



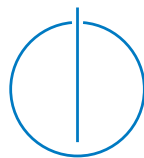
DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

# **Tracing and Analysis of Microglia**

Christoph Dehner





DEPARTMENT OF INFORMATICS

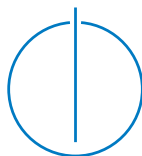
TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

# **Tracing and Analysis of Microglia**

## **Tracing und Analyse von Microglia**

Author:	Christoph Dehner
Supervisor:	Prof. Dr. Björn Menze
Advisor:	Markus Rempfler, Dr. Vasileios Zografos
Submission Date:	15.01.2018



I confirm that this master's thesis in informatics is my own work and I have documented all sources and material used.

Munich, 15.01.2018

Christoph Dehner

## Acknowledgments

Foremost, my sincere thanks go to Prof. Dr. Björn Menze of the Image-Based Biomedical Modeling Group at Technische Universität München and Dr. Ali Ertürk of the Acute Brain Injury Research Group at the Institute for Stroke and Dementia Research, Ludwig-Maximilian-Universität München for giving me the opportunity to write this thesis and for their support.

Likewise, I would like to express my sincere gratitude to Dr. Vasileios Zografos for the continuous support of my research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

Furthermore, I would like to acknowledge Leander Mrowka and Marika Ruiyao Cai for conducting all biological experiments and imaging-related tasks for this thesis. Their data and passionate participation laid the foundation for my work.

I would also like to thank Markus Rempfler for his input and valuable comments on this thesis.



# Abstract

Analyzing microglial morphology can reveal information about mechanisms in the brain e.g. during injury, neurodegeneration or aging. However, this is challenging as manually tracing microglia from scans is error-prone and tedious.

In this thesis we introduce a new approach to automatically extract microglia from volumetric confocal microscopy scans. First we first segment somas, then we extract processes and finally we combine the obtained structures complete cells. By visual evaluation our approach provides a new level of accuracy for extracted microglia.

Applying our tracing framework to scans of mouse brain tissue, we extract nearly 3000 microglia cells and analyze their morphology in classification, clustering and regression experiments. In various ways we quantify the morphology of microglia in different activation states and show a continuous morphological transition between the microglia cells close to an injury site and those further away. In this way we demonstrate that microglial morphology can be used to predict injury severity.

# Contents

<b>Acknowledgments</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Contributions . . . . .	2
1.2 Outline . . . . .	3
<b>2 Related work</b>	<b>4</b>
2.1 Soma segmentation . . . . .	4
2.2 Process extraction . . . . .	5
2.2.1 Measuring tubularity . . . . .	5
2.2.2 Tracing . . . . .	6
2.3 Feature extraction and Analysis . . . . .	7
<b>3 Methodology</b>	<b>8</b>
3.1 Soma segmentation . . . . .	8
3.2 Process extraction . . . . .	9
3.2.1 Centerline extraction . . . . .	11
3.2.2 Partitioning . . . . .	13
3.2.3 Pruning . . . . .	16
3.3 Feature extraction . . . . .	16
3.4 Analysis . . . . .	20
<b>4 Results</b>	<b>22</b>
4.1 Datasets . . . . .	22
4.2 Evaluation . . . . .	25
4.2.1 Soma segmentation . . . . .	25
4.2.2 Centerline extraction . . . . .	27
4.2.3 Partitioning and Pruning . . . . .	28
4.3 Classification . . . . .	29
4.4 Clustering . . . . .	35

## *Contents*

---

4.5 Regression . . . . .	38
<b>5 Conclusion and Outlook</b>	<b>41</b>
<b>Bibliography</b>	<b>43</b>

# 1 Introduction

Microglia are immune cells in the central nervous system and are responsible for a broad range of maintaining and protective cellular responses. Microglia typically consist of one soma in the center of the cell and tree-shaped processes emanating from the soma. The primary purpose of these processes is to scan the surrounding tissue. During development, homeostasis and disease microglia appear in altering phenotypes, something which is also linked to their functionality [CB17]. For example, in healthy tissue the cells usually have small somas and long, thin processes. Conversely, activated cells around an injury are typically characterized by enlarged somas and thick, but very short processes. Beyond these common cases, studies have observed a range of in-between phenotypes of microglia for different circumstances and injury states [CB17]. The left image of Fig. 1.1 depicts for example a scan of microglia which are neither completely ramified nor fully activated.

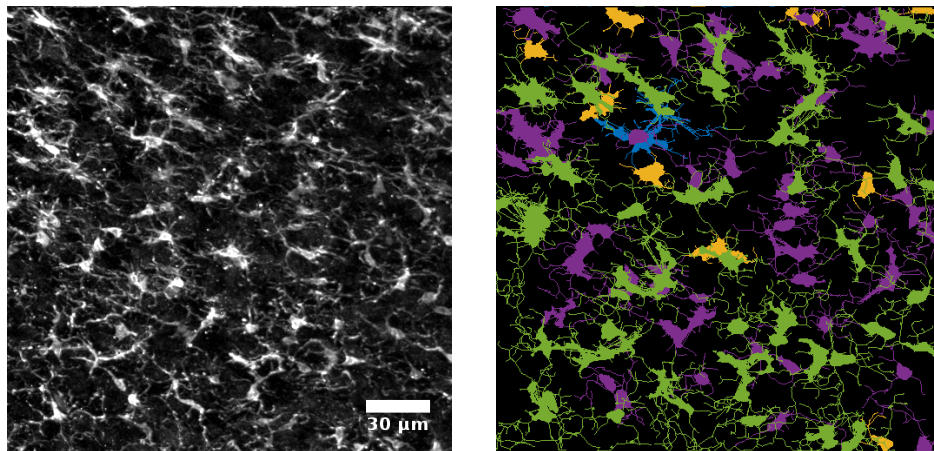


Figure 1.1: Examples for microglia in different shapes (as max-projection in the left image) and corresponding morphological classes (color-encoded in the right image) that are found by clustering a large amount of automatically extracted microglia.

Thus, analyzing microglial morphology can reveal information about mechanisms in the brain e.g. during injury, neurodegeneration or aging. However, such studies are very challenging for the following two reasons: First, microglial outlines are very difficult to extract from scans, since they appear in highly complex and varying shapes. Even for a human expert it is hardly possible to identify unambiguously all microglial processes from a scan. Second, a large amount of cells is needed to do comprehensive and statistically significant analysis on microglial phenotypes. Generating a sufficiently sized dataset for such analysis by manual tracing is very tedious and time consuming.

Both of the above reasons motivate us to automate the extraction and analysis of microglial cells. A computer can process multidimensional data more easily and is able to capture more complex characteristics than a human. Furthermore it is capable to scale up to large amounts of input data in a straightforward way. The right image of Fig. 1.1 for example shows morphological classes of microglia that are obtained by clustering a large amount of automatically extracted cells.

Existing approaches for automatically extracting microglia from volume scans are limited in the quality of the images they work with and in the accuracy of the computer-vision algorithms that they use [Din+17; Xu+16; Meg+15]. In this thesis we address those limitations and introduce a new approach to extract microglial morphology from confocal scans of mouse brain tissue in an automated way. Subsequently, we demonstrate that microglia cells which are extracted by our tracing framework are applicable to a broad range of analyses.

## 1.1 Contributions

In the first part of the thesis we combine various state-of-the-art computer vision algorithms to segment and trace microglia from volumetric confocal microscopy scans. In the second part we apply our tracing framework to scans of microglia in mouse brains and analyze the morphology of extracted cells. The scanned tissue is treated by the uDISCO tissue clearing protocol [Pan+16], a new and highly effective tissue clearing procedure. More precisely, we contribute in the following way to the research about microglia:

- We develop an accurate method to segment the somas of microglia from volumetric images, building upon a method for neurons from [KL16].
- We extract process outlines by regression similar to [Sir+16]. Following our methodology, ground truth training data is automatically deduced from input images. Thus, microglia do not have to be traced manually.

- We obtain extractions of individual microglia cells by combining the results of the previous steps and refining them based on an approach from [Tür+16].
- In classification experiments we analyze the morphological classes of "ramified", "partially activated" and "fully activated" microglia.
- Applying clustering techniques to extracted cells we detect new morphological classes of microglia.
- In regression experiments we use microglial morphology as a predictor of injury severity by relating it in a continuous way to the distance to an injury site.

## 1.2 Outline

Fig. 1.2 illustrates the main steps of our microglia tracing and analysis framework. Capturing the shape of microglia from 3D scans is split up into two consecutive tasks. Firstly, somas are segmented from the data. Secondly, we extract the process outlines from the scans and combine the results to complete extractions of microglia cells. The

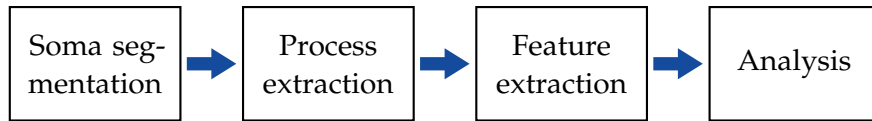


Figure 1.2: Outline of the main steps in this thesis to trace and analyze microglia. The first two steps deal with extracting microglia from volumetric scans. Subsequently, we analyze the morphology of extracted cells in the remaining two steps.

successive analysis of microglial morphology also consists of two parts. Firstly, features are extracted from the traced cells to capture their shapes numerically. Secondly, the obtained data is analyzed in classification, clustering and regression experiments.

The ordering of Fig. 1.2 is followed in all subsequent chapters of the thesis. In the next part we give an overview about related work. Following that, in Chapter 3 we explain our methodology in detail. Concrete experiments and results are presented in Chapter 4. Finally, we conclude the thesis with a summary and an outlook about possible follow-up projects in the last Chapter 5.

## 2 Related work

Previous work on extracting biological structures from brain scans mainly focuses on neurons and vessels. However, since microglia and neurons have similar shapes and involve tasks like segmenting tubular structures and tracing tree-shaped cells, existing approaches can be also adopted for this purpose. This chapter gives a summary of how the topics of soma segmentation, process extraction, feature extraction and analyzing microglia are addressed in literature.

### 2.1 Soma segmentation

The purpose of soma segmentation is to distinguish soma voxels from the background or other structures in the scan. In the following, two recent approaches originally designed for neurons are presented. [Ozc+15] processes two-dimensional neuron images obtained by projecting three-dimensional scans into the x-y-plane. In a first step, a support vector machine is used to segment neuronal structures from the background. Then, a measure of isotropy called directional ratio is evaluated to distinguish blob-shaped somas from thinner processes. This method works really well for voxels within a soma, but fails to accurately grasp the soma border. Therefore, in a third step the initial soma localizations are refined by the level set method [OF02] to better fit to the actual soma edges. In [KL16], this approach is improved. In particular, the orientable filters used within the directional ratio are reformulated, enabling a fast implementation by separable convolution.

The most recent approach in literature for soma segmentation uses deep learning. In [Wan+17] a cascaded convolutional neural network is trained to segment mouse brain nuclei. An initial network learns a first probability mapping between the image and the intended cell segmentations. The output from the network is used to generate contextual features. A second convolutional neural network learns the final segmentations from both the image signal and the generated contextual features. By combining image and contextual features, the learning method can detect varying object shapes and is robust against inhomogeneities in the data, e.g. introduced by increased noise in confocal

scans on deeper z-levels.

## 2.2 Process extraction

A further central challenge for analyzing the morphology of microglia is to accurately extract their processes. As the spatial outline of processes is critical for the morphology, process extraction aims at capturing their progression instead of segmenting the branches voxel by voxel. Incorporating that, an extracted process is modeled as a tubular structure. It is represented as a centerline proceeding along the middle of a branch and a radius for each point on the centerline describing the width.

In related literature, process extraction is usually accomplished in two steps. First, the appearances of processes are captured by a tubularity score. Such score values are calculated for all voxels in a scan and indicate the spatial progression of centerlines. Second, a network of centerlines is traced from a calculated tubularity image. In combination with the corresponding soma localizations, the extracted structures represent reconstructions of complete cells.

### 2.2.1 Measuring tubularity

State of the art methods to measure tubularity can be classified into two techniques. One option is to use hand-crafted, often Hessian-based filters such as the Frangi filter [Fra+98] or the multi-directional oriented flux (MDOF) filter [Tür+13]. These filters give tubularity measures for all voxels of a scan and produce maximal responses on the respective centerlines of vessels or cell processes. Since they are designed for ideally round tubes, they work only to a limited extent for irregular structures and give weaker responses at bifurcations. A big advantage of such filters is that they can be applied to any data without the need of preprocessing. Taking only local image information, they require few computational resources and can be parallelized easily.

A second and more recent technique is to learn the characteristics of relevant tubular structures from ground-truth data. Most of the successful learning approaches formulate centerline extraction as a regression problem and outperform filter methods. In [DH16] vessel segmentation is tackled amongst others using structured regression. Since tubular structures are not explicitly modeled as centerlines, this is however not directly applicable in this thesis. [Sir+16] trains regressors on multiple scales to learn the inverse distance transform to the closest centerline using gradient boosted trees. The output of the regressors can then be interpreted as a tubularity score. Applied to a test



image, centerlines appear as maximal ridges, which can be extracted by non-maximum suppression. Both filter and regression approaches can be applied to multiple scales and thereby allow us to capture the information of processes with different radii.

### 2.2.2 Tracing

As the second part of process extraction, a network of centerlines is traced from the obtained tubularity measures. Whereas finding the appearances of processes in the previous step operates independently of intended cell structures, shape constraints as for example a tree structure are enforced in this tracing step.

The review paper [ASI16] gives an overview of automated state-of-the-art tracing techniques for neurons. Different methods vary considerably with respect to the algorithms that they deploy and the information they incorporate into their decision making. Related algorithms can be classified into local and global approaches. Local methods follow greedy strategies and grow processes or cell parts step by step using only local image information. As a consequence, they allow on the one hand fast and efficient implementations, but on the other hand are very sensitive to errors caused by noise and discontinuities in the data.

Global methods usually yield more robust solutions. They aim at finding an optimal result for a whole cell or volume simultaneously. Tracing is formulated on a more abstract level as a global optimization task. Frequent techniques in literature are Markov Chain Monte Carlo algorithms or graph-based approaches like finding the minimum spanning or shortest path tree.

A very innovative approach for obtaining an accurate centerline network from a tubularity image is presented in [Tür+16]. After extracting an over-complete graph from the tubularity responses, an optimal subgraph is found by integer programming. In this sense, the final centerline network is obtained by pruning an initially over-complete solution. The information about which parts of the graph best represent the underlying structure and which ones should be pruned is encoded in weights. These weights can be defined not only on single edges, but also on longer paths in the graph. The underlying probabilistic model allows us to incorporate structural constraints in a very flexible way. When solving the integer program, a weighted sum of the introduced weights is optimized while enforcing the defined constraints. In summary, the introduced method provides an advanced formulation of tracing and outperforms other state-of-the-art approaches on several public datasets. This integer linear programming (ILP) formulation has also been applied for vascular networks [Rem+15], vascular labeling [Rob+16] and artery-vein separation [Pay+16].

[RAM16] investigates various formulations of the connectedness constraint in such an ILP framework and their scalability. [MTF17] integrates the approach into an active learning framework.

### 2.3 Feature extraction and Analysis

Features are deduced from extracted cells to capture their morphology numerically. In related literature, microglial features are defined in several ways. For example, different arbor properties are quantified, the spatial extent of the cells is measured and fractal scores of the extracted cells are considered.

Studies on a large number of automatically extracted cells in literature involve either classification or clustering techniques. While classification approaches aim at analyzing and predicting known phenotypes of microglia, clustering algorithms are used to detect unknown structures in the data and find new morphological classes of microglia. In the following, we give a summary of three recent and state-of-the-art publications performing analyses on microglial morphology based on automatically extracted cells.

[Xu+16] compares confocal scans of healthy and injured rat brain tissue. 136 features are defined and calculated for approximately 7700 traced microglia cells. Most of the features are L-measures [SPA08] and describe the morphology of a cell on different levels. Furthermore convex hull features and measurement from ellipsoids, fitted to the cells are added. The obtained data is clustered using an infinite Gaussian mixture model [Ras00]. In this way, six distinct phenotypes of microglia are extracted. Mapping the found clusters back to the tissue scans reveals insights about the spatial distribution of different phenotypes in healthy and injured tissue.

In [Meg+15], a similar approach is applied, though using a harmonic co-clustering algorithm and extracting four different shape appearances. In [Din+17], microglia are extracted from healthy and Alzheimer’s mouse brains in the following four activation states: Resting, partially ramified, slightly ramified and activated. A support vector machine with a Gaussian Radial Basis Function kernel is used to classify the extracted cells. Hereby, the authors do not only consider fractal and Fourier features that are deduced from the cell extractions, but also include Gabor features, which they calculate directly from the grayscale images. Classification accuracies of around 94% are achieved by fractal features. The grayscale features could not capture any information about the different morphologies of the cells.

## 3 Methodology

In this chapter we explain the methodology for all tasks in this thesis, following the outline of Fig. 1.2.

Most accurate methods for soma segmentation and process extraction in literature rely on machine learning techniques. However, due to highly complex and varying appearances of microglial structures in scans, it is not feasible to extract manually sufficiently accurate ground truth data. Even a human expert cannot unambiguously identify all microglial soma edges and processes from a scan.

In this thesis we solve this problem as follows: For soma segmentation we apply a filter approach and do not require training data. For centerline extraction we use ground truth data that is generated from filter responses. We describe our method in the remainder of this chapter.

### 3.1 Soma segmentation

Somas usually appear as spherical and intensity-based dense blobs in the volumetric confocal scans. Therefore, our core strategy for soma segmentation is to detect isotropic and dense objects. All involved steps of the presented approach are outlined in Fig. 3.1.

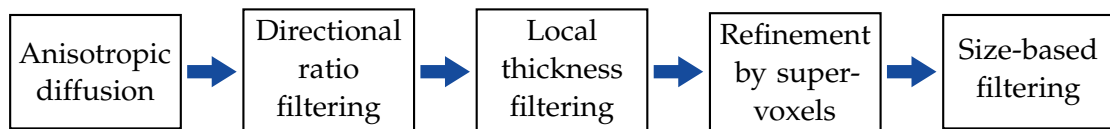


Figure 3.1: Outline of the main steps for segmenting somas. The presented approach bases on filters that detect isotropic and intensity-based dense objects.

In a first step, we preprocess input images by anisotropic diffusion [KSM10] to remove noise. Then, directional ratio filtering from [KL16] is applied to localize somas and suppress processes. This is a state-of-the-art technique for isotropy-based detection and works well in related literature [Ozc+15]. The directional ratio is defined as the

quotient of two Shearlet transformations [KL12] and provides for each voxel a measure of isotropy between 0 and 1. By thresholding the acquired isotropy image, elongated and anisotropic processes can be filtered out from more ellipsoidal shaped somas with larger directional ratio scores.

Subsequently, these initial, yet over-complete and noisy detections are improved further by considering density. Local thickness scanning [HR97] measures for each pixel in a binary mask the maximal radius of a sphere, which contains the pixel and covers only non-zero grid points. Applying the local thickness algorithms to the directional ratio output from the previous step yields density measurements for all voxels in the scan and allows to distinguish between dense structures, which are more likely to be somas, and thin processes. The concrete decision of what is considered to be a soma is again made by an appropriately chosen threshold.

In a last step, the detections are refined again. Supervoxels [HD14] adhere to the locally most dominant image edges and can thus be used to fit the segmentations to the borders of the underlying soma signal more precisely. For that, supervoxels on multiple granularity levels are calculated. Likelihood scores for belonging to a soma are obtained for all supervoxels by their overlaps with the soma segmentations so far. Averaging the scores of different supervoxel levels yields a combined likelihood image. Thresholding this image reveals the final segmentation.

In combination with size-based filtering at the end to avoid small false detections, the explained steps yield the soma locations of a scan. Fig. 3.2 illustrates the soma segmentation algorithm with intermediate results for the directional ratio and local thickness measurements.

A huge advantage of the presented method is that it does not require ground truth data and can be applied to any new scans instantly. On the other hand, the drawback of such a non-learning method is that parameters have to be adjusted manually for changing data. For example, all three thresholds mentioned above have to be set to appropriate values depending on the noise level and the magnification of the scan. However, if the parameters are adjusted for a specific imaging scenario, they are robust against varying cell shapes in the data and work for activated as well as non-activated cells.

## **3.2 Process extraction**

The goal of process extraction is to capture the spatial outlines of processes in a scan and to combine these outlines with the soma segmentations to extractions of complete microglia cells. The presented methodology consists of three parts. First, we capture

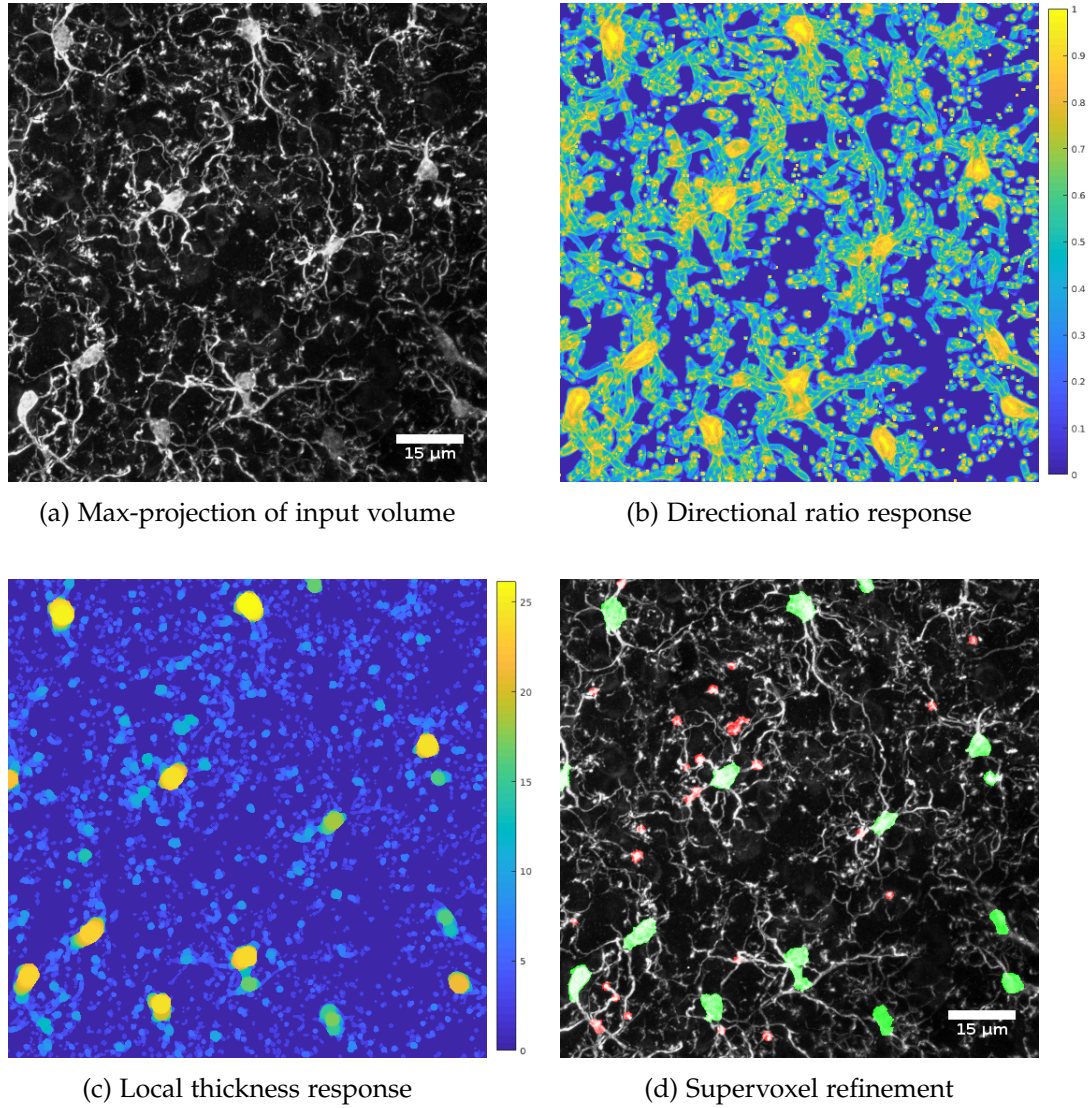


Figure 3.2: Main steps for segmenting somas: a) Max-projected confocal scan of microglia. b) Directional ratio response that measures isotropy in the scan. c) Local thickness response of the thresholded directional ratio output to detect intensity-based dense objects. d) Final segmentation (in green) obtained by thresholding the result from the local thickness algorithm, refining the detection with supervoxels and filtering small detections (in red). The original scan is underlaid for a visual evaluation of the results.

the process appearances during centerline extraction. Second, we partition the found structures into separate cells. Third, we refine these cells and remove spurious parts from the extracted centerlines by pruning.

### 3.2.1 Centerline extraction

Extracting the centerlines of microglial processes is a challenging task. Highly complex shapes, varying appearances of different activation states and weak signal to noise ratios make it difficult even for human experts to identify unambiguously all branches in a scan. The most accurate automatic detection frameworks for tubular structures are based on learning and require ground truth data. However, as explained previously such training data cannot be obtained manually for microglia. We overcome this obstacle by introducing a framework in which centerlines obtained from filters are used as ground truth. An overview of the presented method for centerline extraction is illustrated in Fig. 3.3.

Following the approach from [Sir+16], we formulate centerline extraction as a multi-scale regression task. For each scale, we train a regressor that outputs the inverse distance transform to the closest centerline. The required multi-scale ground truth centerlines to train the regressors are obtained by multi-directional oriented flux (MDOF) [Tür+13] and Frangi [Fra+98] filters. After these filters are applied to prospective training images, ground truth centerlines are extracted by binarizing and skeletonizing [LKC94] the filter responses. The radius of each centerline voxel is specified by the scale of the corresponding maximal filter response.

The presented way of generating ground truth can be interpreted as learning a combined generalization of various filters. Due to irregular shapes and noise, a hand-crafted filter working on local image information cannot capture all tubular structures, but returns an incomplete subset of all branches. By feeding the obtained centerline fragments into the regressor, it can incorporate information from multiple image locations and learn more general characteristics of tubular structures in the data. Since different filters (in the concrete case MDOF and Frangi filters) capture different branch fragments, it improves the performance to incorporate multiple of them into the framework.

Even though we obtain incomplete ground truth data, we have enough training images to compensate missing parts. Rather than having complete training data, the precision of (possibly incomplete) ground truth centerlines is crucial for the accuracy of the presented centerline extraction approach. The regressors can only learn the correct appearances of relevant structures, if the found ground truth data precisely matches with true centerlines in the scans.

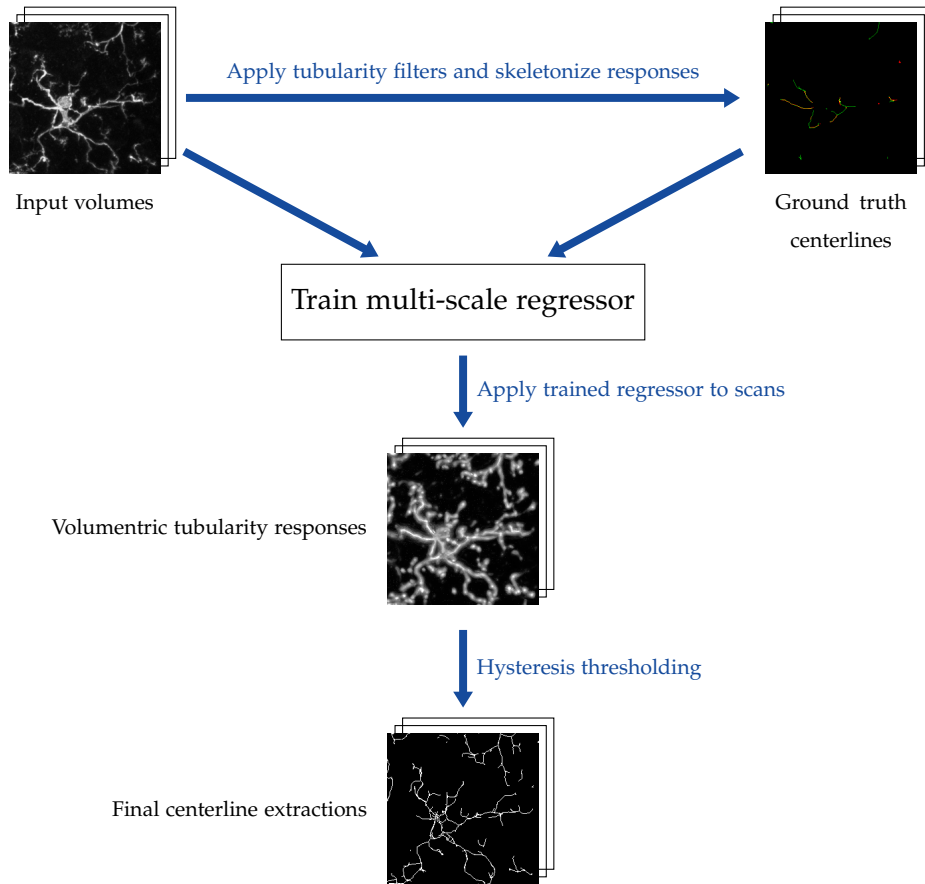


Figure 3.3: Methodology to extract process centerlines from volumetric confocal scans of microglia. Ground truth training data is automatically deduced from the input images by filters and does not have to be obtained separately. Exemplary image crops are depicted as max-projections.

As MDOF and Frangi filters are specifically designed to capture tubular structures and to return maximal values for voxels on the respective centerlines, the fragmented ground truth centerlines deduced from those filters are sufficiently accurate. Manually obtained ground truth data on the contrary fails to outline true centerlines in the scan with a similar precision.

Applying the trained regressors to a test volume results in a four-dimensional output with three spatial and one scale dimension. Taking the maximum over the last dimension combines all scales to a volumetric tubularity output. The final skeleton is obtained by binarizing and skeletonizing the tubularity response volume using hysteresis thresholding [Xie] and medial axis thinning [LKC94]. In Fig. 3.4 we give an example of extracting the centerlines from a volume with our approach and show intermediate results of all core tasks.

A strength of our presented approach is that it is highly generic and can be applied easily to any data. Since the ground truth is generated in an unsupervised way, no manual centerline extraction is required. The only data-dependent parameters are the scales, which can be found easily by inspecting the data visually, and the thresholds to skeletonize the filter responses with. These thresholds are quite robust and can be chosen rather restrictive for a sufficient amount of training images.

### **3.2.2 Partitioning**

Having extracted an over-complete skeleton of process centerlines, we further refine the results to obtain accurate extractions of individual cells. In this section, the process centerlines are merged with the corresponding soma detections and partitioned into separate cells. In the next section, the partitioned cells are pruned.

A significant attribute of microglia is their hierarchical structure. Taking the soma as root node, each microglia cell forms a tree-shaped network. In order to model this tree structure, we merge the skeleton from the centerline detection with the corresponding soma segmentations and then transform it into a graph. Bifurcations and end points in the skeleton become nodes; direct connections between two nodes are modeled by edges. Nodes representing a soma are labeled as root nodes. All subsequent partitioning and pruning steps are applied on the resulting graph.

We extract individual cells by partitioning the graph into multiple components containing exactly one root node each. Since the extracted graph is considered to be an over-approximation of the underlying microglial structures, graph components with no root node are discarded. Connected components with more than one root node



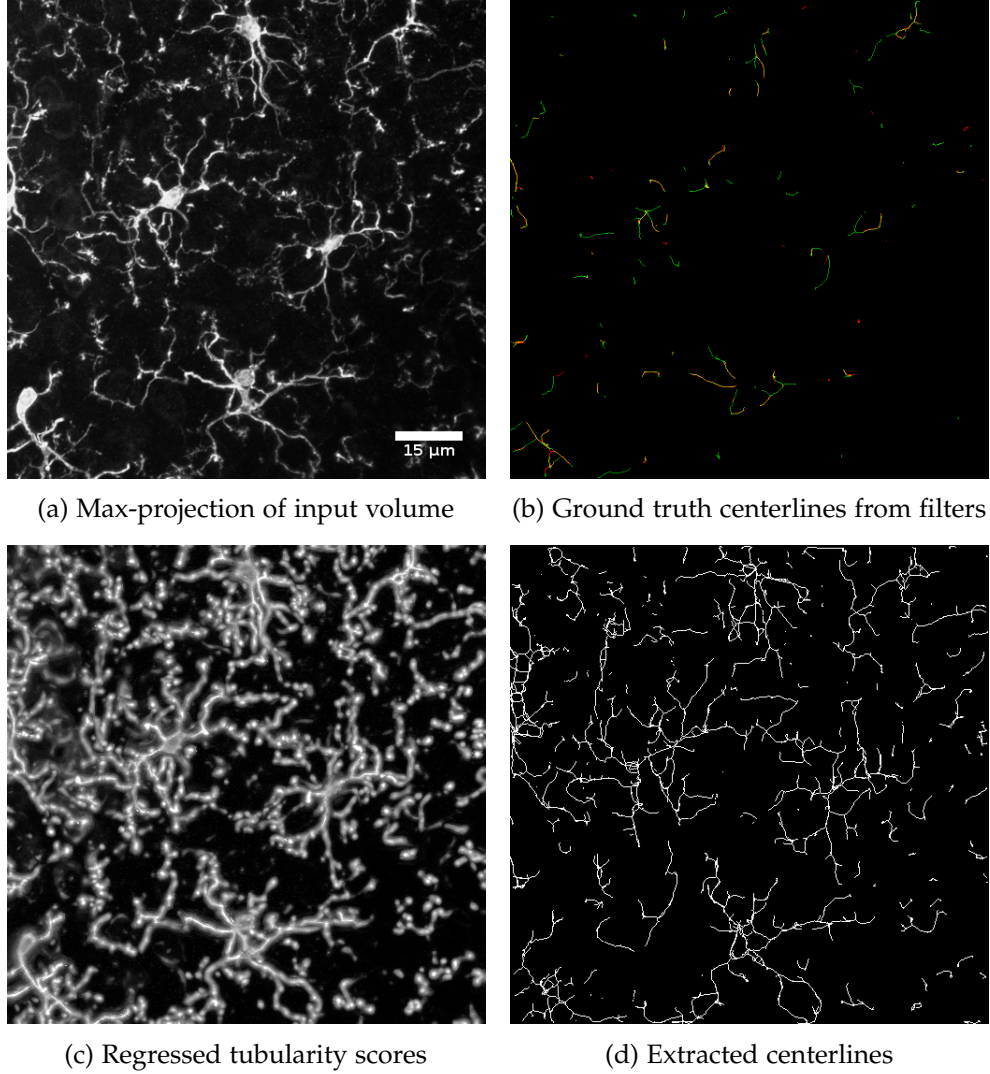


Figure 3.4: Example for centerline extraction: a) Max-projected confocal scan of microglia. b) Fragmented ground truth centerlines obtained by MDOF filters (green) and Frangi filters (red). Overlaps are marked in yellow. c) Tubularity response of the trained regressor. d) Extracted skeleton from (c) by hysteresis thresholding.

are partitioned by topology. Each edge in the component is assigned to the root node that provides the lowest topological score. Ties are resolved randomly. The topological score of a root node to an edge in this context is defined as the number of edges on the shortest path from the root node to one of the two edge nodes. Formulated differently, we assign edges to the closest root node on an equidistant graph.

Applied to a connected component with  $n$  root nodes, the presented approach results in  $n$  partitions with exactly one root node per partition. Each of these partitions is a connected subgraph of the initial structure and describes one microglia cell. In contrast to cutting the graph, a big advantage of subdividing in this way is that no edges are removed from the graph. Instead, neighboring partitions in the result share common nodes. In Fig. 3.5 we visualize cell separation for a synthetic graph. On the top a subgraph of a component with two root nodes is given. Edge weights in red and blue provide the topological distances from the first and the second root nodes in the graph. The second drawing below shows the resulting partitions after each edge is assigned to the root node, which provides the lowest topological distance. Thereby, the assignment of the edge, whose distances to both root nodes have the value 3, is resolved randomly. Nodes in black represent touching points of the two resulting components.

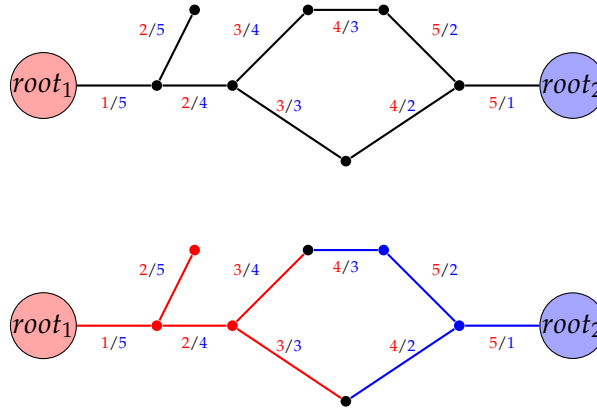


Figure 3.5: Synthetic example for cell separation based on topology. Top: Subgraph of a component with two root nodes. Edge weights in red and blue provide the topological distances to both root nodes. Bottom: Resulting partitions after each node has been assigned to the root node, which provides the lowest topological distance. One tie is resolved randomly; touching points of the components are labeled in black.

In real tissue microglia cells might appear very close to each other, but are not connected. This fact is ignored in our presented approach for cell separation. Nevertheless, we can

justify an output of the partitioning algorithms, in which extracted cells touch each other (as for example the two cells in Fig. 3.5 at the black nodes), by the following reasoning: One possibility is that a shared node represents two disconnected but very close process endpoints, which the centerline extraction method could not distinguish. In such a case it is reasonable that the node is part of two partitions, since it actually models two very close nodes. Considering the two partitions as disconnected separates the cells successfully. As a second possible situation, one of the links to the common node might be spurious. If this scenario takes place and one of the edges has been detected falsely, the subsequent pruning step resolves the issue.

### 3.2.3 Pruning

Finally, the obtained graphs are pruned to remove spurious links and cycles. Having extracted separate graphs for all microglia cells, this is done individually for one cell at a time.

Analogous to [Tür+16], pruning is formulated as a integer programming task. Assuming one root node per graph, the involved constraints enforce a tree-shaped and thus acyclic structure of the graphs. Solving the individual integer programs of all partitions returns optimally pruned microglia cells based on affinity weights of their graph edges. These affinities model for all edges the likeliness of representing a true process in the underlying scan. As source for that information the image intensities, the response of the centerline regressor and the local thickness score of the scanned volume as a measure for the radius are taken. Concrete values are computed by averaging the respective signals along the outlines of the graph edges. The affinity scores of the three different signals are combined to the final affinity weights by a weighted sum.

Exemplary illustrations of the partitioning and pruning steps are given in Fig. 3.6.

## 3.3 Feature extraction

We use various properties to describe the morphology of extracted microglia. More precisely, we distinguish in this thesis between L-measure, spatial extent and hierarchical features.

The software tool L-Measure [SPA08] is originally designed to characterize neuronal morphology, but works also well for microglia in related literature [Xu+16; Meg+15]. Using the library, a large number of mostly arbor measures are computed from the extracted cells on compartment, branch, bifurcation and cell level. A compartment

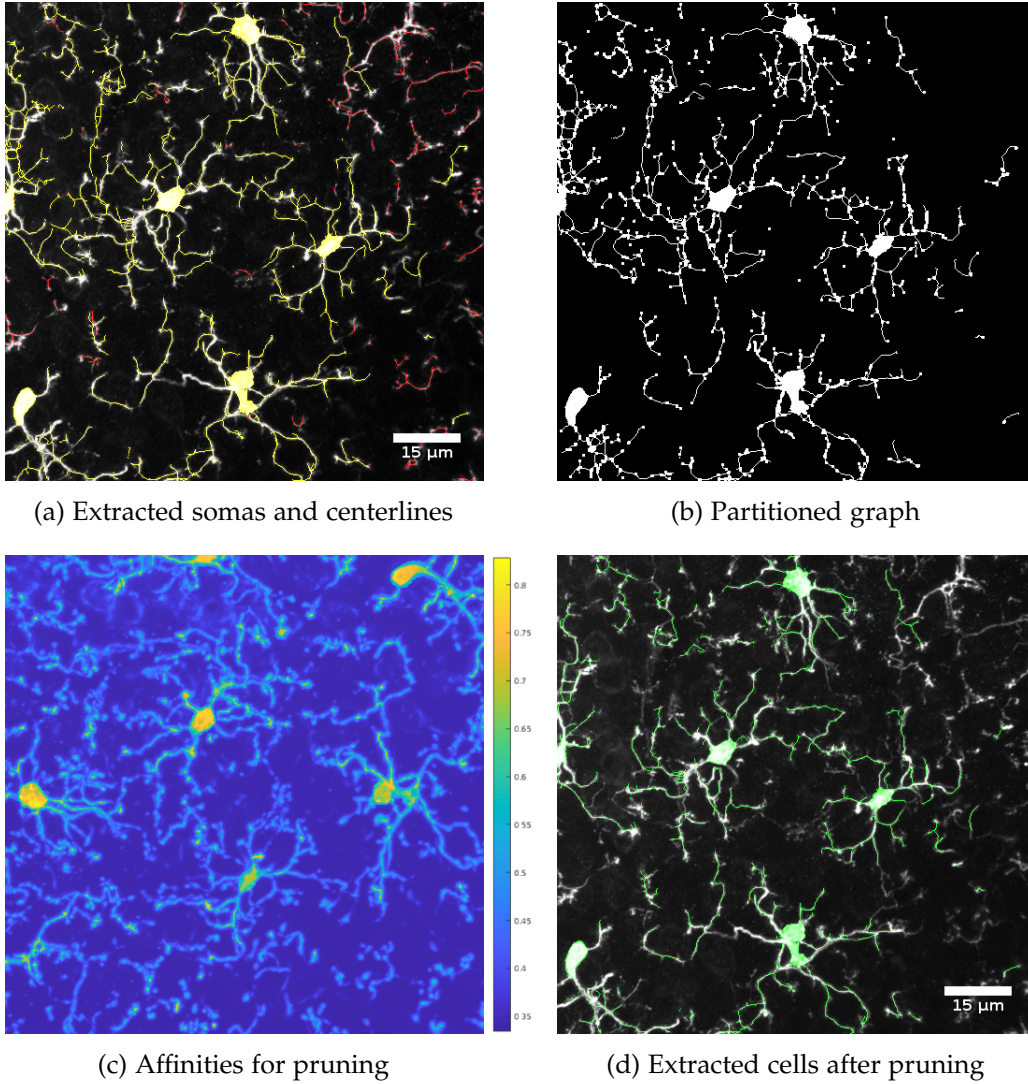


Figure 3.6: Example for partitioning and pruning: a) Max-projection of the results from soma segmentation and centerline extraction before pruning. In order to keep the images uncluttered, we only show a subset of 60 slices of the volume. Disconnected components are labeled in red and will be discarded immediately. The original scan is underlaid for a visual evaluation of the results. b) Partitioned graph calculated from all connected components in (a). Thicker parts of the skeleton represent nodes. c) Projection of the combined affinity scores, which are used to prune the graph from (b). Weights for edges in (b) are calculated by averaging the scores along their outlines. d) Resulting cells after pruning (b) with affinity weights deduced from (c). Seemingly loose fragments in the image are connected to a soma in slices that are outside of the shown range.

represents the smallest unit of a detected process and corresponds to one centerline voxel in a volumetric scan. A branch consists of all compartments lying between two bifurcation points or between one bifurcation and a termination point. Properties on sub-cell levels are transformed to features for a whole cell by calculating the sum, minimum, maximum, average and standard deviation of the measured values. An overview about all L-measures is given in Table 3.1. In total, 178 features are defined in this way.

Nr.	Feature name	Level	Nr.	Feature name	Level
1	Soma surface	compartment	22	Branch path length	branch
2	Number of processes per soma	cell	23	Contraction	branch
3	Number of bifurcations	cell	24	Fragmentation	branch
4	Number of branches	cell	25	Daughter Ratio	branch
5	Number of terminal endpoints	cell	26	Parent daughter ratio	branch
6	Width	cell	27	Partition asymmetry	branch
7	Hight	cell	28	Rall power	branch
8	Depth	cell	29	Pk	branch
9	Diameter	compartment	30	Pk classic	branch
10	Diameter pow	compartment	31	Pk 2	branch
11	Length	compartment	32	Bifurcation angle local	bifurcation
12	Surface	compartment	33	Bifurcation angle remote	bifurcation
13	Section area	compartment	34	Bifurcation tilt local	bifurcation
14	Volume	compartment	35	Bifurcation tilt remote	bifurcation
15	Euclidean distance to soma	compartment	36	Bifurcation torque local	bifurcation
16	Path distance to soma	compartment	37	Bifurcation torque remote	bifurcation
17	Branch order	compartment	38	Last parent diameter	branch
18	Terminal degree	compartment	39	Diameter threshold	branch
19	Terminal segment	cell	40	Diameter Hillman threshold	branch
20	Burke taper	branch	41	Helix	compartment
21	Hillman taper	branch	42	Fractal dimension	branch

Table 3.1: L-measures. Properties on sub-cell levels are transformed to features for a whole cell by calculating the sum, minimum, maximum, average and standard deviation of all measured values. A more detailed documentation of all listed L-measures is given in [Pol].

As the L-measure library does not define features on complete paths from the root to leaf nodes, such measures are computed separately. In this way another type of hierarchical information is added to the feature space. Table 3.2 gives an overview about the considered properties. Summed up, 13 hierarchical features are defined.

Spatial extent features are added to capture the shape and elongation of extracted cells more precisely. Since somas as well as processes appear differently in microglia of altering activation states, we define these features twice: On the one hand for the whole cell and on the other hand for the soma separately. Volume and surface measures

Nr.	Feature name	Level
1	Topological depth	path
2	Number of paths	cell
3	Number of child nodes	path
4	Geodesic length	path

Table 3.2: Hierarchical features. All measures defined on path level are transformed to features for a whole cell by computing the minimum, maximum, average and standard deviation of all measured values.

are deduced from their convex hulls and their boundary points [Mat]. Shape-related size measures are obtained from two ellipsoids fitted to the soma and the whole cell. Table 3.3 lists in its first 13 lines all spatial extent features, which are calculated for somas and cells. By setting the computed values for the whole cell in proportion to the ones for the soma, additional spatial extent descriptors are obtained. In this way, the last two features of Table 3.3 are defined, yielding 28 spatial extent features in total.

Nr.	Feature name	Level
1	Convex hull volume	cell and soma
2	Boundary volume	cell and soma
3	Ratio boundary volume and convex hull volume	cell and soma
4	Convex hull surface	cell and soma
5	Boundary surface	cell and soma
6	Ratio of convex hull surface and boundary surface	cell and soma
7	Volume of fitted ellipsoid	cell and soma
8	First ellipsoidal radius	cell and soma
9	Second ellipsoidal radius	cell and soma
10	Third ellipsoidal radius	cell and soma
11	Ratio of first and second ellipsoidal radius	cell and soma
12	Ratio of first and third ellipsoidal radius	cell and soma
13	Ratio of second and third ellipsoidal radius	cell and soma
14	Ratio of convex hull volume of the whole cell and the soma	cell
15	Ratio of boundary volume of the whole cell and the soma	cell

Table 3.3: Spatial extent features. Most of the features are computed for the whole cell and the soma separately.

Combining all listed properties yields a 219-dimensional feature space. Since some of the defined features describe similar properties, feature selection is performed to select a subset of relevant measures. As a first advantage, a smaller feature space reduces the computational cost of the subsequent analyses. Secondly, removing noisy features allows an algorithm to focus on relevant properties and may improve the performance. Amongst all methods compared in [Rof+17], Minimum Redundancy Maximum Relevance (mRMR) feature selection performs best for the data of this thesis.

### 3.4 Analysis

Experiments in the thesis are subdivided into three parts: Firstly, we study well-known phenotypes from literature by classification. Secondly, we apply clustering techniques to the extracted microglia to discover new morphological classes. Finally, we use microglial morphology as a predictor of injury severity by regressing the distance to the brain injury site.

In the first part, extracted cells are classified into the three activation states "ramified", "partially activated" and "strongly activated". These classes are typically mentioned in literature and represent well-accepted microglial phenotypes at a coarse level. They are easily identified by human experts and thus make it possible to generate labeled training data for the classification. For the experiments, we use standard classifiers from the Matlab toolbox "prtools" [PD]. In order to cover a broad range of different techniques and approaches, we choose the following set of classifiers: Fisher's linear least square classifier (fisherc), linear Bayes normal classifier (ldc), quadratic Bayes normal classifier (qdc), random forest classifier (randomforestc), k-nearest neighbor classifier (knn) and support vector machines using several polynomial kernels (linSVM, p2SVM, p3SVM, p5SVM, p7SVM) as well as a Gaussian radial basis function kernel (rbfSVM).

In the second part of the analysis we apply clustering techniques to discover more precise microglial phenotypes. Using the k-means algorithm, Gaussian mixture models and spectral clustering we try to discover any structure (i.e. clusters) in the feature space of the extracted cells. This is because additional phenotypes will naturally manifest as clusters in the feature space.

The k-means algorithm divides the data into  $k$  clusters by minimizing the sum of squared Euclidean distances of all data points to their closest cluster center. The method represents a very simple approach and fails to capture complex non-spherical cluster shapes. The Gaussian mixture model overcomes this limitation by modeling the data as a mixture of  $k$  Gaussian distributions with separate parameters for the means and the variances. In this way clusters can have ellipsoidal shapes. Beyond Gaussian mixture models we apply spectral clustering, as the method is not restricted to ellipsoidal shapes and can detect non-intersecting clusters of any shape. In spectral clustering the data is partitioned based on a similarity measure. A Laplacian matrix is deduced from the similarity matrix of the data. Then, a standard clustering technique as for example k-means is used to cluster the data in the transformed space spanned by the Laplacian.

In the third type of analysis, we apply regression and predict the distance to a brain

injury from microglial morphology. Instead of classification and clustering approaches that operate on nominal classes, we quantify in this way a continuous shape progression between microglia cells close to an injury and those further away. From a more abstracted point of view, predicting the distance to the injury can be considered as estimating injury severity of a scanned area.



## 4 Results

In this chapter we present our results from applying the introduced methodology. At the beginning, we give an overview about the scanning modalities and the structure of the processed data. Next, we evaluate the introduced methodology for extracting microglial structures from confocal scans. Ultimately, we analyze extracted microglia by classification, clustering and regression experiments in the last part of the chapter.

### 4.1 Datasets

The volumetric data for the experiments in this thesis are obtained from mice expressing green fluorescent proteins. Traumatic brain injury in the controlled cortical impact model is caused in two severity degrees to analyze microglial morphology in various activation states. Injuries with a depth of 2 *mm* are considered as severe, injuries with a depth of 1 *mm* as mild. Brain tissue is cleared using the uDISCO tissue clearing protocol [Pan+16] and is scanned by a confocal microscope. All tasks related to animal treatment, tissue clearing and imaging have been executed by members of the Acute Brain Injury Research Group at the Institute for Stroke and Dementia Research, Ludwig-Maximilian-Universität München.

All retrieved scans have a resolution of  $1024 \times 1024 \times (150 - 200)$  voxels (the exact depths vary in the scans). Before we process the scans with our tracing framework, we resize all scans to the half of their original size and adjust their voxel spacing to be isotropic. Finally, all scans have a resolution of  $512 \times 512 \times (150 - 500)$  voxels. Using the methodology introduced in Chapter 3, we extract microglia from the volumetric confocal scans. In this way, we generate the following two datasets:

A first dataset is generated from 38 scans of three healthy, one severely injured and one mildly injured mouse brains. The voxel size during imaging is  $0.1038 \times 0.1038 \times 0.5 \mu\text{m}^3$ , so that one scan covers an area of approximately  $106 \times 106 \times 100 \mu\text{m}^3$  of the tissue.

Scans are obtained in all brains from the same region on the injury site, allowing to compare the data of different animals to each other. Since the locations and distances to the injury are approximately equal for all captured microglia cells, we can label their

current morphology by the injury degree of the respective mouse. In scans of healthy tissue microglia are resting and appears ramified. A mild injury leads to partially activated cells in the surrounding tissue at the lesion site, a severe injury to fully activated cells. Based on this reasoning, a labeled dataset with the classes "ramified", "partially activated" and "fully activated" is generated from microglia of healthy, mildly injured and severely injured tissue. Exemplary scans from the dataset are shown in Fig. 4.2 of the subsequent evaluation section.

A second dataset is obtained from one severely injured brain. Scans are imaged with a lower magnification of  $0.2012 \times 0.2012 \times 0.5 \mu\text{m}^3$  per voxel. Thus, one tissue scan covers a larger area of approximately  $206 \times 206 \times 100 \mu\text{m}^3$ .

Depending on the distance to the injury, microglia appear in altering shapes in the severely injured brain of dataset 2. Therefore, we can image a transition between different phenotypes of microglia close to and further away from the injury. To capture these morphological changes, scanning is started at the injury and continued perpendicularly to the opposite end of the lesion site in the brain. We obtain scans in regular intervals approximately  $600 \mu\text{m}$ ,  $1200 \mu\text{m}$ ,  $1800 \mu\text{m}$ ,  $2400 \mu\text{m}$  and  $3000 \mu\text{m}$  away from the injury. Imaging four neighboring tile scans for each of the five distance levels return in total 20 scans for dataset 2. As one tile has a width of approximately  $200 \mu\text{m}$ , there is a gap of around  $400 \mu\text{m}$  between two subsequent distance levels. Fig. 4.1 shows all scans of dataset 2 as max-projections.

Table 4.1 summarizes the scanning modalities and the extracted microglia cells of the two datasets.

	Dataset 1	Dataset 2
Number of scans	38	20
Approx. volume of a scans	$106 \times 106 \times 100 \mu\text{m}^3$	$206 \times 206 \times 100 \mu\text{m}^3$
Number of extracted cells	1143	1849
Additional property	Activation state: <ul style="list-style-type: none"> <li>- Ramified (260 cells)</li> <li>- Partially activated (185 cells)</li> <li>- Fully activated (698 cells)</li> </ul>	Distance to injury: <ul style="list-style-type: none"> <li>- <math>600 \mu\text{m}</math> (434 cells)</li> <li>- <math>1200 \mu\text{m}</math> (455 cells)</li> <li>- <math>1800 \mu\text{m}</math> (383 cells)</li> <li>- <math>2400 \mu\text{m}</math> (323 cells)</li> <li>- <math>3000 \mu\text{m}</math> (254 cells)</li> </ul>

Table 4.1: Summary of the two datasets we use in our analyses.

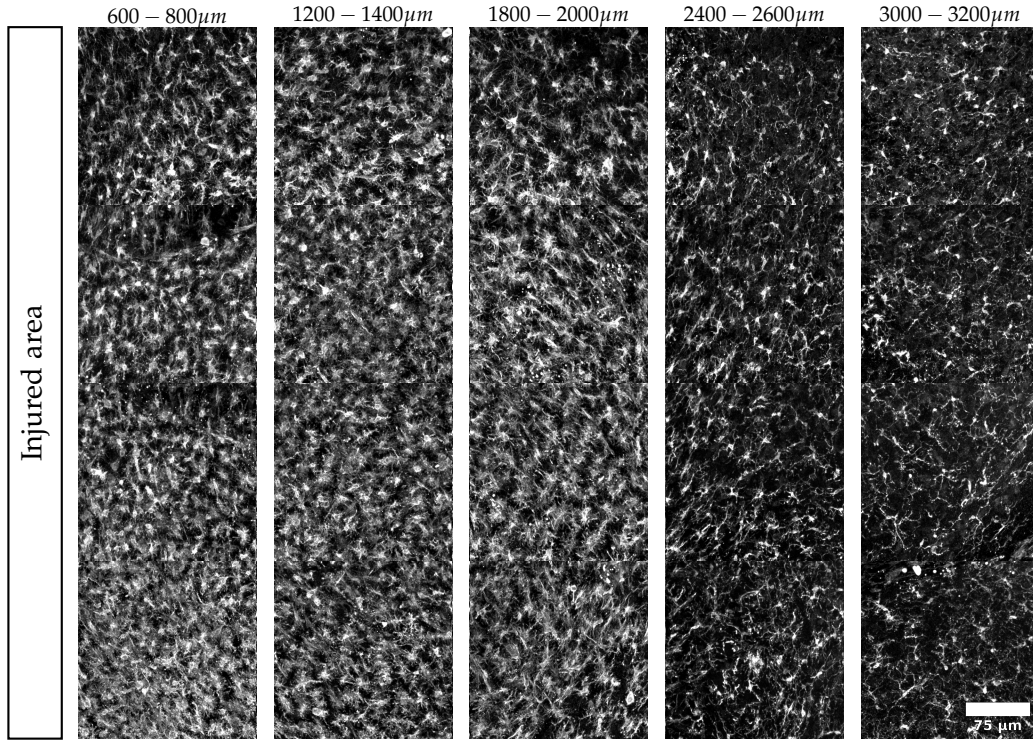


Figure 4.1: Max-projection of the scans from dataset 2. Scanning is started at the injury site and continued perpendicularly to the opposite end of the lesion site in the brain in regular intervals approximately  $600 \mu m$ ,  $1200 \mu m$ ,  $1800 \mu m$ ,  $2400 \mu m$  and  $3000 \mu m$  away from the injury. We image four neighboring tile scans for each of the five distance levels.

## 4.2 Evaluation

As there is no manual ground truth available, we cannot evaluate our microglia extractions numerically. Nevertheless, visual inspection provides sufficient evidence about how meaningful and correct microglia are captured. Fig. 4.2 shows extracted microglia in scans of healthy, mildly injured and severely injured tissue from dataset 1.

In the following we first evaluate the tasks of soma segmentation and centerline extraction individually. Then, we discuss the combination of their results to complete cell extractions during partitioning and pruning.

### 4.2.1 Soma segmentation

The introduced approach for soma segmentation is able to capture somas in various activation states. Fig. 4.2 demonstrates that the method segments somas in altering shapes, ranging from small spherical somas in healthy tissue to bigger, highly irregular and sometimes protracted somas in injured tissue.

As a non-learning method, our algorithm requires its parameters to be tuned manually and is sensitive to changes in the data as e.g. the signal-to-noise ratio or the magnification of the scan. Having adjusted the method appropriately for a given dataset, it reliably extracts somas. However, while the approach works well for the majority of the cells, it still has limitations. Fig. 4.3 depicts common errors of the algorithms.

Found segmentations do not always fit exactly to the signal boundaries in the scans. In some situations the algorithm fails to detect the soma edge but includes parts of attached processes to the segmentation. An example is given in Fig. 4.3a.

The exact opposite can happen in scans with low signal-to-noise ratios, as for example depicted in Fig. 4.3b. Here, the found segmentation does not completely cover the soma structure in the image signal. Such a situation often raise further negative consequences during centerline extraction. For the one thing, uncovered parts of the soma may cause false centerline detections. For the another thing, a too small soma may lead to disconnected processes, which are then discarded.

So far discussed limitations alter some features of affected cells, but do not change their morphology fundamentally. In contrast, the errors depicted in Fig. 4.3c and Fig. 4.3d introduce more problematic situations. As somas are segmented by considering isotropic and solid structures, thick processes and bifurcations are sometimes detected falsely. An example is shown in 4.3c. Though it is possible to leverage between false and missing detections by filtering small somas after the segmentation, such situations

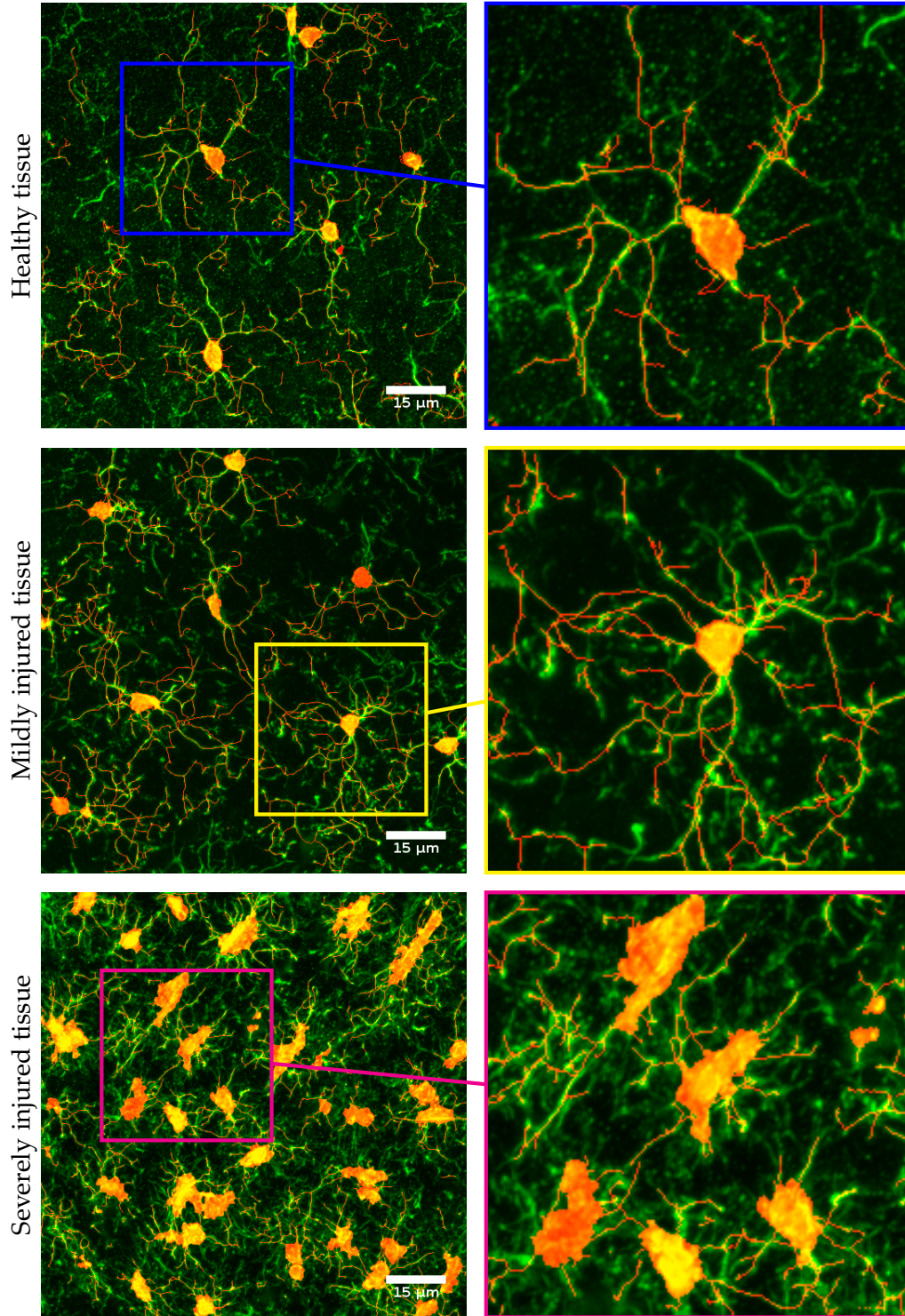


Figure 4.2: Exemplary extractions of microglia from healthy tissue (top), mildly injured tissue (middle) and severely injured tissue (bottom). All volumes are visualized as max-projections in the x-y-plane.

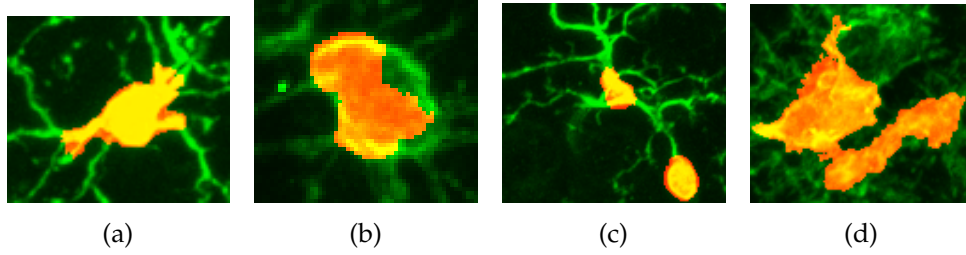


Figure 4.3: Limitations of soma segmentation: Exemplary visualizations.

cannot be prevented completely.

Another problematic situation is shown in Fig. 4.3d. Here, the soma segmentation algorithm cannot distinguish the two neighboring somas in the scan but segments them as one. As activated microglia cells move closer to each other and appear in highly ramified shapes, this is more likely to happen in tissue close to an injury. These two lastly mentioned kind of faults are especially problematic because they affect multiple cells. As they accumulate during partitioning, consequences are discussed in detail in the subsequent Section 4.2.3.

#### 4.2.2 Centerline extraction

Our approach for centerline extraction based on regression works well for the processed data. Tubular structures are enhanced and produce maximal responses on their centerlines. Intensity inhomogeneities along processes are compensated, resulting in smooth tubularity responses. We observed visually that the algorithm is robust against noise and can be easily transferred to different data. In the zoomed regions of Fig. 4.2 we give an impression of precision of extracted centerlines.

In Fig. 4.4 we demonstrate the effectiveness of our presented method to enhance tubular structures. The first image (Fig. 4.4a) shows a projected scan of a microglia cell with a very weak signal intensity; the second image (Fig. 4.4b) the corresponding regression-based tubularity response. For comparison, the MDOF filter output of the scan is given in as well (Fig. 4.4c). While the result from the MDOF filter only contains fragments of the processes, the approach of this thesis can boost all tubular structures of the scan.

Although our approach can deal with noise and weak signals, it cannot compensate all limitations of scanning. Strong flaws as for example gaps in the signal along tubular structures are propagated and lead to missing parts in the extracted centerlines. We give an example for this in Fig. 4.4d, depicting the skeleton extracted from the tubularity

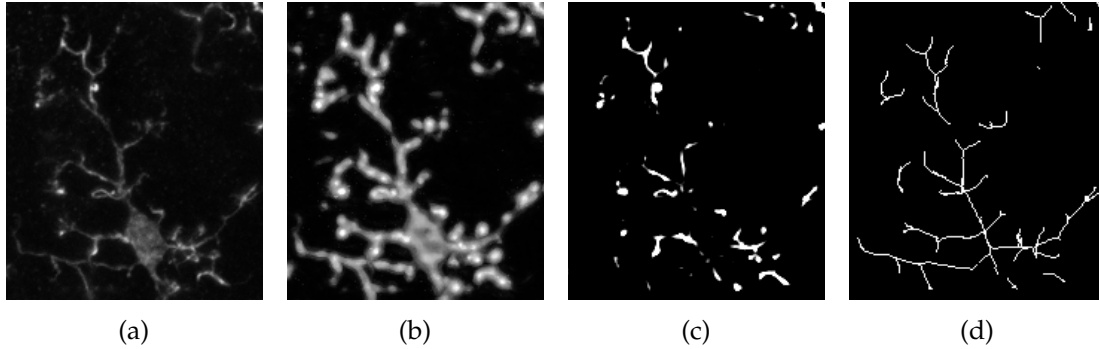


Figure 4.4: Evaluation of centerline extraction: a) Projected scan of microglia cell. b) Regression-based tubularity score of the methodology in this thesis. Almost all tubular structures from (a) are boosted. c) MDOF tubularity response that only captures fragments of all processes from (a). d) Extracted centerlines by skeletonizing the tubularity response from b).

response of Fig. 4.4b. Almost all processes are captured by centerlines. However, since there are spots with almost no intensity in the underlying scan, the extracted structure also contains gaps at those positions.

### 4.2.3 Partitioning and Pruning

The introduced partitioning approach works reliably and yields visually reasonable results. If two or more somas are connected to each other by the detected processes, the algorithm separates them into topologically similarly sized partitions. As our method does not cut connections, no additional errors are introduced in this way.

However, during the partitioning step we combine the results from soma segmentation and centerline extraction. Thereby, also the mistakes from these previous steps are propagated and accumulate.

Soma detections are taken as root nodes to separate multiple cells in a scan based on topology. Therefore, in the case of a false detection as shown in Fig. 4.3c, the single cell in the scan is partitioned into two separate structures. In this way, the extracted morphology changes fundamentally.

A similar situation occurs if two soma segmentations are merged accidentally as illustrated in 4.3d. Then, two microglia cells in the scan are represented as one structure, leading again to a significant difference between the extracted shape and the true outlines of the cells in the underlying scan.

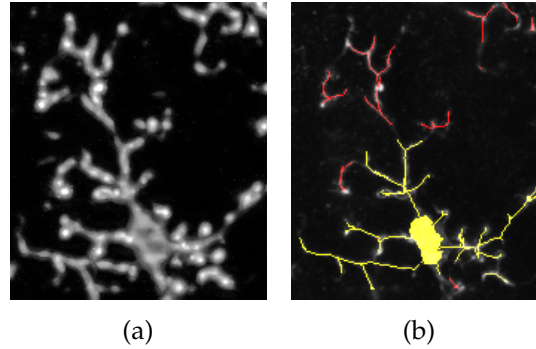


Figure 4.5: Limitation of the introduced approach for process extraction that accumulates during the partitioning step: a) Projected tubularity measure of microglia cell. b) Reconstructed cell in yellow. Disconnected components in red are discarded, since they do not connect to the soma.

A second type of inaccuracy that accumulates in the partitioning step affects the processes of microglia. If there are small gaps in the extracted centerlines, the resulting disconnected components are not included in any extractions but are ignored. Picking up the exemplary cell of Fig. 4.4, its discarded components are highlighted in red in Fig. 4.5. For comparison, the tubularity response of the regressor is shown as well.

The pruning algorithm we use works directly on information from the imaged data. We observe that our presented method is robust against small flaws that are introduced during scanning and does not cause additional errors.

### 4.3 Classification

We perform classification to gain insights into the morphology and characteristics of ramified microglia in healthy tissue, partially activated microglia in mildly injured tissue and fully activated microglia in severely injured tissue. As the cells from dataset 1 contain labels for these nominal classes, classification techniques are directly applicable. We train a range of standard classifiers to distinguish the three classes "ramified", "partially activated" and "fully activated" microglia. Evaluation is performed by leave-one-out cross-validation.

The classification errors of all approaches are summarized in Fig. 4.6. We can successfully distinguish the three classes with accuracies up to 95%. As a linear Bayes classifier performs best, the classes seem to be linearly separable. This is confirmed by



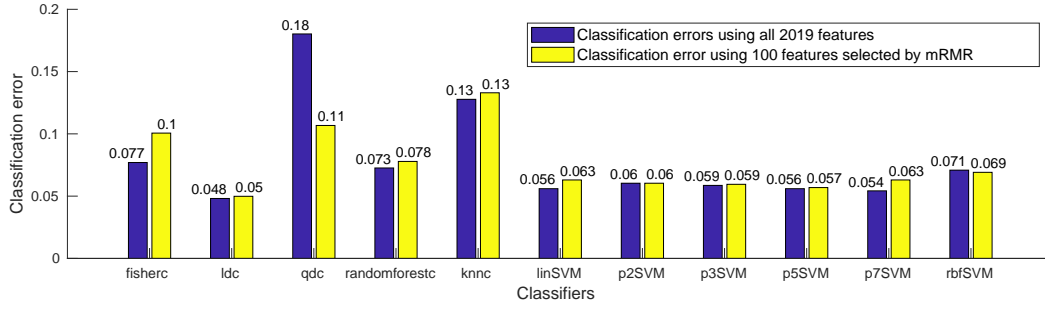


Figure 4.6: Classification errors for distinguishing extracted microglia from dataset 1 into ramified, partially activated and fully activated cells. We show our results of experiments with all features (blue) and with a subset of 100 features selected by mRMR (yellow).

a linear support vector machine (SVM) that achieves an accuracy score of 94%. Various polynomial SVMs can approximate linear functions and thus as well achieve high accuracy scores of 94 – 95%. Fisher’s linear least square classifier, a random forest and a SVM with a Gaussian kernel also gain acceptable classification scores of 92 – 93%. The quadratic Bayes and the k-nearest neighbor classifiers model the data insufficiently and produce significantly worse results.

We can illustrate the linear separability of the three classes in a scatter diagram. Fig. 4.7 plots the average process diameter of all cells against the standard deviation of their fragmentation score (number of compartments between two bifurcations or a bifurcation and a termination point). It reveals that the classes can be roughly distinguished by lines using only these two features. Ramified microglia have relatively thin branches with strongly varying fragmentation scores. Activated cells on the contrary consist of thicker processes and deviate less in their fragmentation scores.

Since approximate classes can already be deduced from a scatter plot of two features, we apply feature selection. Fig. 4.8 demonstrates that using more than 100 features found by Minimum Redundancy Maximum Relevance (mRMR) feature selection does not significantly improve the classification accuracy of the linear Bayes classifier any more. Utilized for all classifiers in Fig. 4.6, the computational costs of the algorithms can be reduced significantly, while the error rates increase only marginally.

For some classifiers, feature selection even improves the classification result. Selecting a subset of relevant features can decrease noise in the data and contribute in this way to higher classification accuracies. Taking the optimal number of 170 features for the linear Bayes classifier reduces for example its error rate from 5% to 4%.

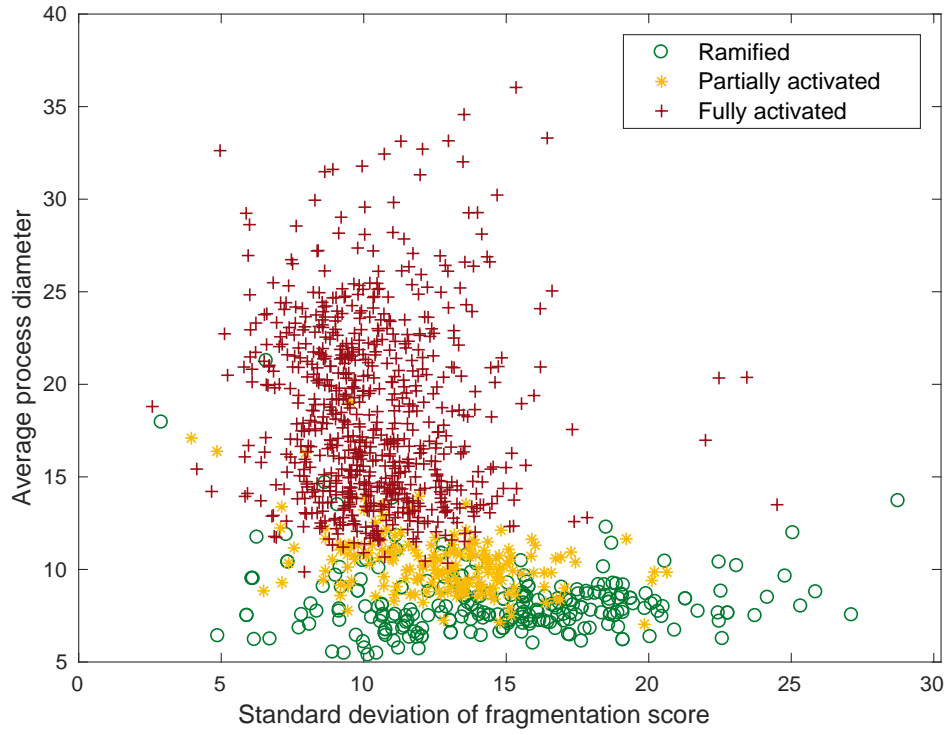


Figure 4.7: Two-dimensional scatter plot of the extracted microglia from dataset 1. The plot reveals that the classes "ramified", "partially activated" and "fully activated" are nearly linearly separable by considering only two features.

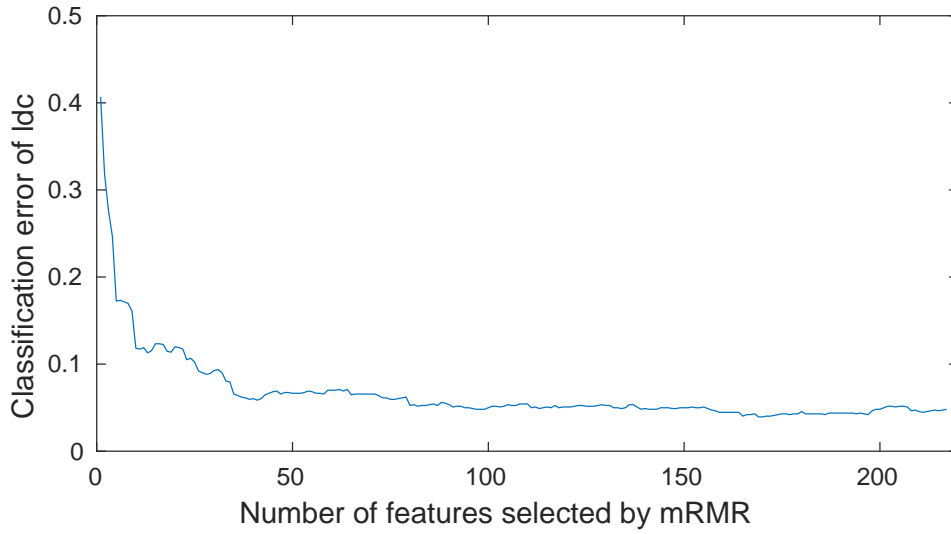


Figure 4.8: Illustration of the classification error of the ldc classifier as a function of the number of features selected by the mRMR feature selection. Considering more than 100 features does not significantly improve the classification performance any more.

We deduce further insights from the per-class classification errors illustrated in the Fig. 4.9. Whereas fully activated microglia are detected almost perfectly, partially activated and ramified cells are misclassified with error rates of approximately 8% and 20%. These inaccuracies do not considerably influence the overall classification score because the respective classes are three to four times less frequent than the partition of fully activated cells.

However, the per-class classification errors reveal that extracted cells from healthy and mildly injured tissue have a partially similar morphology and can only be distinguished up to a limited extent. According to the confusion matrix in Table 4.2 the two classes are mixed up with each other evenly during classification.

As of yet, all analysis is done on cells from dataset 1. Having trained a classifier on that data, it is further possible to apply it to the second dataset. Since dataset 2 contains cells roughly 600  $\mu m$ , 800  $\mu m$ , 1200  $\mu m$ , 1800  $\mu m$ , 2400  $\mu m$  and 3000  $\mu m$  away from the injury, we can relate the three known classes to these distances. Using a linear Bayes classifier trained on dataset 1, we classify the 1849 cells from the second dataset into 1198 fully activated, 109 partially activated and 542 ramified cells.

Fig. 4.10 summarizes the relative frequencies of predicted classes in relation to their

True labels	Estimated labels			Totals
	ramified	partially	fully	
ramified	239	16	5	260
partially	21	151	13	185
fully	1	1	696	698
Totals	261	168	714	1143

Table 4.2: Confusion matrix of the linear Bayes classifier from leave-one-out cross-validation. Ramified and partially activated microglia are mixed up with each other evenly during classification.

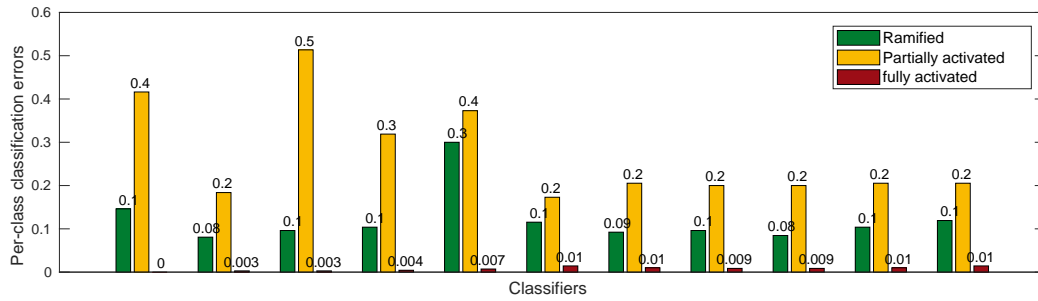


Figure 4.9: Per-class classification errors when using 100 features selected by mRMR. Ramified and partially activated cells are misclassified with considerably higher error rates than fully activated microglia.

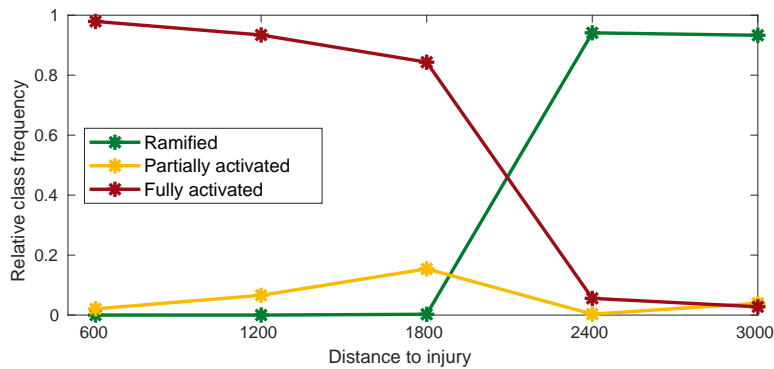


Figure 4.10: Quantification of predicted classes in dataset 2 relative to the distance to the injury. The class frequencies coincide perfectly with expected occurrences of ramified, partially activated and fully activated microglia.

distance to the injury. The classification results coincide perfectly with expected occurrences of ramified, partially activated and fully activated microglia. Within a distance of  $600 - 1200 \mu m$  away from the injury fully activated microglia are dominant. As the distance to the injury goes to  $2400 \mu m$ , the partition of partially activated cells increases as well. Even further away at the other end of the brain ( $2400 - 3000 \mu m$  away from the injury) microglia are mostly classified as ramified.

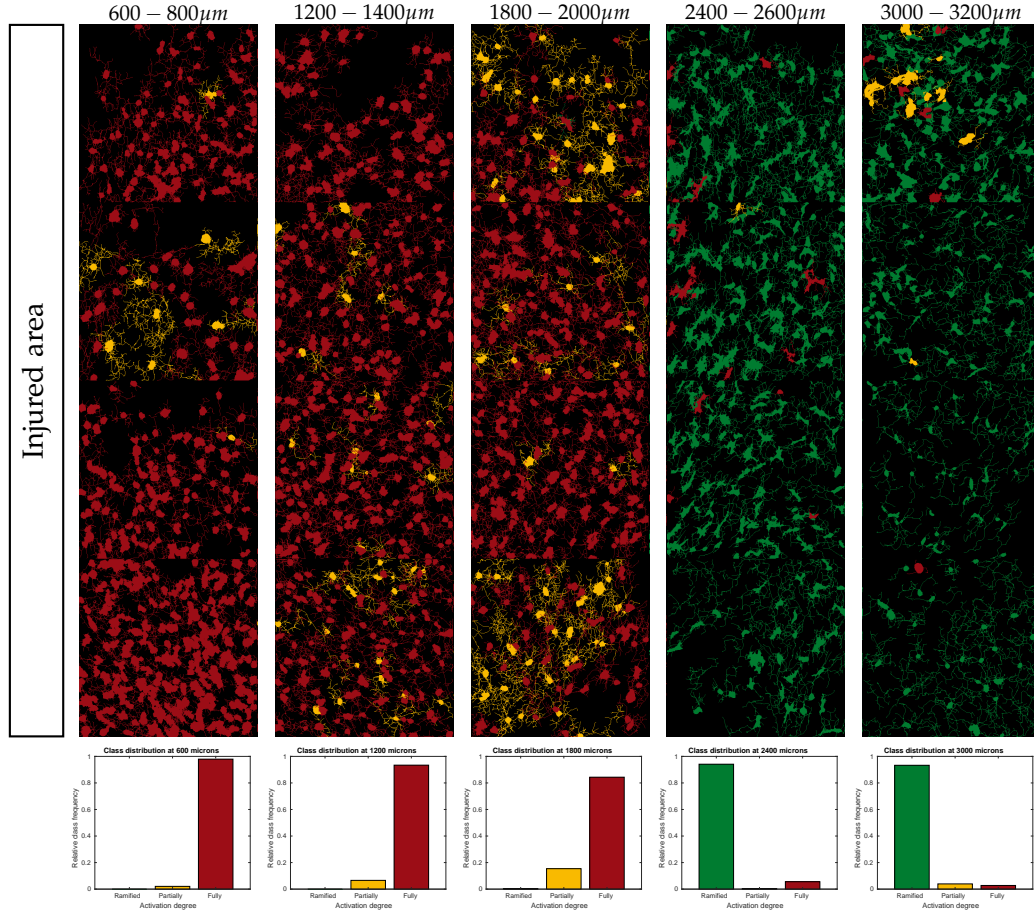


Figure 4.11: Spatial class distribution in dataset 2 obtained by a classifier which is trained on dataset 1. The histograms at the bottom show the class distributions at all distance levels, pointing out an abrupt change between  $1800 \mu m$  and  $2400 \mu m$ .

The spatial distribution of the three classes in dataset 2 is illustrated in Fig. 4.11. Additionally, the figure displays separate histograms of the class distributions at all

distance levels. These histograms point out an abrupt change in the predicted classes between 1800  $\mu m$  and 2400  $\mu m$ . The histograms of 2400  $\mu m$  and 3000  $\mu m$  are roughly identical. In the scans at 1800  $\mu m$  and closer to the injury the class distributions are also relatively similar, although there is some difference visible.

The absolute class frequencies reveal furthermore that fully activated cells appear most often. We explain this with the following two reasons: First, as a severe brain injury affects the whole brain in a serious way, it is from a biological perspective reasonable that a majority of the microglia is activated, also cells further away from the injury. Second, activated microglia cells have shorter processes and can thus appear closer to each other. As there are more cells in scans close to the injury, which are more likely to contain activated microglia, the classifier predicts more activated than ramified cells. Scans of dataset 2 with a distance to the injury of 600  $\mu m$  and 1200  $\mu m$  contain for example in average 110 cells per scan, whereas scans, which are 2400  $\mu m$  and 3000  $\mu m$  away from the injury, consist of only 72 cells in average.

## 4.4 Clustering

Apart from visually distinguishable classes in microglial morphology, studies assume a range of in-between phenotypes [CB17]. We apply clustering techniques in this section to analyze the structure of the obtained datasets more detailed and possibly discover new morphological classes. For that, cells from both datasets are considered together.

Using Gaussian mixture models we partition the extracted microglia into 3 – 7 clusters. The mixture model with three components detects the same classes as the classification experiments in the previous section have revealed. Clustering with five mixture components separates the extracted microglia visually into most distinct and precise morphological classes. Fig. 4.12 illustrates the five found clusters. The rows in the depicted matrix represent individual cells in the feature space and are grouped according to the detected classes. In this way, all five clusters appear as notably homogeneous blocks.

The relative frequencies of the found clusters are summarized in Fig. 4.13. For dataset 1 the frequencies are measured in relation to the known classes "ramified", "partially activated" and "fully activated"; for dataset 2 the clusters are quantified in relation to the distance to the injury. The plot of dataset 1 reveals that three of the five components from the Gaussian mixture model mainly cover activated cells. The other two clusters mostly consist of ramified cells. This differentiation emerges in a weaker manner also from the relative cluster frequencies of dataset 2. However, the classes change smoothly

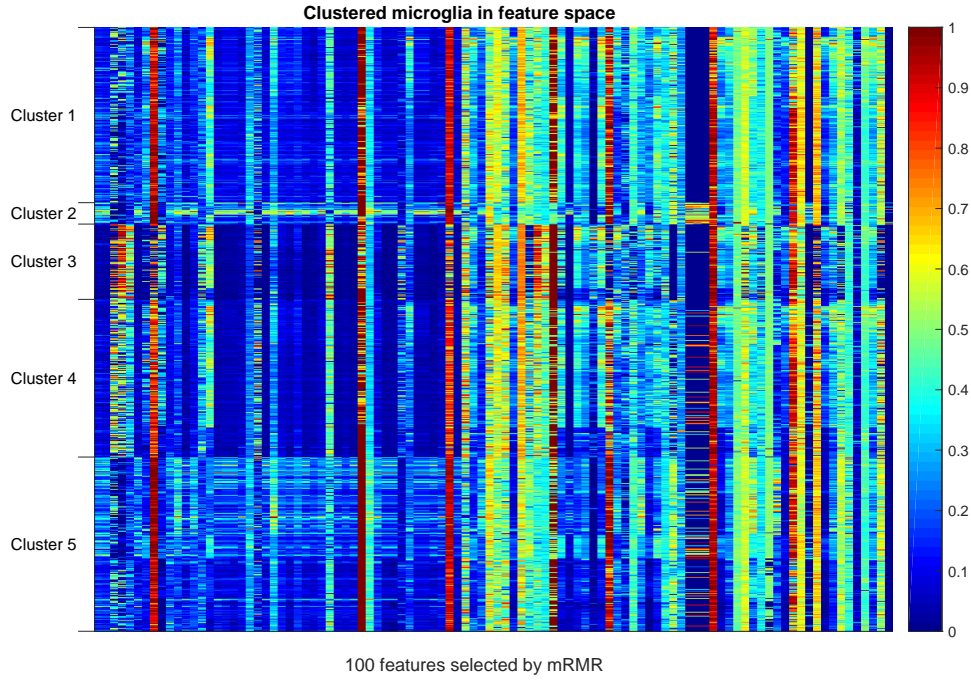


Figure 4.12: Visualization of the five morphological classes of microglia found by a Gaussian mixture model. The rows in the depicted matrix represent individual cells in the feature space and are grouped according to the detected classes. In this way, all five clusters appear as notably homogeneous blocks.

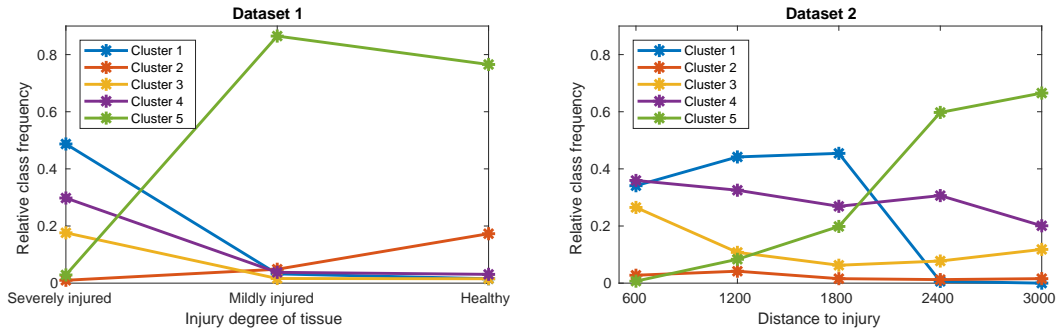


Figure 4.13: Relative frequencies of five classes detected by a Gaussian mixture model in the two datasets. In dataset 2 the classes change smoothly throughout the different distances and appear intermixed.



throughout the different distances of the dataset and appear more intermixed.

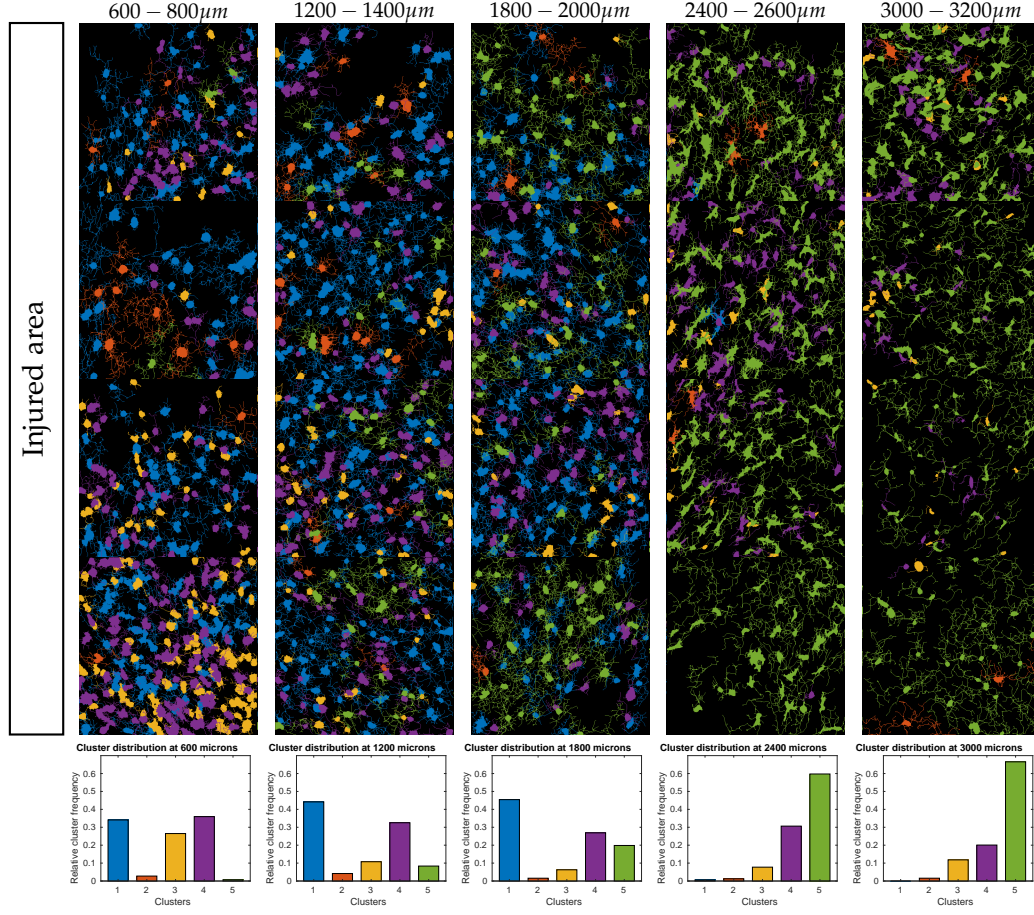


Figure 4.14: Spatial class distribution in dataset 2 obtained by clustering with a Gaussian mixture model with five mixture components. The per-distance class distributions depicted in the histograms do not only allow us to clearly distinguish the different distance levels, but further point out a roughly continuous transition of the class distributions from 600  $\mu\text{m}$  to 3000  $\mu\text{m}$ .

The spatial distribution of five found clusters in dataset 2 is shown in Fig. 4.14, revealing a strong correlation between the distance level of the scans and the classes of the microglia that they contain. This connection is highlighted in histograms showing the relative class distributions at all distance levels. These histograms do not only allow us to clearly distinguish the different distance levels, but further point out a continuous transition of the class distributions from 600  $\mu\text{m}$  to 3000  $\mu\text{m}$ . Especially the clusters 3



(yellow), 4 (green) and 5 (purple) indicate such a progression. Analogous to Fig. 4.11, the class frequencies change less homogeneously between the scans at 1800  $\mu m$  and 2400  $\mu m$ .

As the conducted analysis in this section is exploratory, results are difficult to evaluate. Without ground truth data we cannot unambiguously evaluate the quality obtained clusterings. A method to assess different clusterings numerically is presented in the next section. In regression experiments we try to capture the transition in the class distributions of different scans numerically. The Gaussian mixture model with five components achieves thereby the best scores and outperforms other clusterings obtained by the k-means algorithm, spectral clustering or a different number of clusters.

## 4.5 Regression

The found clusters in dataset 2 imply a continuous morphological change of the processed microglia. In the regression experiments of this section we quantify this transition and use microglial morphology to predict the distance to the injury site. In this way we demonstrate that it is possible to deduce properties of the injury from the morphology of microglia. The distance to the injury site, which is used in our experiments, can be considered from a more abstract point of view as a measure for the injury severity in a processed scan.

Fig. 4.13 and Fig. 4.14 indicate that the frequencies of the five classes that are found by the Gaussian mixture model change smoothly throughout the distance levels of dataset 2. Therefore, we link these two properties in a first regression experiment. Randomly and separately for all distance levels, we partition the cells of dataset 2 into subsets of size 80 and compute the class distributions of these subsets. This can be interpreted as generating subsamples from the histograms of Fig. 4.14 with respectively 80 cells. Then, we train a regressor to predict the distances of the obtained subsets (i.e. histogram subsamples) from their class distributions.

We evaluate the regression performance by measuring the mean absolute error during leave-one-out cross-validation. In order to be resilient against the randomness that is introduced by the sampling step, we repeat all experiments 100 times. The presented error rates are calculated by averaging the individual results of all repetitions.

Using linear regression we achieve a mean absolute regression error of 175 $\mu m$ . Detailed results are summarized in Table 4.3 and visualized in Fig. 4.15 (blue bars). More complex regression models as for example regressors with a polynomial or a radial basis function kernel return comparable or worse results.

Measurement	Global value	Values broken down for all distances				
		600 $\mu m$	1200 $\mu m$	1800 $\mu m$	2400 $\mu m$	3000 $\mu m$
Number of samples	24	6	6	5	4	3
Mean absolute regression error	175 $\mu m$	123 $\mu m$	124 $\mu m$	166 $\mu m$	260 $\mu m$	285 $\mu m$

Table 4.3: Summary of regressing the distance to the injury in dataset 2 from the distribution of five classes that are detected by a Gaussian mixture model.

Considering that the scans have a width of approximately 200  $\mu m$  and are taken with gaps of 400  $\mu m$  between two distance levels, a mean absolute regression error of 175  $\mu m$  yields an acceptable degree of accuracy. Interestingly, the regression errors are not similar for all of the five distance levels. While the samples of the distances until 1800  $\mu m$  are regressed correctly with an mean absolute error of 136  $\mu m$ , the distances of the samples at 2400  $\mu m$  and 3000  $\mu m$  are approximated poorly with an average error of 271  $\mu m$ . A likely reason for that is the discontinuity of the class distribution between 2400  $\mu m$  and 3000  $\mu m$ , which is also visible in Fig. 4.13 and Fig. 4.14.

Obtained regression errors can be also used to numerically evaluate different clustering solutions. The lower the regression error with a specific clustering is, the better it models the morphological transition between activated cells close to the injury and ramified cells further away. Applying the same regression experiment to the classes that we obtain by classification (Fig. 4.10 and Fig. 4.11) returns an mean regression error of approximately 350  $\mu m$ . Likewise, other clustering solutions that we obtain by the k-means algorithm, spectral clustering or a different number of clusters also lead to significantly higher errors.

In our second type of regression experiment, we predict the distances to the injury in dataset 2 in a more direct way. Instead of abstractly capturing microglial morphology by class distributions, regression is performed directly on the features of the extracted microglia. Again, the cells of dataset 2 are divided randomly into batches of 80 cells with the same distance to the injury. Considering these batches as individual samples, we then train a regressor to predict the distances of the batches to the injury from the mean feature values of their cells.

As summarized in Table 4.4, this approach improves the regression accuracies for all distance levels significantly. In total, the error of the linear regressor goes down to 128  $\mu m$ . Fig. 4.15 visualizes the total and the per-distance regression errors of the experiment (yellow bars) and compares them to the results from the first regression experiment. Unlike the class distributions in the first experiment, regressing from mean feature vectors allows us to predict all distance levels with a similar precision. All per-distance error rates are within an acceptable range of 117 – 149  $\mu m$ . In this way we

## 4 Results

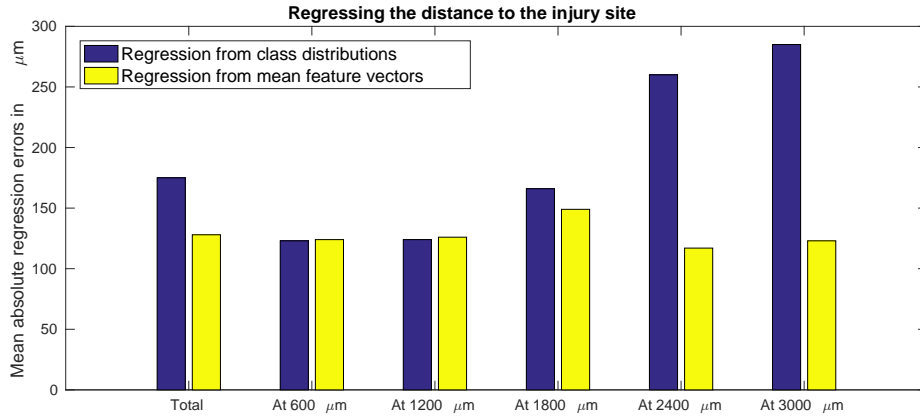


Figure 4.15: Comparison of the total and per-distance regression errors when regressing the distance to the injury from class frequencies of five classes detected by a Gaussian mixture model (blue) and from averaged morphological features that are deduced from the extracted cells (yellow).

show a linear progression in the shape features of the extracted microglia from dataset 2.

Measurement	Global value	Values broken down for all target distances				
		600 $\mu m$	1200 $\mu m$	1800 $\mu m$	2400 $\mu m$	3000 $\mu m$
Number of samples	24	6	6	5	4	3
Mean absolute regression error	128 $\mu m$	124 $\mu m$	126 $\mu m$	149 $\mu m$	117 $\mu m$	123 $\mu m$

Table 4.4: Summary of regressing the distance to the injury in dataset 2 from averaged morphological features that are deduced from the extracted microglia cells.

In summary, the conducted regression experiments reveal a linear progression of microglial morphology from scans close to the injury to scans further away. We can quantify this transition in two way: Firstly by analyzing the distribution of morphological classes that are obtained by clustering techniques; secondly by regressing from averaged morphological features that are directly deduced from extracted microglia.

## 5 Conclusion and Outlook

In this thesis, we introduce an accurate and robust framework to automatically trace microglia from volumetric confocal microscopy scans. We segment somas by using filters that detect isotropic and intensity-based dense objects. We implement process extraction as a multi-scale regression task, in which ground truth training data is automatically obtained by filters. In this way we overcome the obstacle of manually acquiring microglial tracings to train from. Finally, after partitioning and pruning found structures we obtain extractions of individual microglia cells. Our approach provides very accurate cell extractions by visual evaluation and can scale up to large amounts of input data in a straightforward way.

In the second part of the thesis, we demonstrate that cells traced by the presented framework are applicable to a broad range of analyses. Analyzing nearly 3000 extracted cells by classification, clustering and regression experiments reveals interesting insights about microglial morphology. We successfully classify processed microglia into "ramified", "partially activated" and "fully activated" cells. Furthermore we show that the three classes are linearly separable. Applying clustering techniques we detect new morphological classes of microglia. We demonstrate that their distribution in the tissue can be related to the distance from the injury site or similarly, the severity of the injury. Using regression we quantify a roughly linear morphological change between activated microglia close to an injury and more ramified cells further way. That means we can use microglial morphology as a predictor of injury severity by regressing the distance to the brain injury.

Our work can be extended in various ways. Addressing the limitations explained in Section 4.2, a first option is to improve the presented tracing framework. In particular soma segmentation has potential for improvement that considerably impacts the success of subsequent analyses. As soma segmentation is implemented using a filter-based approach, which only relies on local image information and requires its parameters to be adjusted repeatedly for different scans, soma detection is the least robust component of the introduced framework. Segmenting somas by a learning method could reduce the importance of parameter tuning and thus improve accuracy as well as robustness. Secondly, the presented results motivate to continue and scale up studies on microglial

morphology. The found insights of this thesis suggest to extend the simple regression of the injury distance to more elaborate tasks. Modern tissue clearing and imaging techniques make it possible to obtain scans of a whole brain in a quality, that can be processed by the introduced tracing framework. Thereby, microglial morphology can be quantified in larger brain regions, which might give new insights about mechanisms in the brain.

Another possible course of research is to extend the presented microglia extraction framework by deep learning techniques. Instead of tracing complete microglia cells, a "deep" classifier (i.e. a convolutional neural network) might already be able to distinguish different types of microglia based on the image patch around the cell.

The overall aim our research about microglia is the use of the cells as biological markers. This could yield completely new ways for investigating brain functionality and predicting brain injuries, diseases or stroke. In this thesis we present initial investigations and findings to achieve that.

# Bibliography

- [ASI16] L. Acciai, P. Soda, and G. Iannello. “Automated Neuron Tracing Methods: An Updated Account.” In: *Neuroinformatics* 14.4 (Oct. 2016), pp. 353–367.
- [CB17] M. Colonna and O. Butovsky. “Microglia Function in the Central Nervous System During Health and Neurodegeneration.” In: 35 (Feb. 2017).
- [DH16] F. Diego and F. A. Hamprecht. “Structured Regression Gradient Boosting.” In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016, pp. 1459–1467.
- [Din+17] Y. Ding, M. C. Pardon, A. Agostini, H. Faas, J. Duan, W. O. C. Ward, F. Easton, D. Auer, and L. Bai. “Novel Methods for Microglia Segmentation, Feature Extraction, and Classification.” In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 14.6 (Nov. 2017), pp. 1366–1377.
- [Fra+98] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. “Multiscale vessel enhancement filtering.” In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI’98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings*. Ed. by W. M. Wells, A. Colchester, and S. Delp. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 130–137.
- [HD14] M. Holzer and R. Donner. “Over-Segmentation of 3D Medical Image Volumes based on Monogenic Cues.” In: *Proceedings of the CVWW’14*. Ed. by Z. Kúkelová and J. Heller. 2014, pp. 35–42.
- [HR97] T. Hildebrand and P. Rügsegger. “A new method for the model-independent assessment of thickness in three-dimensional images.” In: *Journal of Microscopy* 185.1 (1997), pp. 67–75.
- [KL12] G. Kutyniok and D. Labate. *Shearlets: Multiscale Analysis for Multivariate Data*. Birkhäuser Basel, 2012.
- [KL16] C. B. Kayasandik and D. Labate. “Improved detection of soma location and morphology in fluorescence microscopy images of neurons.” In: *Journal of Neuroscience Methods* 274 (2016), pp. 61–70.

- [KSM10] D.-J. Kroon, C. H. Slump, and T. J. J. Maal. "Optimized Anisotropic Rotational Invariant Diffusion Scheme on Cone-Beam CT." In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part III*. Ed. by T. Jiang, N. Navab, J. P. W. Pluim, and M. A. Viergever. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 221–228.
- [LKC94] T.-C. Lee, R. L. Kashyap, and C.-N. Chu. "Building Skeleton Models via 3-D Medial Surface/Axis Thinning Algorithms." In: *CVGIP: Graph. Models Image Process.* 56.6 (Nov. 1994), pp. 462–478.
- [Mat] Matlab. *boundary function*. <https://de.mathworks.com/help/matlab/ref/boundary.html>. Accessed: 2018-01-03.
- [Meg+15] M. Megjhani, N. Rey-Villamizar, A. Merouane, Y. Lu, A. Mukherjee, K. Trett, P. Chong, C. Harris, W. Shain, and B. Roysam. "Population-scale three-dimensional reconstruction and quantitative profiling of microglia arbors." In: *Bioinformatics* 31.13 (2015), pp. 2190–2198.
- [MTF17] A. Mosinska, J. Tarnawski, and P. Fua. "Active Learning and Proofreading for Delineation of Curvilinear Structures." In: *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017 - 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II*. 2017, pp. 165–173.
- [OF02] S. Osher and R. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Applied Mathematical Sciences. Springer New York, 2002.
- [Ozc+15] B. Ozcan, P. Negi, F. Laezza, M. Papadakis, and D. Labate. "Automated Detection of Soma Location and Morphology in Neuronal Network Cultures." In: *PLOS ONE* 10.4 (Apr. 2015), pp. 1–22.
- [Pan+16] C. Pan, R. Cai, F. P. Quacquarelli, A. Ghasemigharagoz, A. Lourbopoulos, P. Matryba, N. Plesnila, M. Dichgans, F. Hellal, and A. Ertürk. "Shrinkage-mediated imaging of entire organs and organisms using uDISCO." In: *Nature Methods* 13 (Aug. 2016).
- [Pay+16] C. Payer, M. Pienn, Z. Bálint, A. Shekhovtsov, E. Talakic, E. Nagy, A. Olschewski, H. Olschewski, and M. Urschler. "Automated integer programming based separation of arteries and veins from thoracic CT images." In: *Medical Image Analysis* 34 (2016), pp. 109–122.
- [PD] E. Pekalska and B. Duin. *Pattern Recognition Tools*. <http://37steps.com/>. Accessed: 2017-12-24.

- [Pol] S. Polavaram. *L-measure help page*. <http://cng.gmu.edu:8080/Lm/help/index.htm>. Accessed: 2018-01-10.
- [RAM16] M. Rempfler, B. Andres, and B. H. Menze. "The Minimum Cost Connected Subgraph Problem in Medical Image Analysis." English. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Lecture Notes in Computer Science. Springer International Publishing, 2016.
- [Ras00] C. E. Rasmussen. "The Infinite Gaussian Mixture Model." In: *In Advances in Neural Information Processing Systems 12*. MIT Press, 2000, pp. 554–560.
- [Rem+15] M. Rempfler, M. Schneider, G. D. Ielacqua, X. Xiao, S. R. Stock, J. Klohs, G. Székely, B. Andres, and B. H. Menze. "Reconstructing cerebrovascular networks under local physiological constraints by integer programming." In: *Medical Image Analysis* 25.1 (2015), pp. 86–94.
- [Rob+16] D. Robben, E. Türetken, S. Sunaert, V. Thijs, G. Wilms, P. Fua, F. Maes, and P. Suetens. "Simultaneous segmentation and anatomical labeling of the cerebral vasculature." In: *Medical Image Analysis* 32 (2016), pp. 201–215.
- [Rof+17] G. Roffo, S. Melzi, U. Castellani, and A. Vinciarelli. "Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach." In: *CoRR* abs/1707.07538 (2017).
- [Sir+16] A. Sironi, E. Türetken, V. Lepetit, and P. Fua. "Multiscale Centerline Detection." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38.7 (July 2016), pp. 1327–1341.
- [SPA08] R. Scorcioni, S. Polavaram, and G. A. Ascoli. "L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies." In: *Nature Protocols* 3.5 (Apr. 2008), pp. 866–876.
- [Tür+13] E. Türetken, C. Becker, P. Glowacki, F. Benmansour, and P. Fua. "Detecting Irregular Curvilinear Structures in Gray Scale and Color Imagery Using Multi-directional Oriented Flux." In: *2013 IEEE International Conference on Computer Vision*. Dec. 2013, pp. 1553–1560.
- [Tür+16] E. Türetken, F. Benmansour, B. Andres, P. Glowacki, H. Pfister, P. Fua, undefined, undefined, undefined, and undefined. "Reconstructing Curvilinear Networks Using Path Classifiers and Integer Programming." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38.12 (2016), pp. 2515–2530.



- [Wan+17] Q. Wang, S. Wang, X. Zhu, T. Liu, Z. Humphrey, V. Ghukasyan, M. Conway, E. Scott, G. Fragola, K. Bradford, M. J. Zylka, A. Krishnamurthy, J. L. Stein, and G. Wu. "Accurate and High Throughput Cell Segmentation Method for Mouse Brain Nuclei Using Cascaded Convolutional Neural Network." In: *Patch-Based Techniques in Medical Imaging: Third International Workshop, Patch-MI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Proceedings*. Ed. by G. Wu, B. C. Munsell, Y. Zhan, W. Bai, G. Sanroma, and P. Coupé. Cham: Springer International Publishing, 2017, pp. 55–62.
- [Xie] L. Xie. *Hysteresis thresholding for 3D images*. <https://de.mathworks.com/matlabcentral/fileexchange/44648-hysteresis-thresholding-for-3d-images--or-2d->. Accessed: 2017-12-23.
- [Xu+16] Y. Xu, M. Megjhani, K. Trett, W. Shain, B. Roysam, and Z. Han. "Unsupervised Profiling of Microglial Arbor Morphologies and Distribution Using a Nonparametric Bayesian Approach." In: *IEEE Journal of Selected Topics in Signal Processing* 10.1 (Feb. 2016), pp. 115–129.