2018)) uses the Bottleneck distance to compare barcodes derived from neuronal trees, and has demonstrated that topological signatures are robust under small deformations or sampling errors in the structure. More recently, Beers et al. (2023) emphasized that persistence-based descriptors such as persistence surfaces and persistence images are not only interpretable but also **stable with respect to the 1-Wasserstein distance**. This property was established formally by Adams et al. (2017), who proved that the distance between persistence surfaces is bounded above by the 1-Wasserstein distance between the corresponding diagrams (Theorem 4). This result ensures that small perturbations in the diagram lead to proportionally small changes in the surface representation, a key requirement for reliable statistical comparison of neuron populations.

## 2.5 Computational Aspects

Computing persistent homology reduces to matrix operations over vector spaces. Specifically:

- Chain groups $C_p$ are vector spaces over a finite field (e.g., $\mathbb{Z}_2$),
- Boundary maps $\partial_p$ are represented as matrices,
- Homology groups are computed as:

$$H_p = \ker(\partial_p)/\operatorname{im}(\partial_{p+1})$$

Matrix reduction algorithms (such as the column-reduction algorithm) enable efficient computation of persistence pairs (birth-death intervals), which are then visualized in the persistence diagram. This pipeline is detailed in the tutorial by Edelsbrunner & Morozov (2010).

## 2.6 Functional Summaries of Persistence Diagrams

Persistence diagrams are powerful descriptors of topological features in data (Berry et al. (2020)), however they are not inherently suited for statistical operations such as averaging, hypothesis testing, or classification. To enable such analyses, persistence diagrams can be transformed into functional summaries, which map the multiset of birth-death pairs to real-valued functions. These summaries provide a stable and differentiable representation, making them compatible with standard statistical and machine learning tools.

Let $\mathcal{D}$ denote the space of persistence diagrams and $\mathcal{F}$ a space of real-valued functions. A **functional summary** is a mapping:

$$F : \mathcal{D} \to \mathcal{F}$$

such that each diagram $D \in \mathcal{D}$ is mapped to a function $F(D)$. When analyzing multiple samples $D_1, \ldots, D_n$ drawn i.i.d. from a common distribution, their functional summaries $F(D_1), \ldots, F(D_n)$ are treated as i.i.d. random functions in $\mathcal{F}$.

### 2.6.1 Persistence Entropy and the Life Entropy Curve

Persistent homology encodes multiscale topological features of data as a collection of birth–death intervals (bars) in a *persistence diagram* $D = \{(b_i, d_i)\}$, where each interval corresponds to a topological feature born at scale $b_i$ and dying at scale $d_i$. However, the space of persistence diagrams lacks convenient algebraic structure, which motivates the need for *vectorization methods* – mappings from barcodes to real-valued features suitable for statistical or learning pipelines.

Among the simplest and most interpretable vectorizations is the **entropy**, introduced by Chintakunta et al. (2015), which applies Shannon entropy to the distribution of bar lifespans:

$$E = -\sum_{[p,q]\in B} \left(\frac{q-p}{L}\right) \log\left(\frac{q-p}{L}\right),$$

where

$$L = \sum_{[p,q]\in B} (q-p)$$

is the total lifespan of all features in the diagram. The entropy $E$ quantifies the diversity of topological features: higher entropy indicates a more uniform distribution of lifespans, while lower values suggest a dominance of a few long-lived bars.

To enrich this scalar summary, Atienza et al. (2020) proposed the **life entropy curve**, also called the entropy summary function. This function measures the entropy contribution of features that are alive at a given filtration scale $t$, and is defined as:

$$S(t) = -\sum_{[p,q]\in B} \mathbf{1}_{[p\leq t<q]} \cdot \left(\frac{q-p}{L}\right) \log\left(\frac{q-p}{L}\right).$$

The function $S : \mathbb{R} \to \mathbb{R}$ is piecewise constant and provides a scale-resolved view of topological complexity. As a functional summary, it belongs to the family of curve-based vectorizations and enables statistical comparisons in function spaces.

These entropy-based descriptors are computationally efficient, interpretable, and stable with respect to small perturbations in data, making them attractive tools in TDA (Ali et al. (2023)).

### 2.6.2 Persistence Landscape

Persistence landscapes, introduced by Bubenik (2015), are functional summaries of persistence diagrams that allow the use of standard statistical techniques in the realm of TDA. While persistence diagrams represent topological features as multisets of birth-death pairs, their non-vectorial nature makes statistical operations, like averaging or computing variances, non-trivial. Persistence landscapes resolve this by embedding diagrams into a separable Banach space, thereby enabling functional data analysis.

Given a persistence diagram composed of points $(b_i, d_i)$, each point is mapped to a triangular function:

$$f_i(t) = \max\{0, \min(t - b_i, d_i - t)\}$$

This function has its peak at the midpoint $(b_i + d_i)/2$ with height $(d_i - b_i)/2$. The $k$-th **persistence landscape function**, denoted $\lambda_k(t)$, is defined as the $k$-th largest value among all $f_i(t)$ at each $t$:

$$\lambda_k(t) = \text{k-th largest value of } \{f_i(t)\}_{i=1}^N$$

The full persistence landscape is the sequence $\{\lambda_k\}_{k=1}^\infty$.

This representation offers several advantages. First, it is stable with respect to perturbations in the data, ensuring robustness. Second, it facilitates statistical analysis, allowing the computation of means, variances, and the use of hypothesis testing procedures. Third, it enables visualization and comparison of complex topological features across samples in a coherent and interpretable manner.

### 2.6.3 Persistence Silhouette

The **persistence silhouette** is a functional summary of persistence diagrams designed to provide a smooth, compact, and informative representation of topological features. As described in Berry et al. (2020) the silhouette belongs to a family of function-based summaries that transform a persistence diagram into a real-valued function over the filtration scale. This transformation enables the application of standard statistical tools to topological data.

Formally, let $D = \{(b_i, d_i)\}_{i=1}^N$ be a persistence diagram. Each interval is associated with a triangular function $\Lambda_i(t)$ centered at $(b_i + d_i)/2$ with height $(d_i - b_i)/2$. The silhouette is then defined as a weighted average of these functions:

$$\text{Sil}_p(D; t) = \frac{\sum_{i=1}^N w_i \Lambda_i(t)}{\sum_{i=1}^N w_i}, \quad \text{where } w_i = |d_i - b_i|^p$$

for some fixed exponent $p > 0$. The weights emphasize the contribution of more persistent features and suppress those closer to the diagonal.

Compared to the persistence landscape, which captures all feature peaks via a sequence of functions $\{\lambda_k\}$, the silhouette aggregates this information into a single smoothed function, offering lower dimensionality and improved interpretability. Unlike scalar summaries like persistent entropy, the silhouette retains the distributional shape of topological features across the filtration axis, making it a useful compromise between expressiveness and simplicity.

Berry et al. (2020) emphasize that the persistence silhouette inherits important theoretical properties such as stability, and it allows for the computation of statistical descriptors including means, variances, and confidence bands. These attributes make it particularly well suited for topological inference and comparison tasks.

## 2.7 Statistical Inference from Functional Summaries

Functional summaries of persistence diagrams not only enable visualization and descriptive analysis, but also facilitate formal statistical inference. Two central tools in this regard are the construction of *confidence bands* for mean functions and the application of *two-sample hypothesis testing* procedures. These approaches extend classical techniques from functional data analysis to the topological domain, enabling rigorous comparison of groups of diagrams.

### 2.7.1 Confidence Bands

To quantify uncertainty around the estimated mean functional summary $\hat{F}(t)$, we construct **confidence bands** using a nonparametric bootstrap procedure:

1. Compute the sample mean function: $\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} F_i(t)$
2. Generate $B$ bootstrap samples by resampling with replacement from $\{F_1, \ldots, F_n\}$, and compute the bootstrap means $\hat{F}^{*(b)}(t)$
3. Determine the $1-\alpha$ quantile of the maximum deviation:

$$\hat{t}_{1-\alpha} = \inf\left\{ s : \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left( \sup_t |\hat{F}^{*(b)}(t) - \hat{F}(t)| \leq s \right) \geq 1-\alpha \right\}$$

4. Construct the uniform confidence band:

$$L_{1-\alpha}(t) = \hat{F}(t) - \hat{t}_{1-\alpha}, \quad U_{1-\alpha}(t) = \hat{F}(t) + \hat{t}_{1-\alpha}$$

Such bands provide a visual and quantitative indication of where differences between groups (e.g., human vs. mouse neurons) exceed expected variability due to sampling.

### 2.7.2 Two-Sample Hypothesis Testing

To assess whether two samples of diagrams originate from the same distribution, we employ a **two-sample hypothesis test** based on the $L^\infty$ distance between mean functional summaries:

1. Let $D^{(1)} = \{D_1^{(1)}, \ldots, D_n^{(1)}\}$ and $D^{(2)} = \{D_1^{(2)}, \ldots, D_m^{(2)}\}$ be two groups of diagrams
2. Compute mean functions $\hat{F}_1(t)$ and $\hat{F}_2(t)$ for each group
3. Define the test statistic: $T = \|\hat{F}_1 - \hat{F}_2\|_{L^2}$
4. Perform a permutation test: shuffle group labels, recompute $T^*$ for each permutation, and estimate the $p$-value as the proportion of permutations for which $T^* \geq T$

This approach, aligned with the framework presented in Robinson & Turner (2017) and Bubenik (2015), enables robust, non-parametric testing for differences in topological structure between populations. It is particularly well-suited when using stable and interpretable functional summaries such as the persistence landscape or silhouette.

# Chapter 3: Packages Overview

## 3.1 TMD

The Topological Morphology Descriptor (TMD) plays a foundational role in our analysis pipeline. While its mathematical formulation and theoretical guarantees were introduced in Chapter 2, this section focuses on its practical implementation and the motivations behind our methodological choices.

TMD is specifically designed for analyzing tree-like structures, making it highly suitable for digitized neuronal reconstructions, which can be modeled as rooted trees. In our study, we restrict the analysis to the apical dendrites of pyramidal neurons. This choice is guided by their functional and anatomical importance, as well as their prominence in recent comparative studies of cortical morphology, such as Kanari et al. (2025).

A key step in applying TMD is the selection of a scalar function to induce the filtration on the tree. We chose the **path distance** from the soma, which corresponds to the cumulative branch length from the root to each node. This choice is biologically meaningful, as it reflects the traversal distance along the dendritic arbor rather than the straight-line (Euclidean) distance. Previous work has shown that using path-based filtrations better preserves the hierarchical and developmental structure of neurons (Kanari et al. (2018).)

We used the open-source `TMD` Python package to compute persistence barcodes from ASC-formatted neuron reconstructions. The software processes the tree, applies the chosen filtration function, and outputs the corresponding persistence barcode, which encodes the birth and death of connected components (branches) as a function of the filtration scale. These barcodes serve as input to subsequent analysis steps, including functional summary extraction via `giotto-tda`.

Each bar in a barcode corresponds to a subtree or branch in the apical dendrite. A bar is born when a new branch appears during the filtration and dies when it merges with a more central component. Long bars correspond to persistent and structurally significant features, while shorter bars often reflect minor offshoots or noise. This encoding allows for a compact, topologically meaningful representation of neuronal shape.

In contrast to traditional morphometric analyses – which rely on hand-crafted features such as total length or number of bifurcations – TMD provides a mathematically principled and biologically grounded summary of neuronal structure. Its use of persistent homology ensures invariance to translation, rotation, and small-scale deformations. As such, it is robust to noise and reconstruction variability, both of which are common in digital morphologies.

Overall, TMD acts as the bridge between raw neuronal structure and statistical analysis. The barcodes it produces serve as the topological input for all downstream stages of our workflow: functional summarization, confidence band construction, and hypothesis testing.

The GitHub repository for the TMD package is accessible at https://github.com/BlueBrain/TMD.

## 3.2 Giotto-TDA

As part of this project, we leverage the Python package `giotto-tda`, a comprehensive TDA toolkit designed to bridge the gap between topological methods and modern machine learning workflows (Tauzin et al. (2021)). TDA offers a powerful mathematical framework for analyzing the shape of data by capturing global structural features such as connectivity, loops, or voids; features particularly well-suited for biological and geometric data such as neuronal morphologies. Despite its mathematical depth and increasing relevance, TDA has long remained underutilized in applied settings due to a lack of accessible, scalable tools. The `giotto-tda` library addresses this need by offering a robust, user-friendly, and performant interface, tightly integrated with the widely used `scikit-learn` ecosystem.

Developed with a strong emphasis on modularity, efficiency, and interpretability, `giotto-tda` provides tools for computing persistence diagrams from a wide range of data types – including point clouds, time series, images, and graphs – as well as for deriving functional and vectorized representations of these diagrams. The library wraps high-performance C++ backends (e.g., Ripser, GUDHI, Flagser) and exposes a Python interface that adheres to scikit-learn's estimator design, making it straightforward to incorporate topological features into classical machine learning pipelines.

In our study, `giotto-tda` plays a crucial role in extending the outputs of the TMD algorithm. While the TMD provides persistence diagrams that encode the branching structures of neurons, `giotto-tda` allows us to compute and explore two central functional summaries derived from these diagrams: the persistence landscape and the persistence silhouette. These functions are particularly well-suited for statistical analysis and averaging. The silhouette function further simplifies this representation by aggregating all features into a single smooth curve, weighted by their persistence. This results in a continuous, interpretable, and stable summary of the data's topology, ideal for comparing groups of neuronal morphologies.

By using `giotto-tda`'s implementation of both the persistence silhouette and the persistence landscape, we are able to conduct further statistical analysis, including the computation of mean functions and the construction of confidence bands. These summaries are used to quantitatively compare neuron morphologies across species, assess intra-group variability, and test the significance of observed structural differences.

Furthermore, the library's design supports hyperparameter optimization, visualization (via Plotly-based interactive plots), and advanced feature extraction techniques (such as persistence images and kernel methods). Although our application focuses on landscape and silhouette functions, the broader set of tools offered by `giotto-tda` could be used in future extensions of this work, for instance to classify neurons using supervised learning on topological features or to explore their connectivity using Mapper algorithms.

While `giotto-tda` is a general-purpose library, it is notably well-suited for neuroscience applications. Persistent homology has already shown promise in the analysis of functional brain networks (Reimann et al. (2017)) and neuron geometries. The authors of `giotto-tda` themselves include researchers from the Laboratory for Topology and Neuroscience at EPFL, and the library has been used in studies that aim to uncover topological signatures in neural data. Its compatibility with outputs from TMD makes it a natural fit for this project, enabling a pipeline from neuron reconstruction to interpretable topological statistics.

In summary, `giotto-tda` provides the essential infrastructure to analyze, summarize, and compare the topological features of neuronal morphologies derived from TMD. Its stability, scalability, and integration with machine learning tools make it an indispensable component in our exploration of neuronal structure using topological methods.

The official GitHub repository of the `giotto-tda` library can be found at https://github.com/giotto-ai/giotto-tda.

# Chapter 4: Application to Neuronal Morphologies

The classification and comparison of neuronal morphologies is a challenging task due to the complexity and high dimensionality of neuronal arborizations. Traditional shape descriptors often fail to capture the global structural patterns that distinguish neuron types or species. Topological approaches, in particular the TMD introduced in Kanari et al. (2018), offer a powerful framework for addressing these limitations. TMD enables the comparison of tree-like neuronal structures using persistent homology, providing a robust and compact representation of branching patterns. These representations have been successfully applied for clustering and classification of neurons, and for comparing morphologies across experimental conditions.

However, we aim to go a step further: rather than focusing only on whether topological methods can distinguish groups of neurons, we investigate the statistical properties of this separation. Specifically, we ask whether persistent homology – through various functional summaries – can reliably capture biologically meaningful structural differences, and whether these differences are statistically significant across species.

This section presents the application of topological descriptors to the analysis of neuronal morphologies. Our goal is to assess how functional summaries of persistence diagrams, including entropy, persistence landscapes, and persistence silhouettes, can capture biologically meaningful structural differences between neurons. These summaries offer complementary perspectives: entropy provides a scalar quantification of topological complexity, while landscape and silhouette functions enable richer, function-based comparisons between neuronal structures. Through this analysis, we aim to determine whether topological features can robustly distinguish neurons of the same type across species and reflect underlying biological variation.

To test the robustness of these methods, we apply them to two datasets of reconstructed neurons from human and mouse cortex. The analysis focuses on Layer 2/3 Pyramidal Cells (L23 PCs), a morphologically and functionally well-characterized type of excitatory neuron. We restrict our topological analysis to the **apical dendrite** of each neuron, which plays a central role in signal integration and is known to exhibit species-dependent morphological variability. In computing persistence diagrams, we use the **path distance** from the soma, rather than the Euclidean (radial) distance, as it more accurately reflects the biological flow of information along the dendritic tree. In what follows, we provide a concise biological background to situate this choice, followed by a detailed presentation of the datasets and experimental pipeline.

### Biological Background

Neurons are among the most structurally diverse cell types in the body. Their branching patterns, comprising basal and apical dendrites, as well as axons, play a central role in signal integration, propagation, and network connectivity. Morphological features such as branch length, bifurcation angles, and arbor complexity influence the properties of neurons and have been shown to correlate with functional roles across brain regions and species (Beers et al. (2023), Elston (2002)).

In particular, L23 PCs are excitatory neurons located in the superficial layers of the cortex. They exhibit rich dendritic trees and are involved in local and long-range cortico-cortical communication. Morphological variations in these cells have been documented both across species and across cortical areas, reflecting evolutionary adaptations and region-specific functional demands. Such variability makes L23 PCs a valuable model for assessing whether topological descriptors can identify biologically grounded morphological distinctions.

### Dataset Description

We analyze two datasets of digitally reconstructed L23 PCs:

- Human dataset: 186 neurons from adult human cortex
- Mouse dataset: 38 neurons from mouse cortex, similarly extracted and curated

All neurons were preprocessed to retain consistent topological structures, ensuring compatibility with the persistent homology pipeline. Although the human dataset is larger, this difference in sample size is addressed in the statistical analysis, where we use permutation testing and confidence bands to assess the robustness and significance of the results.

In the subsequent subsections, we apply the three descriptors – entropy, persistence landscape, and persistence silhouette – to the persistence diagrams obtained from each neuron. We analyze their distributions across species, construct confidence intervals around mean functions, and perform statistical tests to evaluate the discriminative power of each method.

## 4.1 Persistence Entropy

In this section, we apply entropy and the life entropy curve introduced in Section 2.6.1 to persistence diagrams derived from the morphological structure of individual neurons. These topological descriptors allow us to summarize the complexity and variability of branching patterns, which are central to neuronal function. In the context of neuroscience, dendritic and axonal arborizations are not merely geometric: they influence synaptic integration, firing behavior, and ultimately, the neuron's role within a circuit (Elston (2002)).

We begin our application study with entropy-based descriptors for two reasons. First, persistent entropy offers a computationally efficient scalar measure that provides immediate insight into the topological distribution of features. Second, and more practically, this approach is implemented as part of the `TMD` Python library (Kanari et al. (2018)), which makes it a natural entry point for our analysis pipeline.

In addition to the scalar entropy values, we compute and compare the life entropy curves between groups. These functional summaries allow us to track how topological information content varies over the filtration scale, offering a more detailed view of morphological complexity. In subsequent sections, we complement this analysis with persistence landscapes and silhouette functions, which offer richer representations but require more computational and statistical care. The results of this analysis, including statistical comparisons and visualizations, are detailed in the accompanying Notebook.

**Results**

To introduce entropy-based descriptors, we first computed the persistence entropy for each neuron's persistence diagram. This scalar measure summarizes the diversity of bar lengths (i.e., the lifetimes of topological features), providing a compact and interpretable overview of the branching complexity encoded by persistent homology. A high persistence entropy value indicates a more uniform distribution of bar lengths, suggesting a rich and balanced topological structure. In contrast, a lower value reflects the dominance of a few prominent features, implying a more hierarchical or simplified topology.

We applied this descriptor to the apical dendrites of L23 PCs in both human and mouse cortical neurons. As shown in Table 1, the mean persistence entropy is 2.984 for human neurons and 2.617 for mouse neurons, with standard deviations of 0.420 and 0.319, respectively. This difference suggests that human apical dendrites tend to support a more evenly distributed set of topological features, reflecting greater morphological diversity and complexity in their arborization patterns.

Although the absolute difference in entropy values may appear moderate, it is statistically meaningful. A Welch's $t$-test, which accounts for the unequal sample sizes and variances in the two groups, yielded a $t$-value of 13.45 with a $p$-value less than 0.0001, confirming the significance of the observed difference. This result indicates that the topological structure of apical dendrites in human neurons is not only more complex but also more consistently diverse across the population compared to that of mice.

While persistence entropy offers a useful scalar overview, it does not reveal how this topological complexity is distributed across the filtration process. To gain a more detailed understanding, we turn to the life entropy curve, a functional summary that captures the time-resolved

| **Metric** | **Human** | **Mouse** |
|---|---|---|
| Mean Persistence Entropy | 2.984 | 2.617 |
| Standard Deviation Entropy | 0.420 | 0.319 |
| AUC (mean area under entropy curve) | 869 | 490 |
| Welch t-test ($t$-value) | 13.45 | |
| Welch t-test ($p$-value) | $< 0.0001$ | |

Table 1: Summary statistics for entropy-based descriptors, L23 PCs apical dendrites.

evolution of topological entropy. This allows us to track how persistent features contribute to complexity across different filtration scales, offering a finer-grained view of morphological differences between species.

Figure 4 shows the mean life entropy curves for both species. The higher entropy curve for human neurons suggests a more balanced distribution of topological feature lifespans, indicating a potentially richer morphological structure compared to mouse neurons. In contrast, the sharper and earlier peak in the mouse curve, followed by a rapid decay, reflects a dominance of short-lived topological features, suggesting simpler and less distributed arborization patterns.

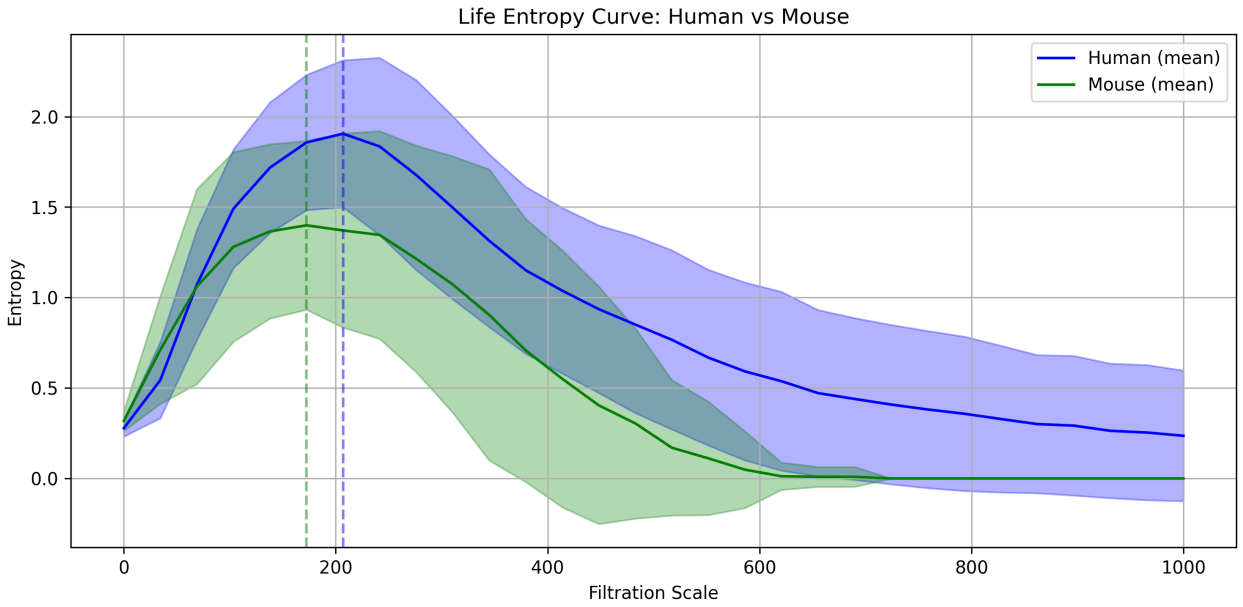



Figure 4: Mean life entropy curves for human and mouse (L23 PCs, apical dendrites). Shaded regions represent one standard deviation. Dashed vertical lines indicate the filtration scale at which each group reaches peak entropy.

## 4.2 Persistence Landscape

In this section, we investigate the use of persistence landscapes introduced in Section 2.6.2 to analyze and compare the morphological structure of neurons. Persistence landscapes provide a powerful vectorization of persistence diagrams, enabling a functional representation of the underlying topological features. This makes them particularly suitable for studying structural variability across populations of neurons and for supporting robust statistical analysis.

Unlike entropy-based summaries that reduce a persistence diagram to a single scalar value, persistence landscapes preserve a richer description by encoding the prominence and localization of topological features through piecewise-linear functions. This allows us to capture not just the quantity of topological information, but also how it is distributed across the filtration scale. For neuronal trees, this can reveal differences in branching depth, extent, and hierarchical organization.

In practice, we proceed as follows. First, we compute the persistence diagrams of the apical dendrites using the TMD framework. These diagrams are then transformed into persistence landscape functions using a transformer from the Giotto-TDA library. This is done separately for the human and mouse datasets. We begin by plotting the individual landscapes for selected representative neurons from each group, based on the area under the first landscape level. In particular, we visualize neurons corresponding to the minimum, lower quartile, median, upper quartile, and maximum landscape area, therefore capturing the full spectrum of topological variability within each group.

Next, we assess the central tendency and variability within each group by computing the average persistence landscape and constructing confidence bands. These bands are obtained via a bootstrap resampling procedure with 500 repetitions, enabling a non-parametric estimate of variability around the mean landscape. Finally, we directly compare the two groups by overlapping their respective confidence bands. This visual comparison serves as a diagnostic tool for detecting structural differences in neuronal topology, which are then further quantified using formal statistical tests.

### Results

To better understand structural variation across individual neurons, we visualize selected persistence landscapes for both human and mouse groups. Figures 5 and 6 display five representative curves for each species, selected according to the area under the first landscape level. This strategy allows us to examine the most and least topologically complex neurons (in terms of persistent features), as well as typical and intermediate configurations.

In the mouse group (Figure 5), the first landscape function $\lambda_1$ consistently displays a pronounced initial peak centered around a narrow filtration range. This concentration of topo-

logical activity suggests that mouse apical dendrites tend to contain one or more features with high persistence and relatively similar birth-death profiles across neurons. The upper levels $\lambda_k$ decay rapidly, reflecting a relatively limited number of persistent features per neuron. By contrast, the human neurons (Figure 6) exhibit broader variability both in peak height and in the spread of topological activity along the filtration scale. Some landscapes show a prominent leading feature (high $\lambda_1$), while others display multiple moderately persistent features (non-negligible $\lambda_2$, $\lambda_3$) contributing to a wider support. This indicates a greater heterogeneity in branching depth or distribution of apical segments, consistent with known morphological complexity in human pyramidal cells.

We also computed 95% confidence bands around the mean landscape functions for both species using a bootstrap procedure with $B = 500$ resamplings. The results, shown in Figure 7, illustrate the mean $\lambda_1$ landscape per group along with their corresponding 95% confidence intervals.

Importantly, the confidence band for the mouse group (red) is substantially wider than the one for the human group (blue). This is mainly due to the smaller sample size for mice ($n = 38$) compared to humans ($n = 186$), which increases the variance of the bootstrap distribution. As a result, the observed band width reflects statistical uncertainty rather than greater morphological heterogeneity. However, notable differences remain visible between the two groups. In the early part of the filtration domain ($t \in [10, 30]$), the human mean landscape exceeds that of the mouse group. While the confidence bands do overlap, the divergence in mean curves suggests that prominent topological features in the human group tend to arise earlier and persist longer in the filtration. This reconfirm the hypothesis that the two species differ in their apical dendritic architecture in a way that is captured by persistent homology.

Thus, persistence landscapes coupled with confidence bands allow not only for central tendency comparison but also for assessing uncertainty, particularly useful when dealing with asymmetric group sizes.

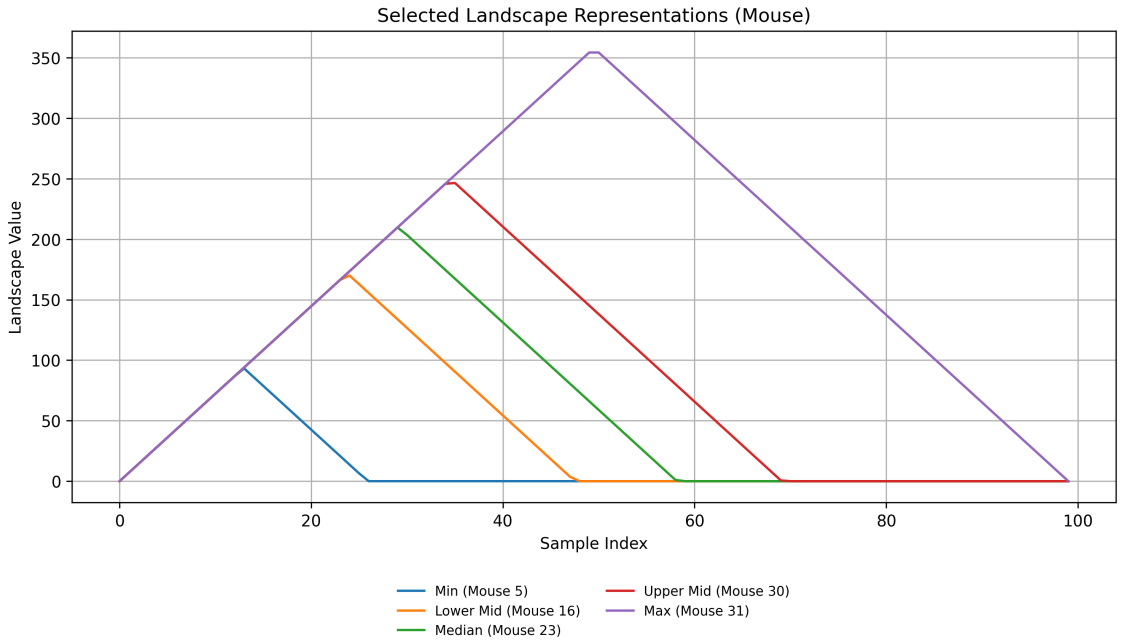



Figure 5: Selected landscape functions for representative mouse neurons (min, lower mid, median, upper mid, max) based on area under the first landscape level.

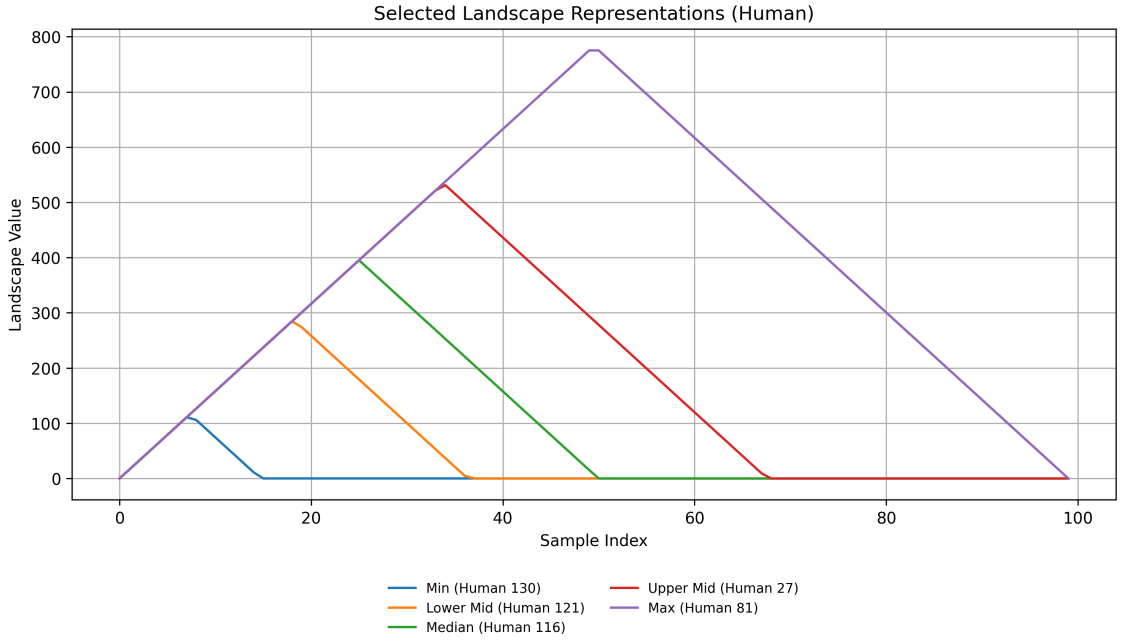



Figure 6: Selected landscape functions for representative human neurons (min, lower mid, median, upper mid, max) based on area under the first landscape level.

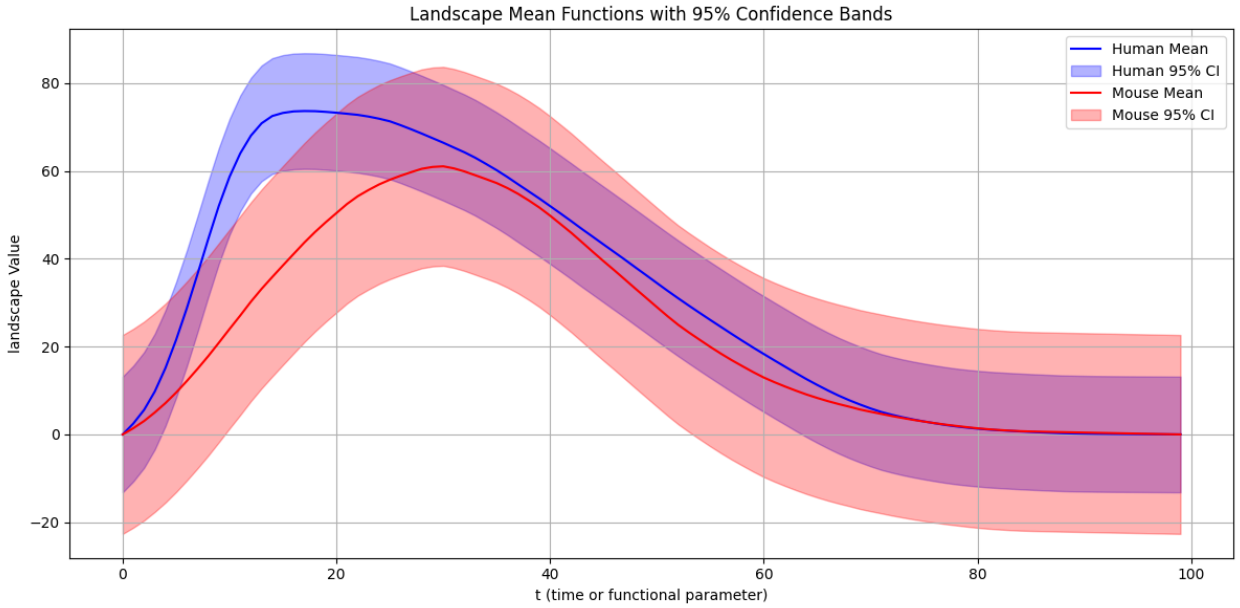



Figure 7: Mean persistence landscapes for human (blue) and mouse (red) neurons with 95% bootstrap confidence bands (apical dendrites).

## 4.3 Persistence Silhouette

In this section, we investigate the use of persistence silhouettes (defined in Section 2.6.3) to capture and compare topological variability in apical dendrites of mouse and human neurons. Building on the same persistence diagrams derived from the TMD framework, we compute silhouette functions using the Giotto-TDA library. These functions provide an interpretable and statistically tractable summary of the topological features encoded in each barcode.

The silhouette complements both entropy and landscape representations: while entropy captures global topological diversity and the landscape retains multiscale structure, the silhouette focuses on persistent features through a weighted averaging scheme. Its single-function format also allows for simpler visual and statistical comparisons across groups.

To analyze these representations, we adopt a two-step strategy. First, we visualize representative silhouette curves for each group by plotting the minimum, median, and maximum (based on the area under the silhouette function) for both human and mouse neurons on the same figure. This side-by-side comparison provides an intuitive overview of group-level differences and intra-group variability. These plots serve as a prelude to a more rigorous analysis.

Second, we replicate the bootstrap procedure as done in Section 4.2. The confidence band is constructed in the supremum norm and reflects the uncertainty of the mean function across the neuron population. We then compare the bands by superposing them, assessing whether they remain disjoint, intersect, or partially overlap, which helps us visually identify significant topological differences.

**Results**

The silhouette functions reveal notable differences between mouse and human neurons. In the min–median–max comparison in Figure 8, human apical dendrites exhibit taller and more widely distributed silhouette curves, particularly in the maximum case. In contrast, mouse neurons show more compact curves with lower overall amplitude. Since the silhouette is a weighted average of persistence diagram features, with greater weight assigned to features of higher persistence, the amplitude and shape of the silhouette reflect both the strength and distribution of topological features. A compact and low-amplitude silhouette, as observed in the mouse group, indicates that persistent features are consistently located in a narrow filtration range, suggesting structural homogeneity. Conversely, the taller and broader human silhouettes indicate the presence of strongly persistent features at varying filtration scales, pointing to greater morphological diversity. This interpretation is consistent with the theoretical framework presented in Berry et al. (2020), which emphasizes that persistence silhouettes summarize both the concentration and dispersion of topological information across the filtration axis. For clarity, we display only selected curves here; additional examples are provided in the Appendix (7.1).

These qualitative differences are reinforced by the analysis of confidence bands (Figure 9). The 95% confidence intervals for the two groups show limited overlap, especially around the first and second peaks, indicating statistically meaningful divergence in the underlying topological structures. While the human sample size is significantly larger ($n = 186$ vs $n = 38$ for mouse), the bootstrap method appropriately adjusts for this discrepancy, and the confidence bands remain robust to such imbalance.

In contrast to persistence landscapes, the silhouette highlights the contribution of high-persistence features more directly and suppresses the influence of short-lived structures. This makes it especially suited for uncovering strong, population-wide patterns. The results confirm and complement our earlier findings: mouse neurons exhibit more consistent topological motifs, while human neurons display a broader morphological diversity.

Together, these results support the use of silhouette functions as a valuable and interpretable descriptor of neuronal shape, capable of differentiating morphological phenotypes in a statistically meaningful way.

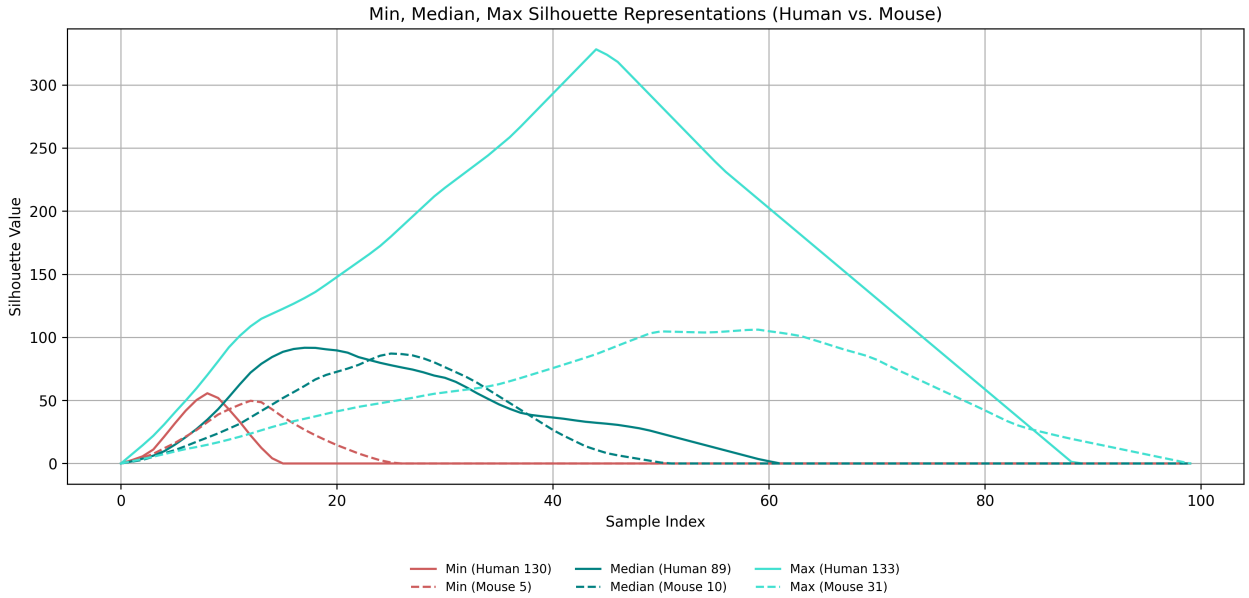



Figure 8: Representative persistence silhouette functions for apical dendrites. The plot compares the silhouette curves corresponding to the minimum, median, and maximum total silhouette area for both human and mouse neurons. Solid lines represent human neurons; dashed lines represent mouse neurons.

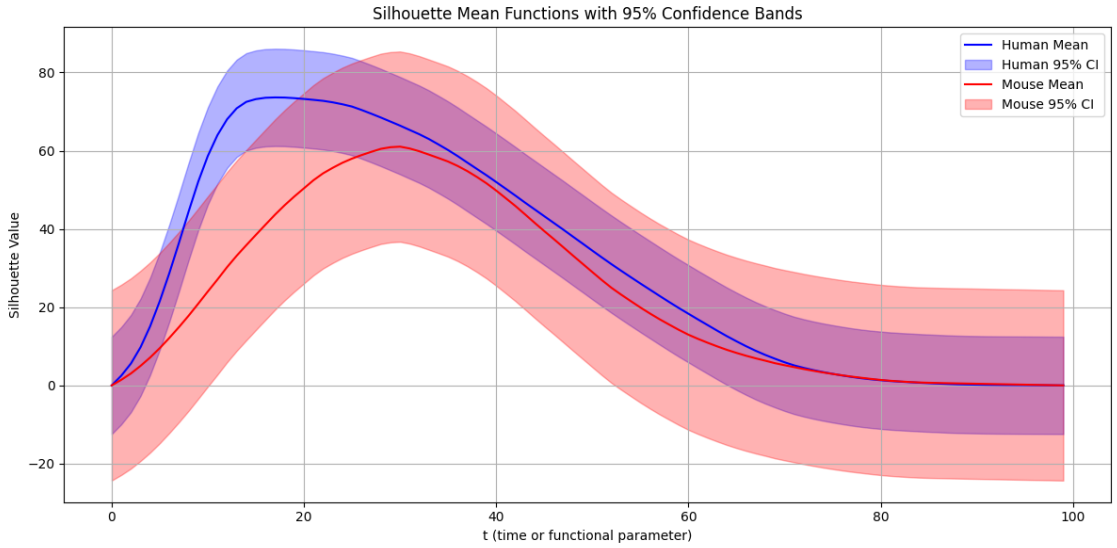



Figure 9: Mean persistence silhouette functions for human (blue) and mouse (red) neurons, with 95% bootstrap confidence bands (path distance, apical dendrites). Confidence bands reflect uncertainty in the sample mean and allow visual comparison between groups. The partial overlap of the bands suggests topological differences in dendritic complexity between species, despite differing sample sizes.

# Chapter 5: Statistical Assessment via Permutation Tests

While Section 4 revealed visible and measurable differences in topological descriptors across species, our third research question asks whether these observed differences are statistically robust or could result from random sampling. To answer this, we apply permutation tests tailored to each descriptor type – including the life entropy curve, persistence landscape, and persistent silhouette – in order to assess whether the inter-group differences are statistically significant.

The permutation test is a non-parametric approach that evaluates whether the difference in topological summaries between groups could reasonably have occurred by chance under the null hypothesis that both groups are drawn from the same distribution. This framework is particularly appropriate in our context, where standard assumptions of normality and homoscedasticity may not hold due to the high-dimensional and non-Gaussian structure of the descriptors. For each functional summary, we use the $L^\infty$ distance between the group-wise mean functions as the test statistic. By repeatedly shuffling the group labels (human vs. mouse) and recomputing the test statistic, we construct a null distribution that reflects the variability expected under the null hypothesis of no group effect. The $p$-value is then estimated as the proportion of permutations for which the permuted test statistic exceeds the observed one.

In addition to inter-group comparisons, we also perform intra-group permutation tests by splitting each species sample into two random halves. This control procedure estimates the baseline variability that arises within a homogeneous population. Comparing intra and inter-group test statistics allows us to verify that the observed group differences are not only statistically significant but also substantially greater than the natural variability found within a single species. This reinforces the robustness of our conclusions.

## 5.1 Inter-groups significance

All three functional descriptors yield statistically significant differences between human and mouse neurons, with $p$-values well below the 0.01 threshold. However, it is important to note that the observed test statistics operate on different numerical scales, reflecting the intrinsic properties of each descriptor. The persistence landscape encodes the lifetimes of topological features and typically takes higher values, often reaching above 100, which explains the larger test statistic observed. In contrast, persistent silhouette values represent weighted averages of persistence and tend to lie within a moderate range (e.g., 0–50), while life entropy curves are inherently bounded by the nature of entropy and rarely exceed 2.5. Consequently, the magnitude of the observed statistic should not be directly compared across descriptors. Instead, the $p$-value offers a normalized measure of significance that accounts for the range and

variability of each functional summary. This distinction is crucial for interpreting the results fairly and underscores the benefit of using permutation-based, distribution-free inference.

| **Descriptor** | **Observed Statistic** | $p$ **-value** |
|---|---|---|
| Life Entropy Curve | 0.597 | 0.0010 |
| Persistence Landscape | 132.00 | 0.0030 |
| Persistent Silhouette | 37.80 | 0.0010 |

Table 2: Permutation test results comparing human and mouse groups using three functional topological descriptors. The test statistic corresponds to the maximum pointwise difference ($L^\infty$ norm) between the group-wise mean curves. Number of permutation = 1000.

Figure 10 shows the null distributions generated by the permutation tests for each of the three functional topological descriptors. In each case, the observed $L^\infty$ test statistic (red dashed line) lies well beyond the range of the null distribution, indicating that the observed differences between the human and mouse groups are highly unlikely to arise by chance. This visual finding is consistent with the results reported in Table 2, where all $p$-values are well below the 0.01 significance threshold.

These results confirm that the differences observed in the average life entropy curves, persistence landscapes, and silhouette functions are not only visually apparent but also statistically robust. Despite the different numerical scales of the test statistics — a consequence of the inherent magnitude and units of each descriptor — the permutation-based framework allows for fair and consistent inference across all three modalities.

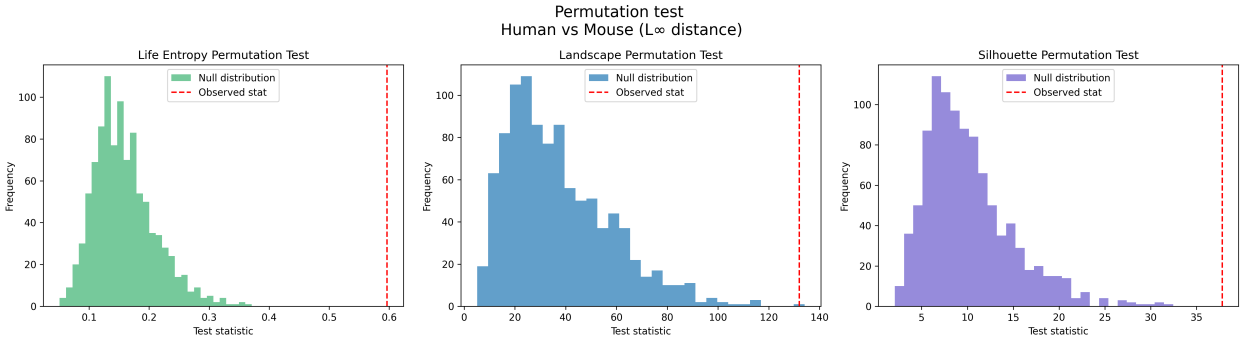



Figure 10: Permutation test (n=1000) null distributions for inter-group comparison (human vs mouse) using three functional descriptors. The red dashed lines mark the observed test statistics, computed as the $L^\infty$ distance between the group-wise mean curves.

## 5.2 Within-group controls

To ensure that the inter-group differences identified in the previous section are not due to random fluctuations or within-species variability, we conducted intra-group permutation tests for each functional descriptor. Specifically, for both human and mouse populations, we randomly split the sample into two equal halves and computed the distribution of $L^\infty$ distances between the resulting group means over multiple trials. These null distributions provide a baseline for the variability expected within a single population under no true group effect.

Figure 11 illustrates the results of these within-group comparisons for life entropy curves, persistence landscapes, and silhouette functions. In each subplot, the red dashed line indicates the observed inter-group test statistic from the human versus mouse comparison, while the histograms represent the null distributions for intra-group variability.

For both life entropy and persistence landscapes, the observed statistic lies well outside the support of both within-human and within-mouse distributions. This indicates excellent stability and strong discriminatory power: the differences observed between species are substantially greater than any natural variability observed within a species. Persistent silhouette descriptors also yield significant inter-group separation, though with a slightly narrower margin. The within-mouse distribution for silhouette statistics displays greater spread, occasionally reaching values near the observed inter-group level. This may reflect increased sensitivity to local morphological variability or higher noise in the silhouette summary.

Together, these results confirm that all three descriptors capture meaningful topological differences between human and mouse neurons, with persistence landscapes and life entropy curves showing particularly strong robustness to intra-group noise.

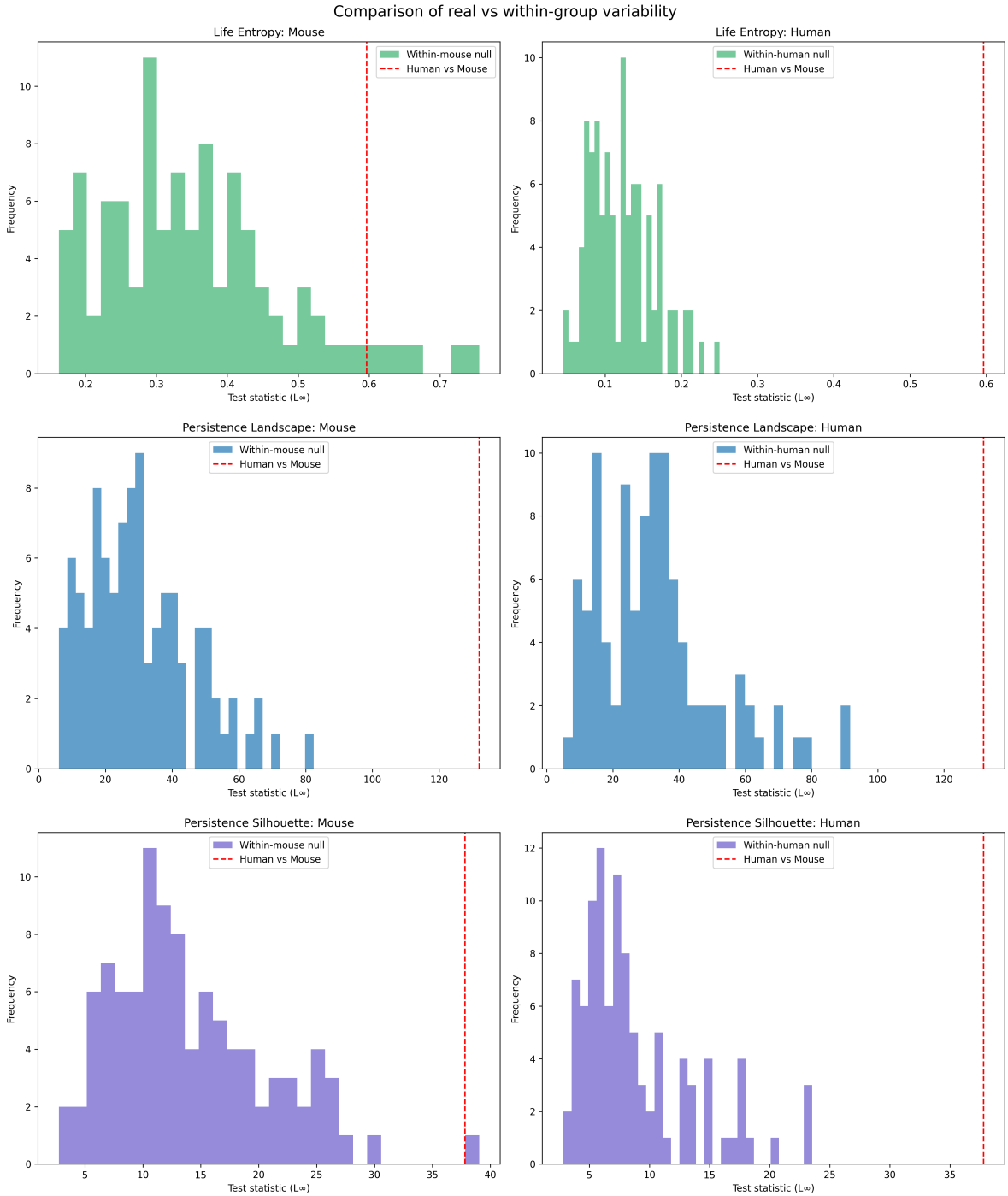



Figure 11: Within-group permutation test results for life entropy, persistence landscape, and silhouette descriptors. Each histogram shows the distribution of $L^\infty$ distances between mean curves computed from random splits of either the human or mouse group. Red dashed lines mark the observed inter-group test statistic (human vs mouse), revealing how it compares to natural within-group variability. Each histogram is based on $n = 100$ random intra-group splits.

# Chapter 6: Conclusion

This project set out to assess whether topological descriptors, in the form of functional summaries of persistence diagrams, can reveal meaningful and statistically robust differences in neuronal morphology between species. We focused on three descriptors: entropy life curve, persistence landscape, and persistence silhouette. They each offers a distinct trade-off between interpretability, complexity, and statistical power.

In response to our first research question, we found that human and mouse Layer 2/3 Pyramidal Cells do exhibit statistically significant differences in their topological structure. These differences are particularly pronounced in the apical dendrites, where persistent homology reveals greater structural complexity in human neurons. This confirms and extends prior findings by Kanari et al. (2025), using a novel statistical approach based on topological descriptors.

Our second objective was to determine whether functional summaries of persistence diagrams can capture these differences in a meaningful and interpretable way. Each summary has its own strengths and weaknesses, which we summarize below:

- **Entropy** provides a compact scalar summary of the disorder within a persistence diagram. While it is computationally efficient and easy to interpret, it loses substantial structural information and is not suitable for tasks requiring nuanced topological discrimination.
- **Persistence landscape** offers a rich functional representation in a Banach space, allowing the use of vector-space operations and statistical tools. It captures the shape of persistence diagrams effectively and showed the strongest discriminative power in our experiments, especially for distinguishing neuron classes with subtle morphological variations.
- **Persistence silhouette** also summarizes persistence diagrams as functions, but with a built-in weighting scheme based on persistence values. It performed comparably to landscapes, with slightly less stability across samples in our experiments.

We further addressed the robustness of these findings through permutation testing and confidence band analysis. These tests confirmed that the observed inter-species differences are not attributable to random sampling effects, despite the disparity in sample sizes. Within-group controls showed low variability, especially for the landscape-based summaries, indicating that our descriptors are both sensitive and stable.

Finally, we compared the effectiveness of the different descriptors in supporting statistical inference. The persistence landscape outperformed both the silhouette and entropy across multiple criteria, including discriminative power, stability across samples, and compatibility with statistical tools. This is summarized in Table 3.

| **Descriptor** | **Discriminative Power** | **Stability** | **Interpretability** |
|---|---|---|---|
| Persistence Entropy | Low | High | High |
| Persistence Landscape | High | High | Medium |
| Persistence Silhouette | Medium | Moderate | Medium |

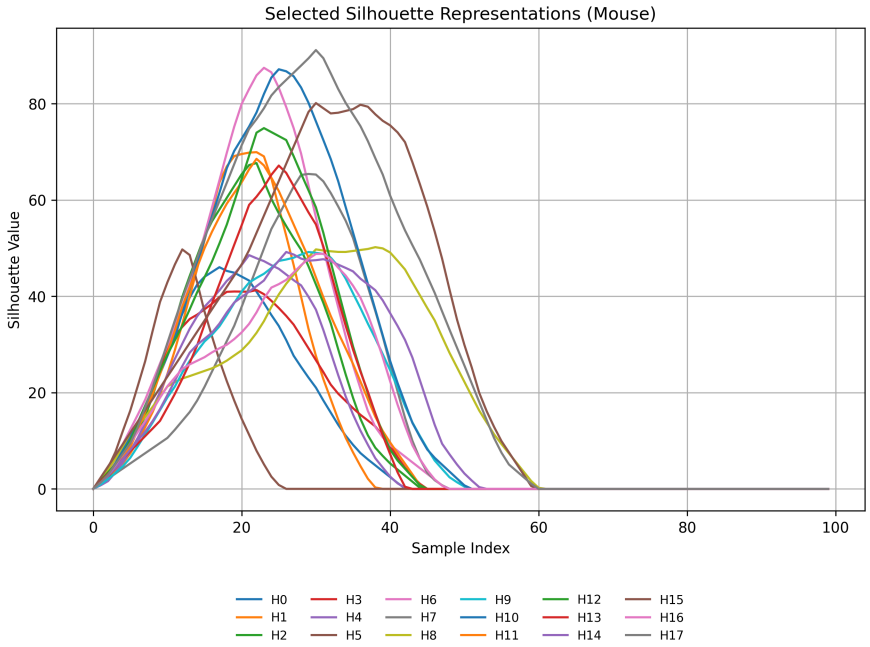
Table 3: Comparative summary of functional summaries of persistence diagrams.

To test the generality of these findings, we extended the analysis to basal dendrites and axons in the Appendix (7.2). The results for basal dendrites largely mirrored those observed in apical domains, confirming the broader applicability of the persistence landscape. For axons, however, statistical differences were weaker, likely due to smaller sample sizes and reconstruction inconsistencies.

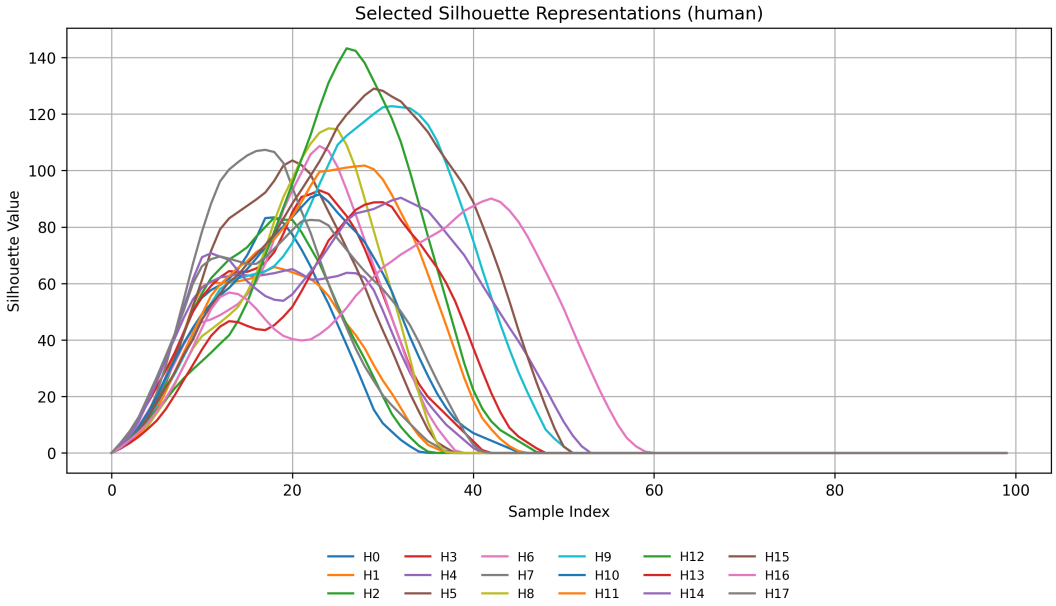
This work opens several promising directions for future research. In particular, given the strong performance of the persistence landscape, one natural extension would be to explore *generalized landscape functions*, as introduced in Berry et al. (2020). These functions allow for more flexible kernel choices and bandwidth tuning, potentially enabling even greater discriminative power or dimensionality reduction. Their ability to adapt more precisely to the distribution of features in the persistence diagram makes them a compelling candidate for further improving statistical analyses.

# Chapter 7: Appendices

## 7.1 Persistence Silhouette plots




((a)) Mouse dataset.




((b)) Human dataset.

Figure 12: Representative persistence silhouette functions for the L23 PCs apical dendrites in the mouse and human datasets. Each plot shows a subsample of 18 diagrams.

## 7.2 Persistence Landscape Analysis for basal dendrite and axons

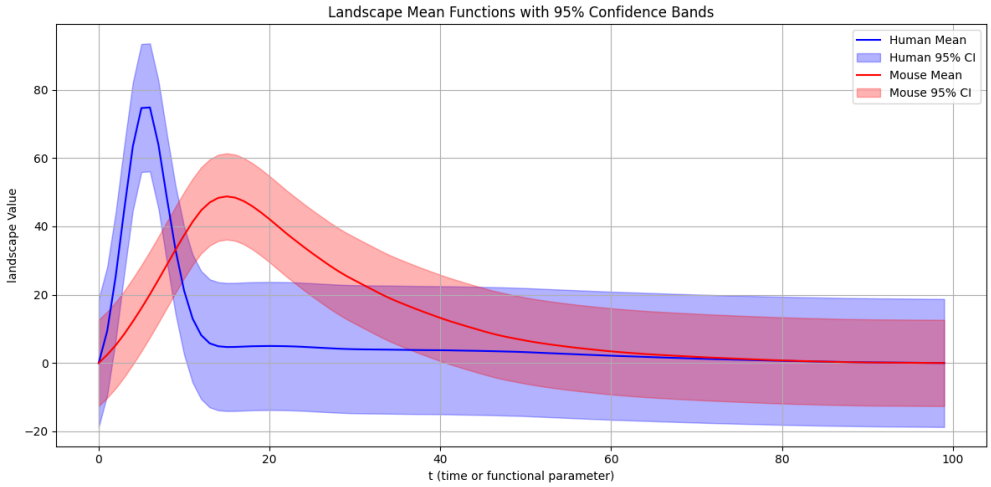
In the main body of this work, we focused on the apical dendrites of Layer 2/3 Pyramidal Cells to evaluate the discriminative power of different functional summaries of persistence diagrams. Among them, the persistence landscape proved to be the most robust and informative descriptor. In this Appendix, we extend the persistence landscape analysis to the other major morphological components of pyramidal neurons: the basal dendrites and the axons. This additional investigation aims to assess the generalization of our approach across different compartments of the same cell type.

Due to quality limitations in the available data, in the case of axons, we had to significantly reduce our sample sizes to ensure consistent data quality. Specifically, the axonal dataset was limited to 39 human neurons and 36 mouse neurons (compared to 186 and 38 respectively in the original dataset). This reduction reflects the greater variability and often lower reconstruction accuracy associated with axonal arbors. We apply the same analytical pipeline described in Section 4.2: computing persistence diagrams using TMD, transforming them into persistence landscapes via Giotto-TDA, constructing bootstrap confidence bands for each group, and performing two-sample permutation tests to quantify inter-species differences.

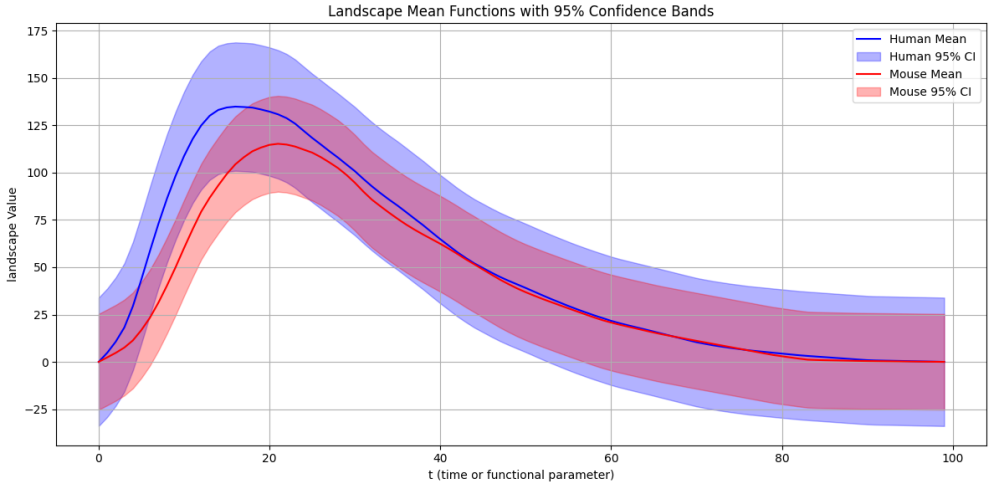
The results of the analysis are summarized in Figure 13. A statistical summary of the permutation test can be found in Table 4. For basal dendrites (Figure 13(a)), we observe a clear separation between human and mouse groups, both visually and statistically. The mean persistence landscape curves show distinct profiles, and their associated confidence bands exhibit minimal overlap. This suggests that persistence landscapes can reliably capture structural differences in basal dendrites between species, consistent with the results obtained for apical dendrites. The permutation test (Table 4) confirms this observation with a test statistic of 101.00 and a $p$-value of 0.003, indicating strong statistical significance.

For axons, however, the situation is more nuanced. While there is some variation in the mean persistence landscapes between human and mouse samples, the overlap of the bootstrap confidence bands is substantial (see Figure 13(b)). The permutation test yields a test statistic of 87.60, but the associated $p$-value is 0.328, indicating that the observed differences are not statistically significant at conventional thresholds (usually $< 0.01$).

This lack of statistical power can be attributed to two factors. First, the sample size for axons is considerably smaller than that of dendrites, limiting the reliability of statistical estimates. Second, axonal reconstructions tend to be more incomplete and heterogeneous, particularly in human data, which introduces additional noise and uncertainty into the topological summaries.




((a)) Basal dendrites.




((b)) Axons.

Figure 13: Mean persistence silhouette functions for human (blue) and mouse (red) neurons with 95% bootstrap confidence bands (path distance metric).

| **Dendrite Type** | **Observed Statistics** | $p$**-value** |
|---|---|---|
| Apical dendrites | 132.00 | 0.0030 |
| Basal dendrites | 101.00 | 0.0030 |
| Axons | 87.60 | 0.328 |

Table 4: Permutation test results comparing human and mouse groups using persistence landscape for three different neuron parts. The test statistic corresponds to the maximum pointwise difference ($L^\infty$ norm) between the group-wise mean curves. Number of permutation = 1000.

# References

Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., ... Ziegelmeier, L. (2017). *Persistence Images: A Stable Vector Representation of Persistent Homology* (Vol. 18; Tech. Rep.). Retrieved from http://jmlr.org/papers/v18/16-337.html.

Ali, D., Asaad, A., Jimenez, M. J., Nanda, V., Paluzo-Hidalgo, E., & Soriano-Trigueros, M. (2023, 12). A Survey of Vectorization Methods in Topological Data Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(12), 14069–14080. doi: 10.1109/TPAMI.2023.3308391

Atienza, N., Gonzalez-Díaz, R., & Soriano-Trigueros, M. (2020, 11). On the stability of persistent entropy and new summary functions for topological data analysis. *Pattern Recognition*, *107*. doi: 10.1016/j.patcog.2020.107509

Beers, D., Goniotaki, D., Hanger, D. P., Goriely, A., & Harrington, H. A. (2023). *Barcodes distinguish morphology of neuronal tauopathy* (Tech. Rep.).

Berry, E., Chen, Y. C., Cisewski-Kehe, J., & Fasy, B. T. (2020, 6). Functional summaries of persistence diagrams. *Journal of Applied and Computational Topology*, *4*(2), 211–262. doi: 10.1007/s41468-020-00048-w

Bubenik, P. (2015). *Statistical Topological Data Analysis using Persistence Landscapes* (Vol. 16; Tech. Rep.).

Carlsson, G. (2009). *TOPOLOGY AND DATA* (Vol. 46; Tech. Rep. No. 2).

Chintakunta, H., Gentimis, T., Gonzalez-Diaz, R., Jimenez, M. J., & Krim, H. (2015, 2). An entropy-based persistence barcode. *Pattern Recognition*, *48*(2), 391–401. doi: 10.1016/j.patcog.2014.06.023

Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. (2007). Stability of persistence diagrams. *Discrete and Computational Geometry*, *37*(1), 103–120. doi: 10.1007/s00454-006-1276-5

Edelsbrunner, H., & Morozov, D. (2010). *Persistent Homology: Theory and Practice* (Tech. Rep.).

Elston, G. N. (2002). *Cortical heterogeneity: Implications for visual processing and polysensory integration* (Vol. 31; Tech. Rep.).

Hatcher, A. (2002). *Preface* (Tech. Rep.).

Kanari, L., Dłotko, P., Scolamiero, M., Levi, R., Shillcock, J., Hess, K., & Markram, H. (2018, 1). A Topological Representation of Branching Neuronal Morphologies. *Neuroinformatics*, *16*(1), 3–13. doi: 10.1007/s12021-017-9341-1

Kanari, L., Ramaswamy, S., Shi, Y., Morand, S., Meystre, J., Perin, R., ... Markram, H. (2019, 4). Objective Morphological Classification of Neocortical Pyramidal Cells. *Cerebral Cortex*, *29*(4), 1719–1735. doi: 10.1093/cercor/bhy339

Kanari, L., Shi, Y., Arnaudon, A., Zulaica, N. B., Piccione, R. B., Coggan, J. S., ... de Kock, C. P. (2025, 6). Of mice and men: Dendritic architecture differentiates human from mouse neuronal networks. *iScience*, 112928. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S2589004225011897 doi: 10.1016/j.isci.2025.112928

Reimann, M. W., Nolte, M., Scolamiero, M., Turner, K., Perin, R., Chindemi, G., ... Markram, H. (2017, 6). Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in Computational Neuroscience*, *11*. doi: 10.3389/fncom.2017.00048

Robinson, A., & Turner, K. (2017, 12). Hypothesis testing for topological data analysis. *Journal of Applied and Computational Topology*, *1*(2), 241–261. doi: 10.1007/s41468-017-0008-7

Tauzin, G., Lupo, U., Tunstall, L., Burella Pérez, J., Caorsi, M., Medina-Mardones, A. M., ... Ch, K. H. (2021). *giotto-tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration* (Vol. 22; Tech. Rep.). Retrieved from https://github.com/giotto-ai/pyflagser .