**Investigation of different cell types through an analysis of persistent images of axonal trees**

Kerem Kurban and Leonardo Pollina

# **Introduction**

The field of neuroscience has seen significant advancements in dendrite analysis through the application of persistent image analysis [1]. However, there remains a noticeable gap in the study of axons [2]. The advent of new long-range datasets using viral tracing has made it possible to obtain high-quality morphology data, offering an opportunity to fill this gap [2,3,5,6].

The complexity of innervation patterns and the lack of readily available projection classification patterns present a challenge, as most neurons are simply labeled as 'projecting' [2]. This work aims to bridge this gap by identifying different classes of projecting axons and exploring the similarities and differences between various morphologies based on their brain region and expert annotation.

We employed dimensional reduction and clustering techniques on radial and path distance on ccf registered morphs. We also utilized region and cell type labels from Neuromorpho to determine if clusters have significant meaning [2]. For Granule Cells (GC) and Purkinje cells, we validated our clustering approach as these cell types exhibit specific morphological characteristics. We then used these clusters to map to specific morphometrics with the goal of identifying fine clusters.

This report also identifies gaps in the current literature and proposes future work using the automated morphology classification tools as the true labels for cell classes [4].

# **Methods**

## **Data Acquisition**

We fetched morphologies from Janelia and converted the JSON format to SWC. We then fetched metadata from Neuromorpho and extracted region information from soma position using the ccfv3 25um atlas. Cell type information was extracted from Neuromorpho metadata, which was not available in Janelia. We combined and cleaned similar cell types and filtered region and cell type classes for samples more than X and Y respectively.

## **Persistent Image Generation**

We used Topological Morphology Descriptor (TMD) to create Persistent Diagrams (PD) and Persistent Images (PI). Each PI was normalized by dividing it by the sum of the pixel values in the image. This means that the values of every PI sum to 1 like in a probabilistic map. Then, each PI was scaled by multiplying it by the number of components found in the PD (i.e., the number of bars or points in the persistent barcode or diagram, respectively) in order to take into consideration the differences in the number of branches. This allowed us to differentiate across morphologies with similar structures but with more or less branches. We then used these scaled-normalized images for further analysis. Also, all persistent images were scaled in order to lie on the same x- and y- axes.

## **Feature Extraction**

One of the main aspects in our analysis is to extract meaningful feature. Our choice is to focus on the persistent images, but these images are each one of a dimension of 100x100. We thus use t-distributed Stochastic Neighboring Embedding (t-SNE) as an unsupervised dimensionality reduction method in order to concentrate the information residing in the persistent images in just 2 components (that is 2 features). The choice of 2 features only is based on convenience also for illustration purposes. The perplexity value of the t-SNE method, one of the main hyperparameters, was fixed to the square root of the number of samples present in our dataset.

## **Clustering and optimization**

In order to identify potential clusters in the data projected on 2 dimensions after dimensionality reduction via t-SNE, we employed two different algorithms performing unsupervised clustering: k-means and GMM. This was due to the desire to counterbalance the pros and cons of these two methods. Indeed, while k-means can result faster and easier to interpret, it is also very dependent on the initialization of the centroids of the clusters. GMM is, on the hand, more robust to this given its probabilistic interpretation. However, both methods require a predefined number of clusters k, which was computed in a data-driven manner by using the elbow method or the silhouette method.

# **Results**

A picture containing screenshot, colorfulness, electric blue

Description automatically generated

**Figure 1:** Examples of persistent images for two different cells: A) a principal projection cell located in the motor area, B) a Purkinje cell located in the cerebellum. Images are normalized and scaled to the same axes. Each image has been multiplied by the number of components in the corresponding persistent diagram. On the top the persistent images obtained via the path distance filtration method are shown, while on the bottom the radial distance has been used.

The focus of our analysis is on the information contained on the persistent images when looking at the axonal trees. Here, we report a few examples of persistent images (after normalization and scaling) (**Figure 1**). Also, for illustration purposes, we report in **Figure 2A** an axogram, that is a dendogram of axons. Finally, we show an example of problematic nomenclature in the notation used in the dataset downloaded. Indeed, the cells in **Figure 2B** show dendrites with different morphological features, but they are all labeled as projecting granule principal cells.

A screenshot of a computer

Description automatically generated with medium confidence

**Figure 2:** Illustration example of a morphological dendrogram, that is the branching structure of an axon (**A**) and examples of dendritic structures of three neurons of our dataset being probably mislabeled as granule cells (note the difference between the two cells on the left and the one on the right) (**B**).

A picture containing diagram, line, text, map

Description automatically generated**Figure 3:** t-SNE projection on two dimensions of the persistent images obtained by radial distance (on the left) and path distance (on the right) filtration over the axonal trees. Each dot represents a neuron and colors are indicating different regions. Very peculiar regions such as the dental gyrus (DG) or the cerebellum show very clustered points.

As a first result, we could observe that specific regions corresponding to specific cell types (such as granule cells for the dentate gyrus or Purkinje cells for the cerebellum) were very well clustered when looking at the scatter plot after dimensionality reduction via tSNE. This is shown in **Figures 3,4**. This indicates that these types of cells can still be very well classified when using the persistent images of the axonal trees and that the dimensionality reduction method is useful and accurate. On the other side, we can observe that other types of cells, especially the ones referred to as pyramidal projection cells or simply projection cells are very spread (even considering their large number).

A picture containing diagram, text, map

Description automatically generated

**Figure 4:** t-SNE projection on two dimensions of the persistent images obtained by radial distance (on the left) and path distance (on the right) filtration over the axonal trees. Each dot represents a neuron and colors are indicating different morphology types. Very peculiar types such as Purkinje cells and granule cells show very clustered points.

Following this observation, we decided to focus on only this population of cells and try to understand if cell sub-types could be identified. The tSNE 2-dimensional projection of only these cells can be observed in **Figure 5**.

**A picture containing diagram, line, text

Description automatically generated**

**Figure 5:** t-SNE projection on two dimensions of the persistent images obtained by radial distance (on the left) and path distance (on the right) filtration over the axonal trees of the neurons labels as simply projection cell or pyramidal cells. Each dot represents a neuron.

With the goal of identifying possible sub-clusters, we employed k-means as an unsupervised clustering method. To optimize and find the number of clusters in a data-driven way, we used the so-named “elbow method”, which consists in evaluating the kmeans algorithm via the sum of squared errors (SSE) for different numbers of cluster. The value of k for which there is no more a significant decrease in SSE is then chosen as the optimal number of clusters to use. In our case, this number corresponded to k = 3. The results from the elbow method are shown in **Figure 6A**. Also, to remove stochasticity, the SSE value for each k was computed by repeating the procedure 20 times. The final average value was then used.

Taking into consideration the potential limitations of k-means such as the constraint of spherical clusters, we decided to double check the clustering approach by using a more flexible algorithm like Gaussian Mixture Models (GMM). Also, in this case we tried to find in a data-driven fashion the optimal number of clusters. We employed the Silhouette method and our results pointed again towards k = 3. This can be observed in **Figure 6B**.

A picture containing line, plot, diagram, screenshot

Description automatically generated

**Figure 6:** Optimization methods employed to find the suitable number of clusters: the Elbow method for k-means in **A**, and the Silhouette score optimization for GMM in **B**. The two curves show the results obtained when using the radial distance or the path distance filtration methods. A number of clusters equal to k=3 seems to be optimal in both cases.

**A picture containing text, screenshot, map, pattern

Description automatically generated**

**Figure 7:** The 3 clusters identified by the two different clustering methods: k-means in **A**, GMM in **B**. Each dot represents a different neuron, and the points are color coded depending on the cluster they have been assigned to. The results for both radial distance and path distance are presented.

In **Figure 7**, we show that the clusters identified from k-means and GMM are practically identical, which reinforces the idea of a true presence of clusters, considering the easy interpretation of k-means and the more robust results of GMM thanks to its minor dependency on initial conditions.

All the procedure was performed by employed as features either the data projected coming from the radial distance filtration method or from the path distance filtration method. Indeed, it is not clear if these two would give similar converging results when using them to

investigate axonal trees. By comparing the percentage of overlap of the clusters obtained via the two different metrics, we can see in **Figure 8** that these are indeed largely overlapping.

A picture containing text, screenshot, number, font

Description automatically generated

**Figure 8:** The intersection (overlapping) between the clusters identified through radial or path distance for k-means (**A**) and GMM (**B**). The high values of overlapping show how the two filtration methods tend to converge towards similar results.

# **Discussion and Future Directions**

Our preliminary findings offer promising insights into the classification of projecting axons, demonstrating the potential of our approach. The identification of distinct clusters based on region and cell type suggests that our methodology can yield meaningful insights into the morphological diversity of axons. However, the biological significance of these clusters and their validation require further investigation.

For this reason, if time will allow, we plan to now compute an analysis over the morphometric variables extracted for the three clusters identified via our method. We plan to investigate whether we can find indeed significant differences for some of the metrics, which would further validate the existence of at least three morphologically different axons.

Moreover, one of the encouraging and reassuring results we obtained is the convergence of results between the persistent images computed via the radial distance filtration method and the path distance filtration method. Indeed, while these two methods are known to yield similar results when analyzing dendritic structures, this result is not yet clear when focusing on axonal trees.

One of the challenges we encountered in our study was the sparse labeling of cell types in the metadata fetched from Neuromorpho for the Janelia dataset. This highlights the need for more comprehensive and consistent labeling in neuronal morphology datasets to facilitate more detailed and accurate analyses.

Looking ahead, we plan to extend our axonal approach to the Morphoclass tool developed by the Blue Brain Project. This tool uses machine learning to classify neurons based on their dendritic morphological features. By applying our axonal approach to these 'true' labels, we can further validate our methodology and refine our classification of projecting axons.

Furthermore, still speculating about the future, we could think of incorporate more long-projecting neuron reconstructions into our dataset. By combining data from different labs, such as the Allen Institute [4,5], we can improve the generalizability of our findings and contribute to a more comprehensive understanding of axon morphology.

In conclusion, our work represents a step towards a more detailed understanding of axon morphology and classification. We believe that our approach, combined with the use of new tools and datasets, could significantly contribute to the advancement of neuroscience research. Our future work will focus on refining our methodology, expanding our dataset, and exploring new ways to classify neurons based on their axonal morphology.

# **References**

1. Lida Kanari, Srikanth Ramaswamy, Ying Shi, Sebastien Morand, Julie Meystre, Rodrigo Perin, Marwan Abdellah, Yun Wang, Kathryn Hess, Henry Markram, Objective Morphological Classification of Neocortical Pyramidal Cells, Cerebral Cortex, Volume 29, Issue 4, April 2019, Pages 1719-1735

2. Ascoli, G. A., Donohue, D. E., & Halavi, M. (2007). NeuroMorpho. Org: a central resource for neuronal morphologies. Journal of Neuroscience, 27(35), 9247-9251.Chicago

3. Gerfen, C. R., Economo, M. N., & Chandrashekar, J. (2016). Long distance projections of cortical pyramidal neurons. In Journal of Neuroscience Research (Vol. 96, Issue 9, pp. 1467–1475). Wiley. <https://doi.org/10.1002/jnr.23978>

4. [https://github.com/BlueBrain/morphoclass](https://github.com/BlueBrain/morphoclass)

5. Muñoz-Castañeda, R., Zingg, B., Matho, K. S., Chen, X., Wang, Q., Foster, N. N., ... & Dong, H. W. (2021). Cellular anatomy of the mouse primary motor cortex. Nature, 598(7879), 159-166.

6. Wang, Y., Xie, P., Gong, H., Zhou, Z., Kuang, X., Wang, Y., ... & Veldman, M. B. (2019). Complete single neuron reconstruction reveals morphological diversity in molecularly defined claustral and cortical neuron types. BioRxiv, 675280.

7. Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11).Chicago