Neuroinformatics (2019) 17:147–161
https://doi.org/10.1007/s12021-018-9388-7

ORIGINAL ARTICLE

Morphological Neuron Classification Based on Dendritic Tree Hierarchy

Evelyn Perez Cervantes[1] · Cesar Henrique Comin[2] · Roberto Marcondes Cesar Junior[1] · Luciano da Fontoura Costa[3]

Published online: 14 July 2018

Abstract

The shape of a neuron can reveal interesting properties about its function. Therefore, morphological neuron characterization can contribute to a better understanding of how the brain works. However, one of the great challenges of neuroanatomy is the definition of morphological properties that can be used for categorizing neurons. This paper proposes a new methodology for neuron morphological analysis by considering different hierarchies of the dendritic tree for characterizing and categorizing neuronal cells. The methodology consists in using different strategies for decomposing the dendritic tree along its hierarchies, allowing the identification of relevant parts (possibly related to specific neuronal functions) for classification tasks. A set of more than 5000 neurons corresponding to 10 classes were examined with supervised classification algorithms based on this strategy. It was found that classification accuracies similar to those obtained by using whole neurons can be achieved by considering only parts of the neurons. Branches close to the soma were found to be particularly relevant for classification.

Keywords Neuron · Morphological reconstruction · Morphometry · Dendritic arborization · Dendritic tree · Digital neuronal reconstruction · Data sharing · Morphological classification · Supervised classification · Feature selection

Introduction

The problem of classifying neurons has been studied since the beginnings of neuroscience, even before the advent of electron microscopy. In particular, Cajal and Azoulay (1955) made a functional interpretation of the nervous system structures and demonstrated the morphological and functional independence of cells using the Golgi stain methodology invented in 1873 by Camillo Golgi. Based on the knowledge provided by Cajal, many researches started developing more quantitative attempts to understand how the brain works. A key aspect is to provide a census of the different types or classes of nerve cells as well as their functionality. However, even now there is no agreement about the total number of neurons, even less about how they should be classified. Classifying neurons, in the strict sense, is the process of separating neurons into groups or classes based on some common characteristics (Armañanzas and Ascoli 2015). This process, which appears to be simple, is in fact far from being fully resolved. Even with the current technological advances, there are limitations from the data acquisition process, disagreements of how to define a neuron discretely as well as on the selection of the most pertinent features used to define the neuron class in the presence of only molecular information. As a consequence, the number of neuronal classes remains unknown (Bernard et al. 2009; Sümbül et al. 2014a). Neurons can be characterized by their morphology, physiology and biochemistry. These characteristics vary substantially for the same neuron type, depending, among other possibilities, on location or function (Barbosa et al. 2003; Ding and Glanzman 2011). Neuronal morphology is relatively easier to access, which makes it a key feature in the study of neurons, and is also directly related to neuronal functionality. In recent

✉ Evelyn Perez Cervantes
eperezc@ime.usp.br

Cesar Henrique Comin
chcomin@gmail.com

Roberto Marcondes Cesar Junior
cesar@ime.usp.br

Luciano da Fontoura Costa
ldfcosta@gmail.com

[1] Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil

[2] Department of Computer Science, Federal University of São Carlos, São Carlos, Brazil

[3] São Carlos Institute of Physics, University of São Paulo, PO Box 369, 13560-970, São Carlos, SP, Brazil

years, neuromorphological information has been shared by researchers using 3D digital reconstructions available in public access databases. Such datasets also present some degree of variability (e.g. with respect to the type of experimental procedure) which makes the process of choosing good features even more challenging (Armañanzas and Ascoli 2015; Bazán and Lolley 2013; Bota and Swanson 2007; Comin and da Fontoura Costa 2013; DeFelipe 2001; López-Cruz et al. 2014). On the other hand, they represent a valuable resource to promote studies and to develop new methodologies.

A neuron is composed of three main structures: soma, axon and dendrites. Each of them has important functions in the transport of the nerve impulse (Kandel et al. 2000). Over the years several proposals have emerged to classify neurons based on the analysis of these structures (Armañanzas and Ascoli 2015; Halavi et al. 2012; da Costa and Velte 1999; Mottini et al. 2014; Ruz and Schultz 2014; Sharpee 2014; Sümbül et al. 2014b). Armañanzas and Ascoli (2015) reported a systematic review of the neuron classification methods using machine learning. Uylings et al. (1989) and Uylings and van Pelt (2002) proposed several measures to quantify dendritic arborizations based on their shapes.

Among the more recent attempts to categorize neurons according to their morphological properties (Uji et al. 1995; McGarry et al. 2010) there are some approaches exploring the neural tree branch density (Teeter and Stevens 2011). Sümbül et al. (2014b) proposed a neuron classification method based on the relative position of the dendritic tree to identify molecularly defined neuron types. The layer-specific distribution and length together with tree density have been described as key parameters to define clusters by Hosp et al. (2014). The density overlap between dendritic trees were proposed by Zhao and Plaza (2014) to express the similarity between neurons. Using the NeuroMorpho.org (Ascoli et al. 2007) dataset and measurements calculated with the L-measure (Scorcioni et al. 2008) software Lu et al. (2015) adopted an harmonic co-clustering algorithm based on diffusion properties. Santana et al. (2013) proposed a clustering algorithm called affinity propagation to identify neuronal types by combining a set of morphological and physiological features. Bayesian networks and clustering algorithms were also applied to categorize interneurons by Mihaljević et al. (2015). Gillette and Ascoli (2015) made a different proposal, which encodes axonal and dendritic trees into sequences of characters representing bifurcations in order to enable similarity measurements. Such measurements have been originally defined to compare gene sequences and allow to perform clustering analysis and to identify tree types.

On the majority of the discussed cases, the complete structures of the neurons was used, i.e. complete axons, whole dendritic tree or whole neuron cell. However, these structures can be partitioned into subcomponents corresponding to sets of axonal or dendritic segments. Because different parts of a neuron have distinct functionality, such a subdivision of the neuronal morphology paves the way to extending and refining studies focusing on the relationship between neuronal shape and function. In addition, the extension and branching pattern of the dendritic structures is directly related to neuronal interconnectivity, especially regarding its locality or long range projections. For instance, dendritic tree shape can have profound impacts on neuronal function (Magee 2000; Rall 1967), including effects on synaptic integration (Mainen and Sejnowski 1996). Thus, it becomes particularly interesting to study neuronal connectivity and shape, from the perspective of partitioning neuronal cells into several subpatterns. This consists in the main objective of the present article. The problem is studied by dismantling Neurons into smaller parts and performing supervised neuron classification to verify the accuracy loss when using only partial information about the neuronal structure. Particularly, the influence of dendritic segments at distinct distances from the soma on the classification is quantified.

The paper starts by presenting the methodology and the data used in the study (Section "Materials and Methods"), which includes the different approaches considered for partitioning the dendritic arborization, a description of the morphological measurements and the methodologies used for classifying the neurons. Next, the results are presented for each neuron partitioning approach (Section "Experimental Results"), followed by a discussion of the main results (Section "Conclusion").

Materials and Methods

Proposed Method

Figure 1 presents an overview of the proposed method, which includes three main steps. The first step implements neuron decomposition. The input digital neuronal reconstruction is represented as an abstract tree structure, which is then decomposed using different strategies introduced in this paper. In the second step, a classification procedure is employed. Shape measures are extracted from the decomposed dendritic trees (different hierarchical levels of decomposition are investigated). The measures are then normalized, subjected to a feature selection process, and used to train a classifier. In the final step, results obtained using whole neurons and those obtained using the proposed decomposition methods are compared in order to determine particularly relevant parts of neurons for morphological classification. In the following sections we describe each step of the methodology in more detail.

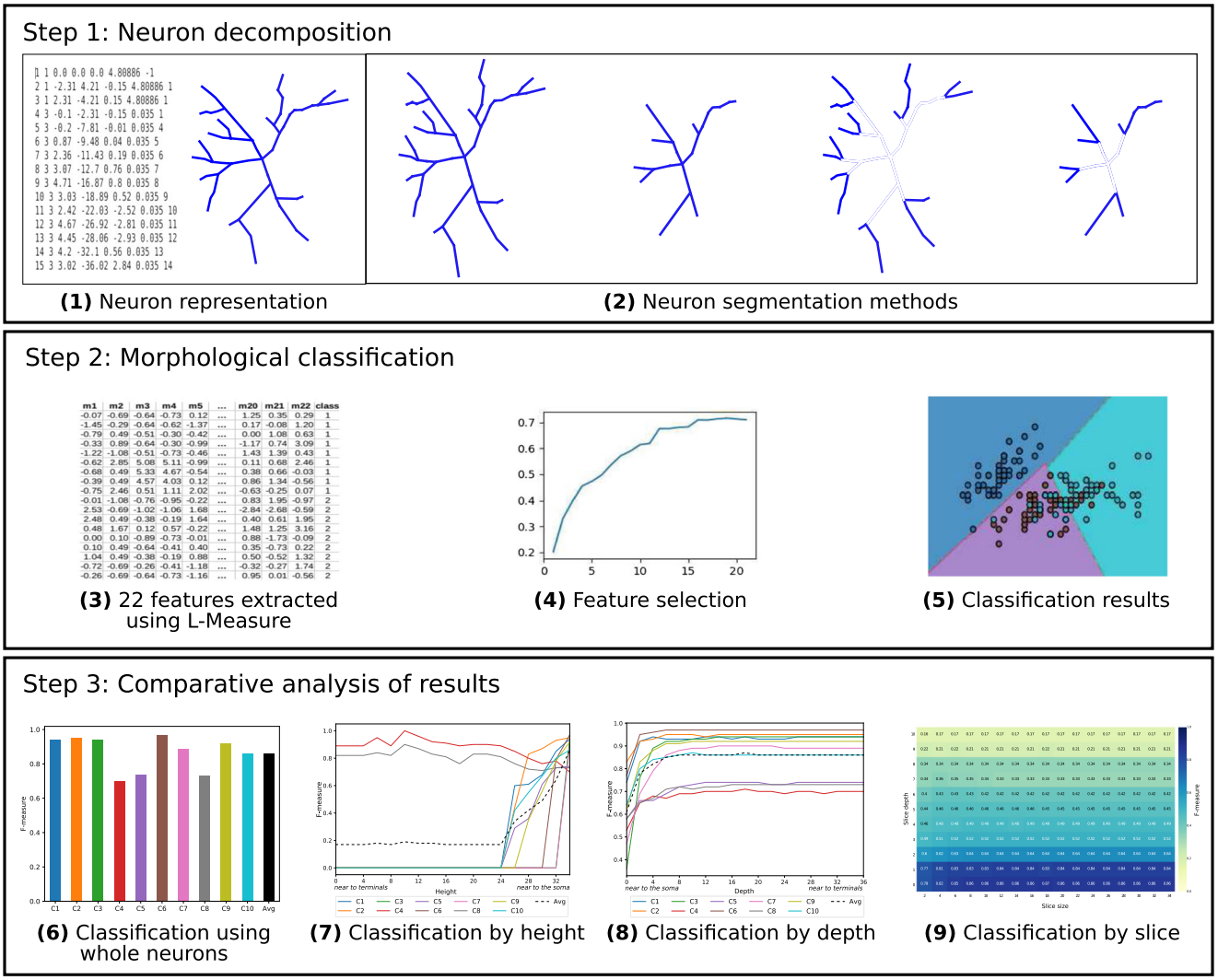


Fig. 1 Overview of the proposed method, divided into three main steps: neuron decomposition, morphological classification and comparative analysis of results. In the neuron decomposition step, a neuron is dismantled into distinct hierarchies, composed of branches at specific topological distances from the soma. The morphological classification step concerns the classification of the dismantled neurons using the classes defined for the whole neurons. In the final step, the classification accuracy of the neuron classes when considering different dismantling strategies is compared

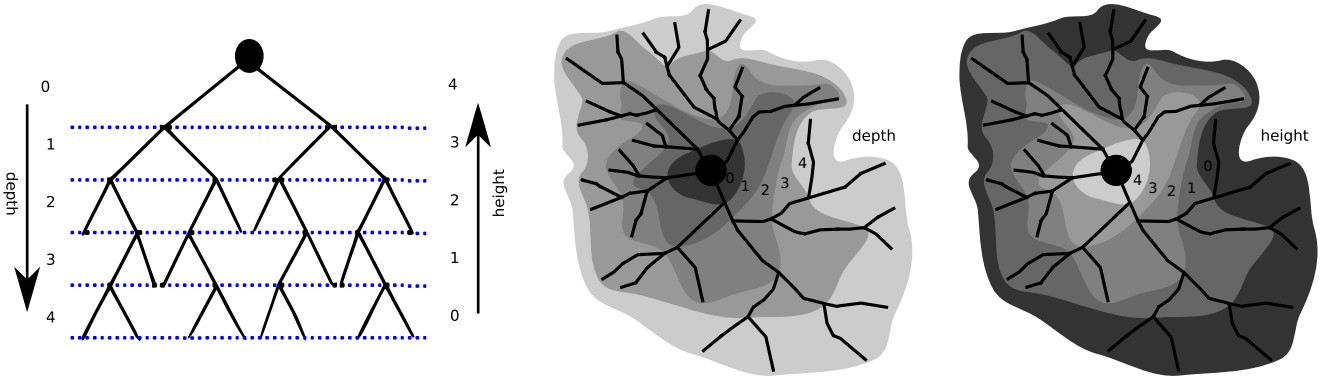


Fig. 2 Illustration of the height and depth of branches in a dendritic tree. The depth of a branch represents the number of bifurcations along the path going from the root to the branch. The height is the complement of the depth. Note that, according to our definition, terminal branches do not necessarily have height equal to zero

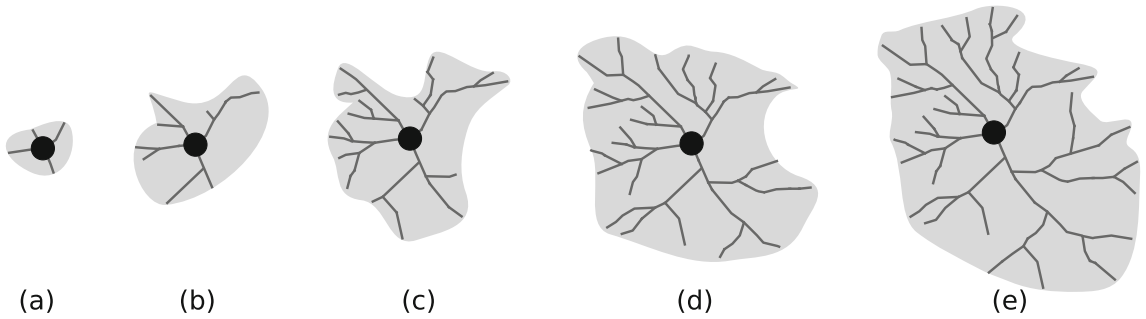


Fig. 3 The 4 new trees created by the process called dismantle by depth from a dendritic tree with depth 4. **a** tree with depth 0, **b** tree with depth 1, **c** tree with depth 2, **d** tree with depth 3, **e** original tree with depth 4

Neuron Decomposition

One focus of interest of the present work is in assessing the effect of using different neuronal parts for morphological classification. The dendritic arborization can be understood as a set of trees. It is hence possible to decompose these structures into subtrees according to distinct depth levels.

The depth of a branch is defined as the number of bifurcations along the path from the root (i.e. the soma) to the branch. Thus, the maximum depth (max-depth) will be reached at the most distant branch from the root. The height of a node is defined as the difference between the max-depth and the depth of the current node. Thus, the height is zero at the most distant branch from the root and achieves its maximum values (max-height) at the root. Figure 2 shows how depth and height are related in this work. Note that our definition of depth and height is different from the standard definition of depth and height of a node from graph theory.

The present work develops from the structure proposed in Torben-Nielsen (2014) in order to represent the dendritic tree using an abstract tree structure. With this structure, it is possible to represent all the neuronal components. The soma is stored as the root, and each dendrite leaving the soma is represented by a binary tree attached to the root. Additionally, two fields are used to store the current depth and height for each compartment. The terms used to describe the components of neurons are supported by L-measure (Scorcioni et al. 2008) tool.

All neurons to be analyzed are initially represented by the tree structure prior to being decomposed. Three decomposition strategies are proposed: 1- decomposition by depth, 2- decomposition by height and 3- decomposition by slice. They are described as follows.

Decomposition by Depth

In the first case, neuron decomposition is guided by tree depth. A set of subtrees is thus created from a tree, one for each level. This process copies the original dendritic tree, keeping the soma and the original branches up to a given depth level. All branches below the given depth level are removed. The process starts at the root and runs towards the leaves. Hence, the tree with depth 0 is composed by the original soma and all branches before the first bifurcation (branches with depth level 0 in the original tree). Level 1 is composed by the tree with depth 0 together with all branches before the second bifurcation (branches with depth level 1 in the original tree) and so on until the leaves are reached. Figure 3 shows the 4 new trees created by the process from a dendritic tree with depth 4. It can be noticed that this process basically accumulates branches from the soma to dendritic leaves in each new iteration. The last tree in Fig. 3e is the same as the original.

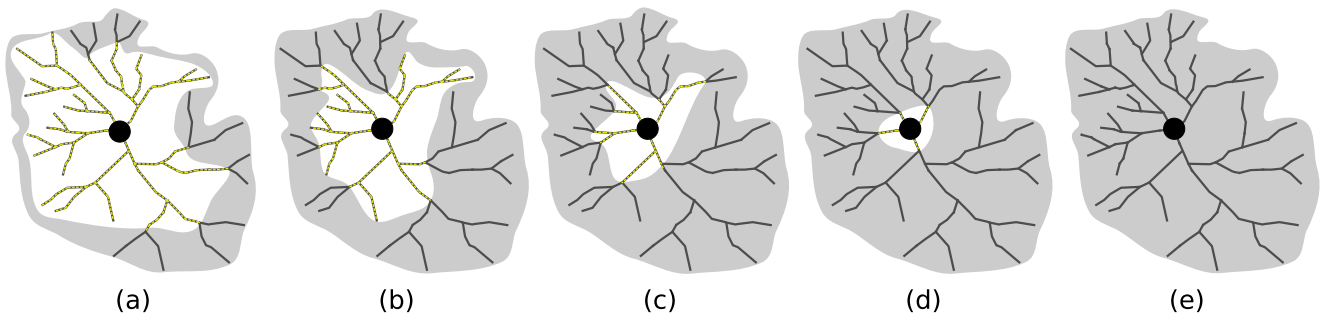


Fig. 4 The 4 new trees created by the process called dismantle by height from a dendritic tree with depth 4, the branches in yellow were disregarded. **a** tree with height 0, **b** tree with height 1, **c** tree with height 2, **d** tree with height 3, **e** original tree with height 4

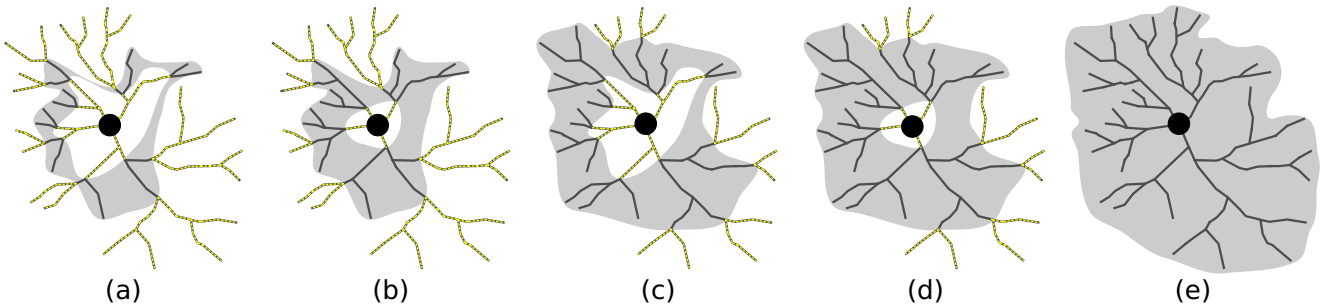


Fig. 5 Subtrees generated by the method of decomposition by slice, **a** slice between levels 2 and 3, **b** slice between levels 1 and 3, **c** slice between levels 2 and 4, **d** slice between levels 1 and 4, **e** the original neuron or a neuron with slice between levels 0 and 5. The branches in yellow are disregarded

Decomposition by Height

The decomposition by height process is guided by the height of the tree from the leaves, rather than depth. By starting with the original neuron, all the branches with height level greater than a given value are removed. Therefore, the tree with height 0 is composed by the original soma and the leaves of the original tree. The tree with height 1 is composed by the tree with height 0 and the branches of height level 1 in the original tree and so on. Figure 4 shows the 4 new trees created by a height decomposition.

Decomposition by Slice

The decomposition by slice mixes both the depth and the height approaches. This method takes as parameters a lower and an upper depth to isolate a slice of the original tree. In order to create the subtree, the soma from the original tree is conserved and only branches between the inputed limits are considered. It is important to notice that this decomposition method is more generic than those described previously. Decomposition by depth is equivalent to decomposition by slice where the lower depth is set to 0 and the upper depth varies between 0 to the max-depth of the tree. On the other hand, decomposition by height is a specific case where the upper depth is fixed with the max-height of the tree and the lower depth varies between 0 and the max-height. Figure 5 shows some examples of subtrees created by slicing a dentritic tree.

Morphometric Measurements

The L-Measure(Scorcioni et al. 2008) software was used in this work to compute quantitative morphological measurements from neuronal digital reconstructions. The set of 22 features shown in Table 1 were chosen. A complete description of the measures here employed can be found

Table 1 Morphological measurements considered for neuronal characterization

Nº	Measure description	Nº	Measure description
1	Soma surface area	12	Maximum Euclidean distance between the soma and leaves
2	Number of stems attached to the soma		
3	Number of bifurcations	13	Maximum path distance between the soma and leafs
4	Number of branches		
5	Neuronal width, difference between maximum and minimum on the x-coordinates	14	Maximum branch order
		15	Average branch path length
6	Neuronal height, difference between maximum and minimum on the y-coordinates	16	Average contraction
		17	Total fragmentation
7	Neuronal depth, difference between maximum and minimum on the z-coordinates	16	Average topological asymmetry
		18	Total fragmentation
8	Average branch diameter	19	Average Rall's power
9	Total arborization/tree length	20	Average local bifurcation angle
10	Total arborization/tree surface area	21	Average remote bifurcation angle
11	Total arborization/tree volume	22	Fractal dimension

in Ascoli (2002), the L-Measure home page[1] as well as the FARSIGHT Project wiki.[2]

It is important to note that the previously defined depth and height measurements are different from the morphological measurements calculated by L-measure. While the former ones are calculated for each branch on the tree structure that was used in this work to represent the neurons, the latter ones are calculated for the whole tree in the coordinate axes that the tool uses to represent the neurons.

Morphological Neuron Classification and Performance Assessment

The morphometric measures underwent a z-score transformation so that each normalized feature presents zero mean and unitary standard deviation. Then, the RFECV method (recursive feature elimination and cross-validated selection) (Guyon et al. 2002) was used. This algorithm performs a recursive feature elimination method with automatic tuning of the number of features. The optimal number of features is selected according with a cross validation procedure which uses the classification accuracy as the scoring measure.

For the classification step, the multiclass C-Support Vector Classifier (SVC) based on LIBSVM (Chang and Lin 2011) was adopted with a linear kernel and default parameters for all tests. Note that since the objective of this work is not to obtain the best possible accuracy, the choice of classifier is not critical for the analysis. A 10-fold cross-validation has been adopted to estimate classification errors based on the original ground-truth (i.e. each neuron class in the database). The precision and recall measures were calculated for the performance assessment, since the objective is to evaluate how precise and robust the classifier is. The F-measure was adopted because it combines both the precision and recall properties. The formula used to calculate the F-measure is:

$$\text{F-measure} = \frac{2 * (\text{precision} * \text{recall})}{(\text{precision} + \text{recall})}$$

Neuron Data

The neurons used in the analysis were obtained from the NeuroMorpho.Org (Ascoli et al. 2007) dataset. This repository provides three dimensional neuronal reconstructions of a large number of neuronal shapes coming from different species, body regions and experimental procedures (Cannon et al. 1998; Parekh and Ascoli 2013). Neurons are made available in the swc text file format. These files contain a description of neuronal structure based on sequences of segments called compartments. Each line in the file describes a compartment using 7 fields: a unique identifier, the type of neuronal compartment, x, y and z coordinates, the radius and a identifier indicating the parent compartment. The types of neuronal compartments considered are: 0 to describe undefined type, 1 for soma type, 2 for axon, 3 for basal dendrite, and 4 for apical dendrite. All compartments have a unique parent.

The present work considers data organized into 6 datasets. The experiments 1 to 4 contain 5000 digital reconstructions belonging to 10 classes presented in Table 2. The selected classes include pyramidal and interneuron cells belonging to chimpanzee, mouse, rat and humans. We also considered granule and medium spiny neurons belonging to different species. All the chosen classes contain at least 500 neurons. For each class, the same number of neurons (500) was randomly selected with uniform probability. Figure 6 illustrates digital reconstructions of each of the 10 neuron classes considered in the first four experiments. The fifth experiment contemplates 224 neurons belonging to mouse and rat species. Finally in experiment 6, 162 neurons belonging to mouse species were examined.

Experimental Results

The morphological measures were computed for the decomposed trees in order to determine how the depth and height of dendritic tree influences neural classification. Note that, since the soma is kept for all levels, some measures involving this structure resulted in the same value. In the following we divide the results according to

Table 2 The neuronal classes considered in this study

Id	Neuron class
C1	chimpanzee principal pyramidal
C2	human principal pyramidal
C3	mouse principal ganglion
C4	mouse principal pyramidal
C5	rat interneuron gabaergic
C6	rat interneuron nitrergic
C7	rat principal pyramidal hippocampus
C8	rat principal pyramidal neocortical
C9	various principal granule
C10	various principal medium spiny

[1] http://cng.gmu.edu:8080/Lm/

[2] http://farsight-toolkit.org/wiki/L_Measure_functions

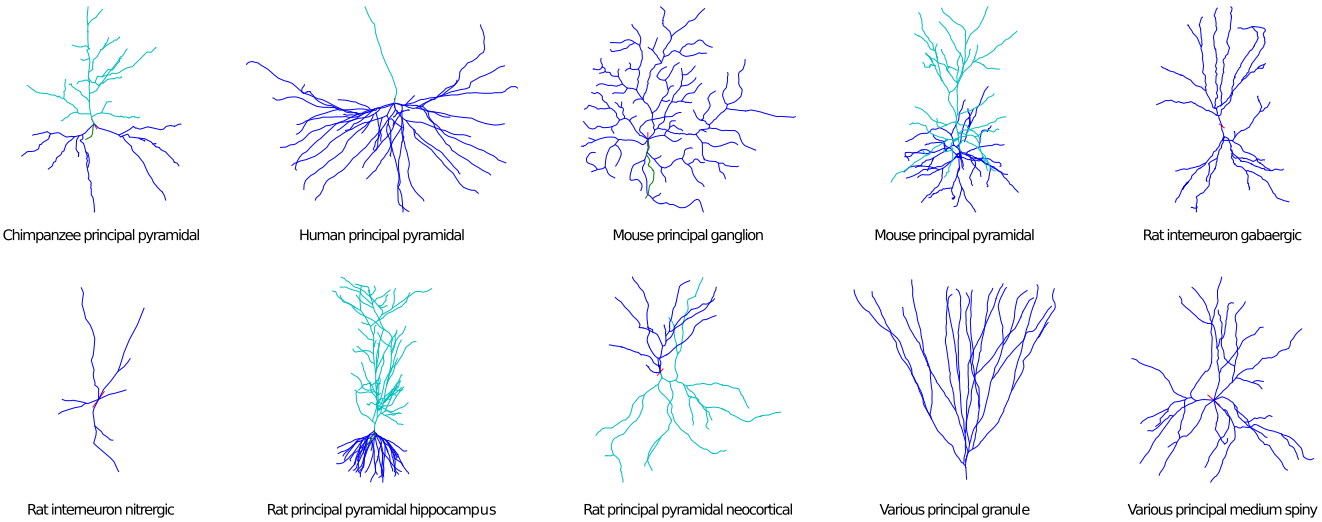


Fig. 6 Digital reconstructions illustrating each one of the 10 neuron classes considered in the first four experiments

the approach used for selecting the relevant branches for classification.

Experiment 1: Whole Neurons

In order to verify the classification accuracy when whole neurons are considered, the 22 features were computed over the entire dendritic tree of the neurons. After the normalization and RFECV feature elimination method, the SVC classifier was used to asses the separability between classes. The F-measure obtained for each class is shown in Fig. 7. The result indicates that the neurons can be classified into the 10 considered classes with a typical F-measure value of 0.86.

Using the Principal Component Analysis (PCA) the dimensionality of the 22-dimensional space was reduced to three dimensions. The weights of each of the three principal components is presented in Table 3. The largest weights are related to measures associated with neuronal size (e.g. path distance, Euclidean distance and height).

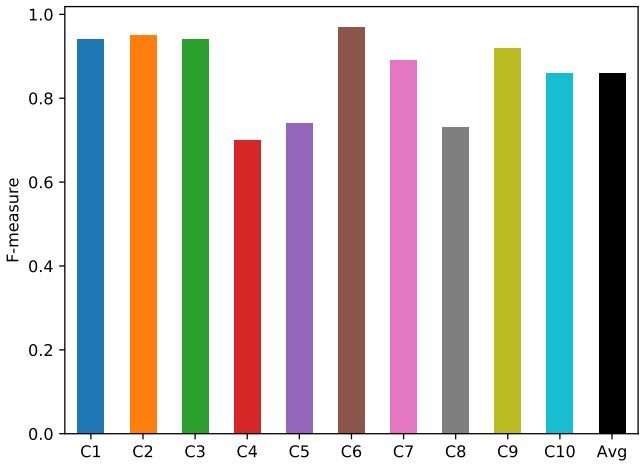


Fig. 7 F-measures obtained for each class after morphological classification. The average F-measure for all classes is shown as a black bar. Whole neurons were used in this analysis

Table 3 Weights of each measurement on the first three principal components obtained from PCA

	PC1	PC2	PC3
Variance explanation	28%	14%	10%
Soma_Surface	0.1	−0.33	0.35
N_stems	0.09	0.08	−0.1
N_bifs	0.26	0.31	0.25
N_branch	0.26	0.31	0.25
Width	0.06	−0.03	−0.09
Height	0.31	−0.1	−0.2
Depth	0.22	−0.09	−0.28
Diameter	0.02	−0.3	0.36
Length	0.36	0.03	−0.07
Surface	0.29	−0.2	0.27
Volume	0.13	−0.35	0.38
EucDistance	0.33	−0.11	−0.23
PathDistance	0.35	−0.06	−0.21
Branch_Order	0.28	0.26	0.08
Branch_pathlength	0.22	−0.32	−0.26
Contraction	−0.19	−0.15	0.04
Fragmentation	0.22	0.19	0.17
Partition_asymmetry	−0.01	0.08	0.12
Rall_Power	0.09	−0.24	0.05
Bif_ampl_local	−0.07	0.05	−0.11
Bif_ampl_remote	0.02	0.28	0.17
Fractal_Dim	0.08	0.16	0.05

The first row of the table shows the percentage of variance explained by each component

Experiment 2: Decomposition by Depth

The second experiment consists in classifying the neurons based on the depth decomposition. This experiment was performed in order to examine how dendritic tree depth influences the classification performance. In other words, we are trying to analyze how much value each depth level adds to the morphological classification. Figure 8a shows the F-measure obtained as a function of depth. Depth 0 represents the soma and the branches with depth 0. Depth 1 represents trees composed of the soma and the branches until the first bifurcation and so on up to depth 35 (largest neurons in the dataset). The results indicate that, as expected, using incomplete neurons can impact the classification performance. Strikingly, around depth 4 the F-measure reaches values that are close to those obtained when using the entire dendritic tree. This indicates that the vast majority of the information used for categorizing the neurons is provided by branches belonging to the first few hierarchical levels. Also, the change in F-measure varies depending on the class. It is possible to observe that the classes mouse pyramidal, rat interneuron gabaergic and rat pyramidal neocortical present the lowest accuracy for the majority of depth values. These three neuron classes yielded lower overall classification accuracy, as can be seen in Fig. 7. This is probably a consequence of the diversity of sub-classes and neuronal geometry observed in this three cases. Figure 8b shows the F-measures normalized by the F-measure obtained using whole neurons. The normalized F-measurements become very close to one for depths larger than 8.

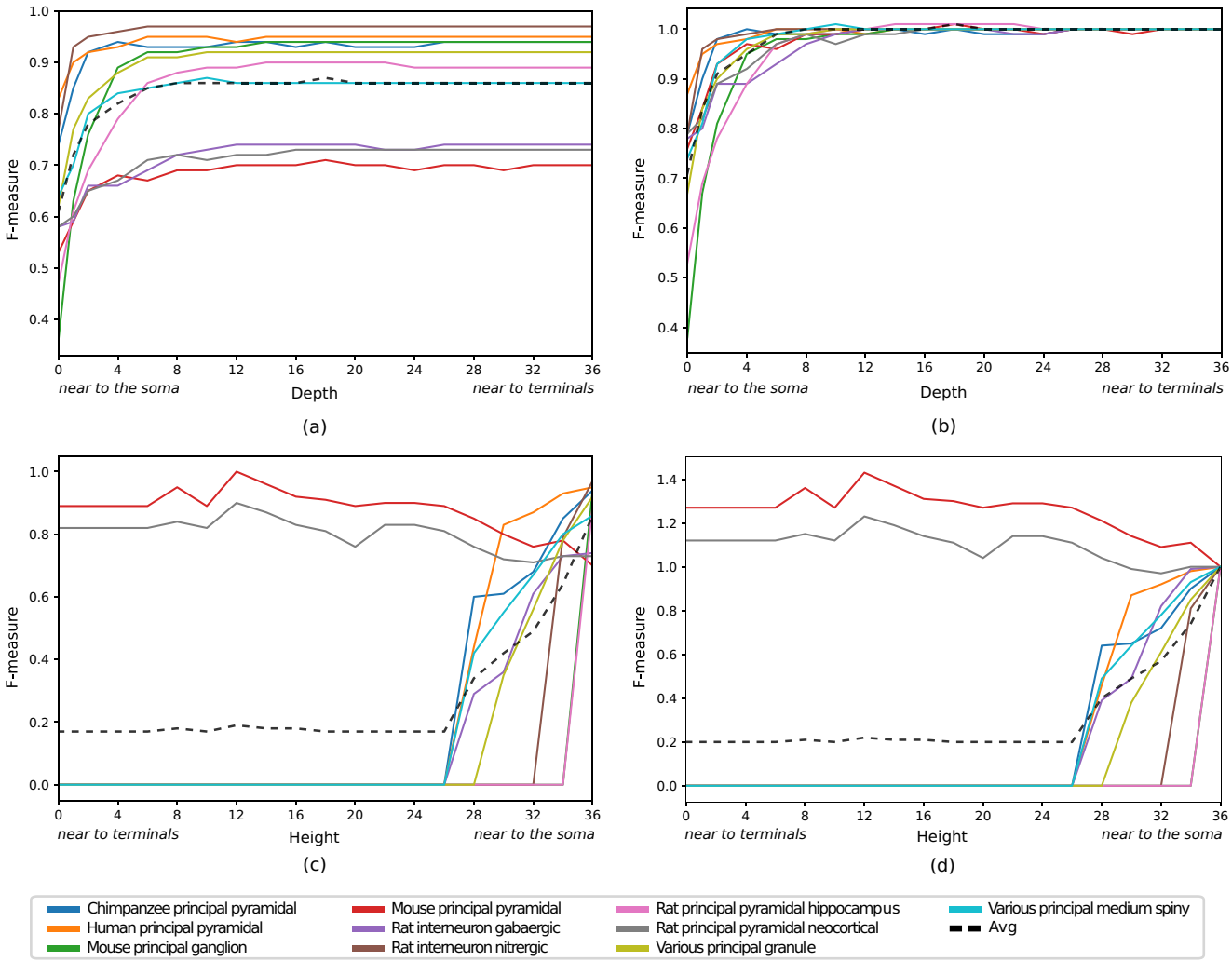


Fig. 8 The F-measure obtained for each class as a function of dendritic tree depth. Lower x values are related to branches that are near the soma. It is possible to observe the growth of the accuracy curves with the increase of the first depth levels until reaching the depth 3 in which the accuracy stabilizes and remains until the end of the graph. **b** shows the results presented in (**a**) normalized by the F-measure value obtained using whole neurons. **c** The F-measure obtained for each class in terms of dendritic tree height. Larger x values are related to branches near the soma. The second graph shows how the absence of the closest levels to the soma impact negatively the accuracy which only improves as these levels are being included.**d** shows the results presented in (**c**) normalized by the F-measure value obtained using whole neurons

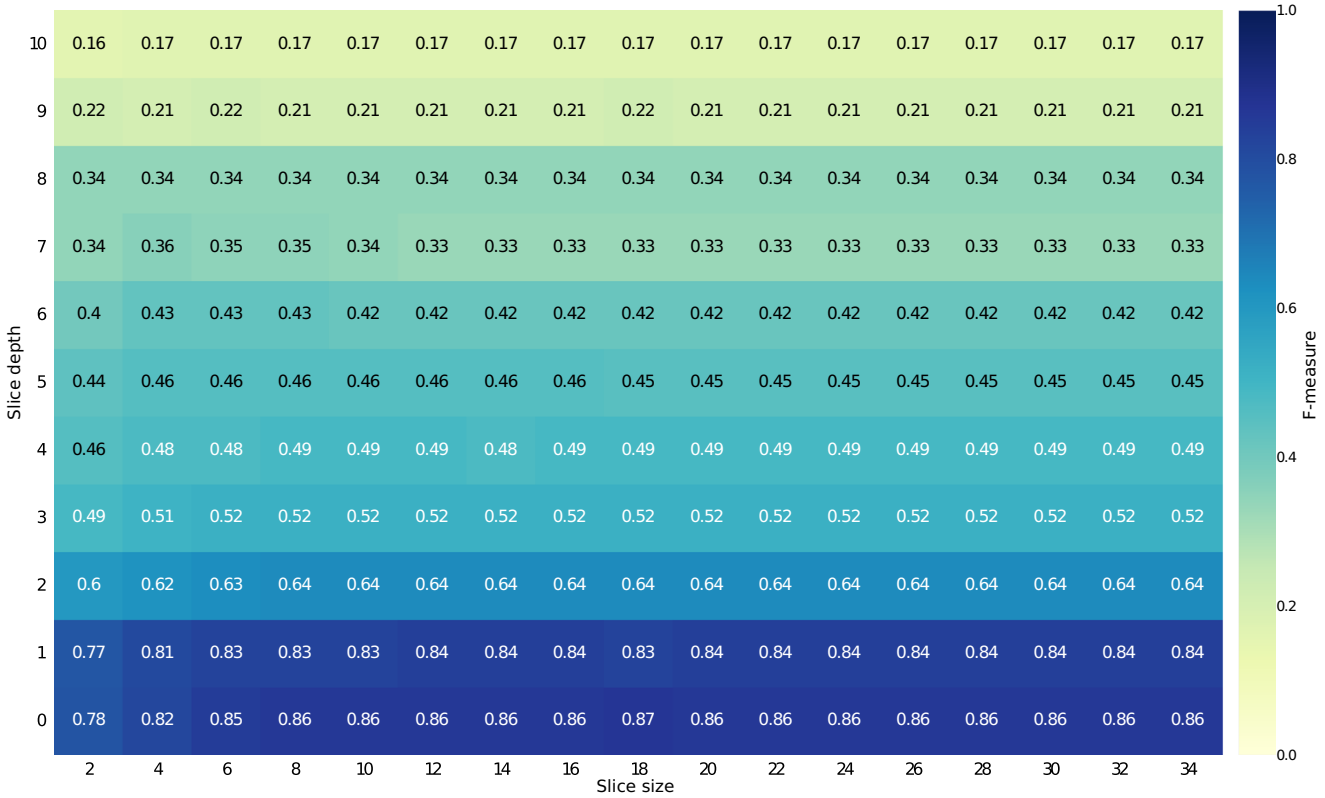


Fig. 9 Average F-measures for the 10 classes when decomposing the neuron by slices and performed the morphologic classification. The best results always contemplate the branches with low hierarchies

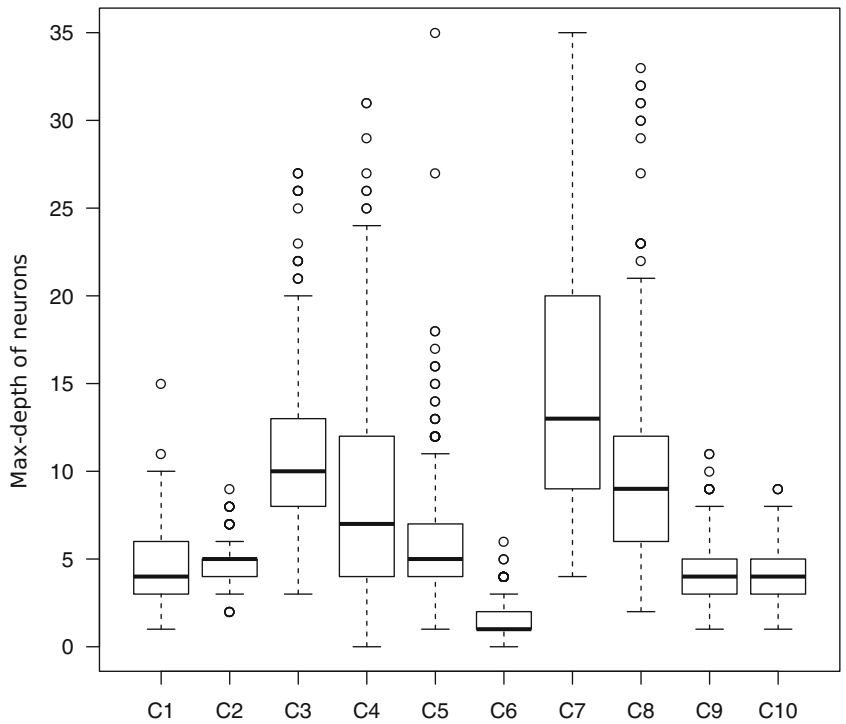


Fig. 10 Maximum branch order for each neuron class. The 10 classes are heterogeneous in depth

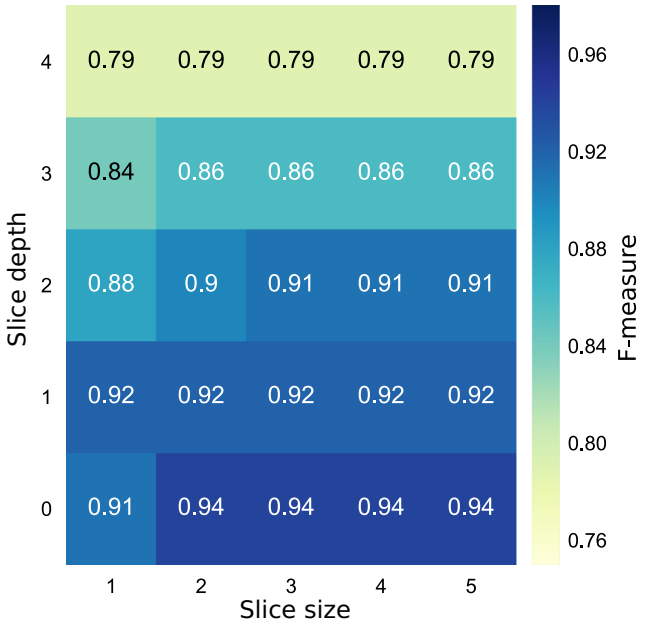


Fig. 11 F-measure values obtained from the classification based on slices for classes chimpanzee pyramidal (C1) , human pyramidal (C2), granule (C9) and medium spiny (C10) from different species

Experiment 3: Decomposition by Height

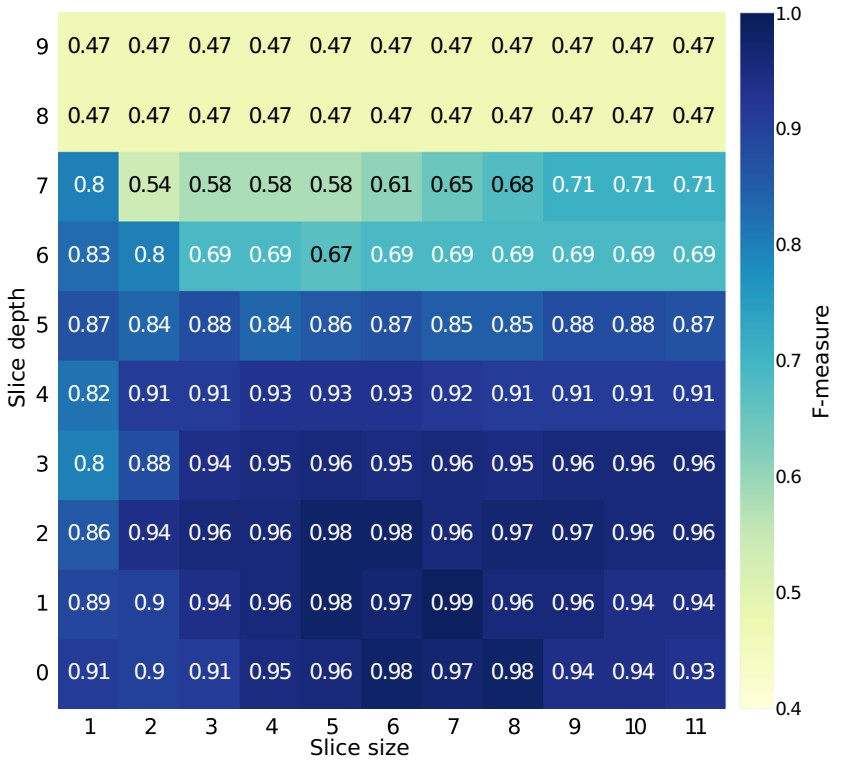
The third set of experiments involve decomposition by height. This test was created to investigate the opposite situation described in the previous experiment. Here, we are interested in analyzing how the dendritic tree height influences the classification performance. In other words, we want to determine how well the dendritic terminals can describe the neurons. The experimental results presented in Fig. 8b and d show that, for most classes, the dendrites with low hierarchy, far from the soma, provide poorer classification. For 8 of the classes, the classification starts to improve only after considering at least a height of 24. This happens because these classes have very few neurons branching at this height, and therefore cannot be properly characterized by such set of branches. Neurons from classes mouse pyramidal and rat pyramidal have long branches and, therefore, can still be differentiated from each other even for small heights.

The results indicate that, for most classes, branches near the leaves are not relevant for classification. This result, together with what was observed in the previous experiment, indicates that branches near the soma are more relevant for classification tasks.

Experiment 4: Decomposition by Slices

The fourth experiment was designed to assess decomposition by slice. This allows investigating if the neurons can be characterized by considering only a section of their dendritic tree. For this experiment, we considered all possible slices with sizes varying between 1 and the max-depth or

Fig. 12 Average F-measures for neurons belonging to mouse and rat species. The accuracy is high in most cases, especially in those involving levels of the lower hierarchy



max-height of the largest neuron in the database. Figure 9 shows the F-measures obtained with different slice sizes and positions. For the position of the slice, we considered the depth of the branches that were closest to the soma. The results indicate that, in most cases, the classification accuracy is not affected by the size of the slice. The only exception being for small slices at low depth values. Interestingly, a large classification accuracy can be obtained for slices having size of roughly 6 levels and being positioned at depth 0. This result confirms that branches that are close to the soma can provide a classification performance that is close to the performance achieved when using the entire neurons.

When considering large slice depths, some neurons are eliminated from the analysis because they do not posses any branch at that depth. In addition, some neuron classes posses many neurons having small depth, while other classes contain neurons having a large number of bifurcations, which usually leads to neurons having large depth. This difference between the classes can be better seen in the histogram of Fig. 10, which shows the maximum branch depth for neurons from each considered class. To verify if the result shown in Fig. 9 is not being influenced by variations on the number of neurons being used for classification, a more restricted set of neurons was considered. This restricted set contains only neurons that have a max-depth of at least 5.

The neurons were randomly selected from classes chimpanzee pyramidal (C1), human pyramidal (C2), granule (C9) and medium spiny from different species (C10) (100 neurons from each class), and the classification process using the decomposition by slice was carried out. The obtained results are shown in Fig. 11. The results are similar to those obtained in Fig. 9. The best classification results are achieved for slice sizes between 2 and 5 and low depth. In order to consider neurons having even more depth levels, a final test involving the classes mouse ganglion, rat interneuron nitrinergic and rat pyramidal with 37 neurons for each class and a fixed depth of 10 was performed. Similar results were obtained. In all cases, the slices closer to the soma led to better characterization of the neuron structure.

Table 4 Comparative table of features used in Experiment 5 with slices between 0-8 and 4-12

	Slice 0-8		Slice 4-12		Slice 0-8		Slice 4-12	
Measurement	Mouse Avg(SD)	Rat Avg(SD)	Mouse Avg(SD)	Rat Avg(SD)	T-test	P-val	T-test	P-val
Soma_Surface	14475.12 (10812.07)	26488.82 (14266.63)	12363.18 (10653.21)	26488.82 (14266.63)	−7.10	0.00	−7.59	0.00
N_stems	3.78 (2.76)	1.03 (0.16)	3.13 (1.8)	4.7 (1.87)	10.53	0.00	−5.60	0.00
N_bifs	10.01 (4.16)	15.91 (4.1)	3.41 (2.78)	10.7 (5.09)	−10.69	0.00	−12.45	0.00
N_branch	23.79 (9.34)	32.85 (8.17)	9.94 (5.99)	26.09 (10.46)	−7.72	0.00	−13.20	0.00
Width	143.8 (50.34)	182.47 (65.59)	98.32 (57.32)	178.74 (69.09)	−4.95	0.00	−8.47	0.00
Height	252.21 (85.74)	304.69 (77.29)	211.08 (116.26)	310.86 (84.52)	−4.81	0.00	−6.19	0.00
Depth	40.57 (17.49)	107.2 (30.91)	31.37 (18.95)	100.2 (32.44)	−19.85	0.00	−18.01	0.00
Diameter	0.77 (0.15)	1.23 (0.22)	1.42 (1.03)	1.15 (0.22)	−18.02	0.00	2.14	0.04
Length	1189.01 (471.55)	1822.8 (495.46)	769.99 (477.22)	2094.47 (682.38)	−9.81	0.00	−15.34	0.00
Surface	3001.15 (1197.17)	7875.18 (1973.1)	2070.93 (1086.77)	8373.26 (2347.49)	−22.35	0.00	−24.47	0.00
Volume	3992.96 (2761.85)	9125.51 (3816.68)	3294.7 (2669.53)	8871.53 (3741.29)	−11.53	0.00	−11.67	0.00
EucDistance	233.71 (96.12)	339.74 (75.27)	235.31 (115.64)	358.72 (81.41)	−9.19	0.00	−7.76	0.00
PathDistance	277.1 (110.63)	396.4 (81.46)	266.38 (132.07)	408.57 (87.32)	−9.19	0.00	−7.94	0.00
Branch_Order	4.42 (1.87)	7.39 (0.98)	1.78 (1.97)	4.32 (1.95)	−14.92	0.00	−8.45	0.00
Branch_pathlength	51.61 (13.09)	56.24 (11.47)	82.67 (40.12)	87.56 (33.7)	−2.82	0.01	−0.85	0.40
Contraction	0.9 (0.03)	0.91 (0.02)	0.93 (0.04)	0.91 (0.02)	−3.05	0.00	3.27	0.00
Fragmentation	267.38 (105.74)	613.17 (192.76)	96.71 (71.12)	510.44 (222.83)	−16.65	0.00	−18.20	0.00
Partition_asymmetry	0.51 (0.15)	0.53 (0.09)	0.68 (0.25)	0.5 (0.14)	−0.64	0.52	5.61	0.00
Rall_Power	1.36 (1.06)	2.54 (0.54)	0.03 (0.21)	2.54 (1.01)	−10.52	0.00	−25.52	0.00
Bif_ampl_local	75.91 (11.24)	71.28 (8.81)	38.67 (31.36)	67.36 (16.47)	3.43	0.00	−7.03	0.00
Bif_ampl_remote	62.8 (13.06)	59.49 (9.05)	37.01 (31.72)	52.43 (13)	2.20	0.03	−3.84	0.00
Fractal_Dim	1.03 (0.01)	1.03 (0.01)	1.01 (0.12)	1.03 (0.01)	2.47	0.01	−1.05	0.30

The table shows the average and standard deviation (between parenthesis) of the measurements for each class and the two slice ranges considered. Also shown is the t-value and p-value of a Student's t-test of the difference between the class averages

Experiment 5: Decomposition by Slice in Rat and Mouse Species

Previous experiments were done on neurons from different species and regions of the nervous system. In this experiment, the number of species was reduced to just two, mouse and rat, and only neurons acquired at specific experimental conditions for both species were considered. The selected neurons belonged to adult male individuals, and are all from the Lee_LJ and Henckens archives. In both cases the physical integrity of the neurons is classified as at least moderate by NeuroMorpho.Org, and all reconstructions posses information about branch diameter, depth and angle. The neurons belong to the prelimbic, layer 2-3 and middle regions of the prefrontal area of the neocortex. The primary cell class is principal cell and the secondary cell class is pyramidal. The experiment protocol was in vitro and the Golgi staining method was used. 112 neurons for each species were analyzed. Figure 12 shows the results regarding neuron classification based on the decomposition by slice. Slice depths going from 0 to 9 and sizes from 1 to 11 were used. The figure shows results similar to those obtained in Section "Experiment 4: Decomposition by Slices". Specifically, low-order branches seem to be better suited for differentiating the neurons.

In order to better analyze the measures that can distinguish the neuron classes, we calculate the average and standard deviation of each measurement for slices between 0-8 and also for slices between 4-12. The values are shown in Table 4. For each slice range, we also show the result of a t-test applied to the difference between class averages. The table shows that the most differentiating measures are those directly linked to the diameter of the dendrites.

Experiment 6: Decomposition by Slice in Neocortical Region

To conclude the experimental part, an even more limited set of neurons was chosen, involving only the mouse species. To homogenize even further the data set of this test, neocortical neurons reconstructed by the same laboratory and corresponding to adult individuals were chosen. This set of neurons belong to the Allen Cell Types archive. According to Neuromorpho.Org, the neurons have as structural domain dendrites, soma and incomplete axons. The provided morphological attributes are branch diameter, depth and angle. Two cell classes were analyzed, pyramidal and interneuron. 81 neurons from each class were included in the experiments.

Figure 13 shows a heatmap indicating the classification performance after decomposing the neurons by slices. In this figure, the importance of both the depth and size of the slices can be observed. Based on the values shown in the bottom right-hand side, it is possible to conclude that slices of sizes greater than six tend to present the best results. In addition, the branches of the first three hierarchies seem

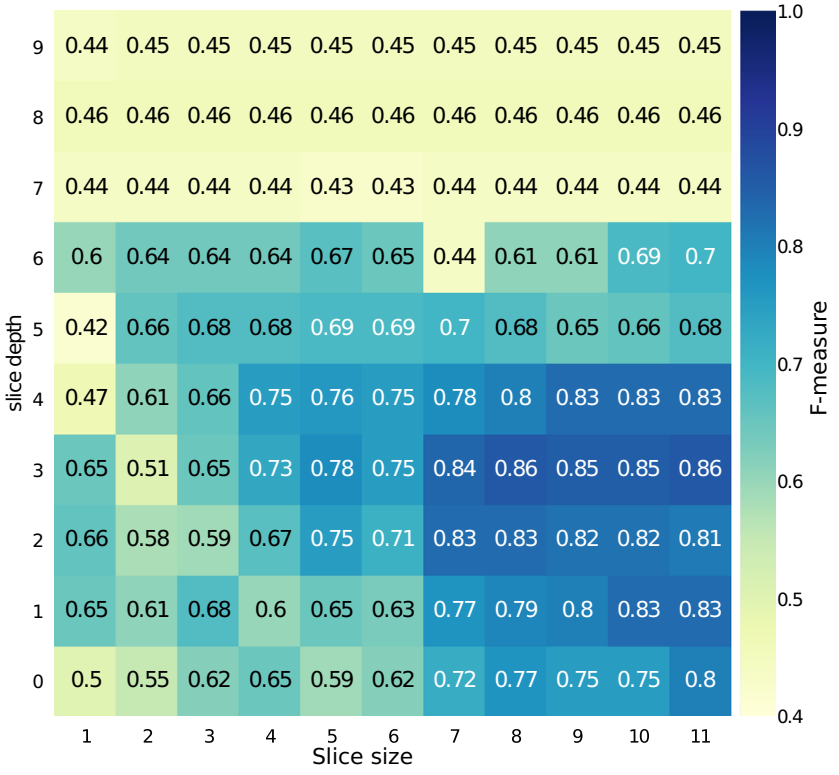


Fig. 13 Average F-measure values for a selected set of neurons belonging to mouse species. The largest F-measures are observed in the bottom right-hand side, which means that slices with sizes greater than six and including less hierarchies are more relevant for classifying pyramidal and interneurons from the neocortex region

Table 5 Comparative table of features used in Experiment 6 with slices between 0-6 and 3-12

	Slice 0-6		Slice 3-12		Slice 0-6		Slice 3-12	
Measurement	Interneuron Avg(SD)	Pyramidal Avg(SD)	Interneuron Avg(SD)	Pyramidal Avg(SD)	T-test	P-val	T-test	P-val
Soma_Surface	529.4 (265.86)	463.39 (155.71)	529.4 (265.86)	463.39 (155.71)	1.93	0.06	1.93	0.06
N_stems	6.04 (1.6)	6.5 (1.96)	9.28 (5.39)	9.7 (4.83)	−1.64	0.10	−0.52	0.61
N_bifs	22.28 (11.5)	25.43 (7.4)	7.52 (9.5)	10.29 (5.2)	−2.06	0.04	−2.30	0.02
N_branch	50.6 (23.11)	57.35 (15.28)	24.32 (22.27)	30.27 (12.82)	−2.19	0.03	−2.08	0.04
Width	215.15 (65.93)	226.06 (80.97)	172.74 (81.07)	200.49 (81.55)	−0.94	0.35	−2.17	0.03
Height	331.59 (103.73)	396.71 (137.05)	278.72 (135.12)	418.24 (140.92)	−3.40	0.00	−6.41	0.00
Depth	59.35 (19.64)	63.54 (22.27)	51.36 (20.66)	60.06 (22.14)	−1.27	0.21	−2.57	0.01
Diameter	0.44 (0.08)	0.5 (0.1)	0.49 (0.23)	0.51 (0.1)	−4.36	0.00	−0.74	0.46
Length	2610.18 (1037.56)	2859.68 (508.46)	1806.68 (1192.93)	2146.62 (626.02)	−1.94	0.05	−2.27	0.03
Surface	4062.34 (1844.32)	4849.14 (1103.72)	2896.51 (1856.69)	3709.44 (1117.7)	−3.29	0.00	−3.37	0.00
Volume	2319.53 (1799.94)	2152.27 (734.97)	2147.76 (1771.26)	1942.49 (716.26)	0.77	0.44	0.97	0.34
EucDistance	266.99 (94.36)	377.89 (116.39)	256.09 (108.66)	414.22 (121.86)	−6.64	0.00	−8.69	0.00
PathDistance	320.38 (105.62)	429.14 (121.5)	296.4 (122.44)	462.38 (128.65)	−6.06	0.00	−8.38	0.00
Branch_Order	5.17 (1.63)	7.26 (1.73)	2.33 (2.08)	4.9 (2.57)	−7.87	0.00	−6.96	0.00
Branch_pathlength	54.09 (14.7)	51.98 (12.54)	84.87 (31.62)	76.76 (26.43)	0.98	0.33	1.77	0.08
Contraction	0.88 (0.05)	0.89 (0.03)	0.89 (0.04)	0.89 (0.03)	−2.49	0.01	0.40	0.69
Fragmentation	2227.73 (951.17)	2412.09 (420.93)	1091.49 (846.97)	1455.35 (419.69)	−1.59	0.11	−3.46	0.00
Partition_asymmetry	0.47 (0.1)	0.48 (0.1)	0.51 (0.22)	0.5 (0.17)	−0.47	0.64	0.19	0.85
Rall_Power	2.31 (0.47)	2.32 (0.47)	1.67 (1.16)	2.23 (0.84)	−0.04	0.97	−3.47	0.00
Bif_ampl_local	74.54 (8.86)	73.86 (6.95)	65.26 (31.43)	70.85 (12.31)	0.55	0.58	−1.49	0.14
Bif_ampl_remote	63.5 (9.41)	69.14 (9.52)	54.59 (26.87)	63.54 (14.94)	−3.78	0.00	−2.624	0.01
Fractal_Dim	1.04 (0.02)	1.04 (0.01)	1.04 (0.02)	1.03 (0.01)	2.79	0.01	2.92	0.00

The table shows the average and standard deviation (between parenthesis) of the measurements for each class and the two slice ranges considered. Also shown is the t-value and p-value of a Student's t-test of the difference between the class averages

to be implying ambiguity, leading to lower classification accuracy. This may be due to the fact that these two neuron classes are morphologically similar in the first three levels of depth, while they tend to be different along subsequent levels. On the other hand, as in the previous results, only very deep branches do not contribute to improving the classification.

Table 5 shows the average and standard deviation of the considered measures for slice ranges going from 0 to 6 and from 3 to 12. The respective t-value and p-value of the Student's t-test for the difference between the averages are also shown in the table. The table suggests that the most discriminative measures are those directly linked to the size of the dendrites, such as height, path distance and euclidean distance.

In order to better understand the results of experiment 6, a final test was performed involving only basal dendrites of both types of neurons. The result is presented in Fig. 14. The figure shows that the classifier does not achieve a high accuracy, which is due the basal dendrites of both neuron

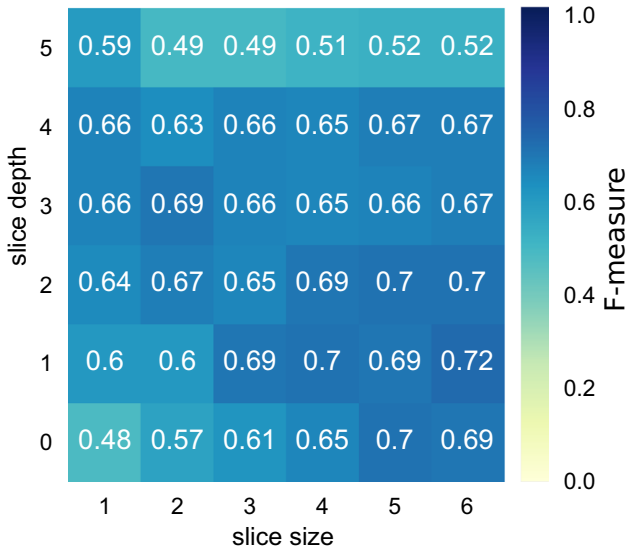


Fig. 14 Average F-measures obtained after classifying pyramidal neurons and interneurons from the mouse species, considering only the basal dendrites

classes have similar characteristics. In the set of selected neurons, not all neurons have apical dendrites and basal dendrites tend to belong to lower hierarchies than apical dendrites. Therefore, due to the the similarity between basal dendrites, the low-hierarchy branches are not as relevant for classification as those belonging to the hierarchies between depth 3 and 12 in Fig. 13.

Conclusion

New advances on neuron morphology analysis have benefited from standardized data made available in different online resources, which favors reproducibility as well as the development of new methods to be applied to much larger datasets. Still, neuron classification remains a challenging problem. The present paper focused on whether parts of neurons can be reliably used for neuron classification. This approach is particularly promising as a means to extend and refine investigations about the relationship between neuronal shape and connectivity/function. A considerable number of neurons (more than 5000) belonging to ten classes and different species was considered. The reported experiments confirm that it is possible to identify relevant parts of dendritic trees allowing classification results equivalent to those obtained by using the whole neuron. Additionally, branches near the soma were found to provide the most discriminative information about the neuronal structure.

The main branches and the branches near to the soma are the most relevant in the classification experiments here reported, as the adopted learning mechanism basically looks for common intraclass patterns and distinct interclass patterns. This result indicates that low-order branches, which are the first to develop in a neuron, tend to follow a common pattern within the same class. From higher orders, the neurons develop branching patterns that are less specific to each neuronal category, perhaps as a consequence of the influence of factors such as the environment. Therefore, they ended up not contributing to improve the classification.

On the other hand, in the case of neurons that are quite similar in the lower hierarchical levels, the learning mechanism looks for a slice with size and depth that allow to clearly differentiate the classes of neurons. In these cases, although the main branches are important because of their morphological information, they are not relevant for the classification since they lead to small interclass difference. That is why in Experiment six the result suggests that the most relevant branches belonging to hierarchical levels greater than three.

The development of dendrites obeys intrinsic/extrinsic, activity-dependent, and dedicated mechanisms that control maintenance of dendritic arbors (Dong et al. 2015; Lefebvre et al. 2015). Respective intraclass patterns were clearly identified in the present work. These patterns show that, from the soma up to a certain depth, the neurons have a kind of class-specific growth. For larger depths, the neurons studied begin to diversify, possibly as a result of extrinsic factors, which complicates the identification of patterns for branches that are distant from the soma.

Future works could consider other types of neuronal cells, as well as cells originating from specific neuronal modules (e.g. hippocampus). It would also be interesting to consider the concept of neuromorphic space (Costa et al. 2010) in the analysis.

Information Sharing Statement

The source code for BTMORPH (RRID:SCR_003566) is available at github: https://bitbucket.org/epcervantes7/btmorph. In the article we worked with data from NeuroMorpho.Org (RRID:SCR_002145): http://neuromorpho.org.

Acknowledgements C. H. Comin thanks FAPESP (Grant No. 15/18942-8) for financial support. L. da F. Costa thanks CNPq (Grant No. 307333/2013-2) for support. This work has also been supported by the FAPESP grant 2015/22308-2, Capes and CNPq.

Compliance with Ethical Standards

Conflict of interests The authors declare no conflict of interest.

References

Armañanzas, R., & Ascoli, G.A. (2015). Towards the automatic classification of neurons. *Trends in Neurosciences*, *38*(5), 307–318.

Ascoli, G.A. (2002). *Computational neuroanatomy: principles and methods*. Springer Science & Business Media.

Ascoli, G.A., Donohue, D.E., Halavi, M. (2007). Neuromorpho.org: a central resource for neuronal morphologies. *The Journal of Neuroscience*, *27*(35), 9247–9251.

Barbosa, M.S., da Fontoura Costa, L., de Sousa Bernardes, E. (2003). Neuromorphometric characterization with shape functionals. *Physical Review E*, *67*(6), 061910.

Bazán, N.G., & Lolley, R.N. (2013). *Neurochemistry of the retina: proceedings of the international symposium on the neurochemistry of the retina held in Athens, Greece, August 28-September 1 1979*. Elsevier.

Bernard, A., Sorensen, S.A., Lein, E.S. (2009). Shifting the paradigm: new approaches for characterizing and classifying neurons. *Current Opinion in Neurobiology*, *19*(5), 530–536.

Bota, M., & Swanson, L.W. (2007). The neuron classification problem. *Brain Research Reviews*, *56*(1), 79–88.

Cajal, S.R.y., & Azoulay, L. (1955). *Histologie du systèeme nerveux de l'homme & des vertébrés*. Instituto Ramon Y Cajal: Consejo superior de investigaciones cientificas.

Cannon, R., Turner, D., Pyapali, G., Wheal, H. (1998). An online archive of reconstructed hippocampal neurons. *Journal of Neuroscience Methods*, *84*(1), 49–54.

Chang, C.-C., & Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 27.

Comin, C.H., & da Fontoura Costa, L. (2013). Shape, connectedness and dynamics in neuronal networks. *Journal of Neuroscience Methods*, *220*(2), 100–115.

Costa, L.D.F., Zawadzki, K., Miazaki, M., Viana, M.P., Taraskin, S. (2010). Unveiling the neuromorphological space. *Frontiers in Computational Neuroscience*, *4*, 150.

da Costa, F.L., & Velte, T.J. (1999). Automatic characterization and classification of ganglion cells from the salamander retina. *Journal of Comparative Neurology*, *404*(1), 33–51.

DeFelipe, J. (2001). Cortical interneurons: from cajal to 2001. *Progress in Brain Research*, *136*, 215–238.

Ding, M., & Glanzman, D. (2011). *The dynamic brain: an exploration of neuronal variability and its functional significance*. USA: Oxford University Press.

Dong, X., Shen, K., Bülow, H.E. (2015). Intrinsic and extrinsic mechanisms of dendritic morphogenesis. *Annual Review of Physiology*, *77*, 271–300.

Gillette, T., & Ascoli, G. (2015). Topological characterization of neuronal arbor morphology via sequence representation. i. Motif analysis.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1), 389–422.

Halavi, M., Hamilton, K.A., Parekh, R., Ascoli, G.A. (2012). Digital reconstructions of neuronal morphology: three decades of research trends. *Frontiers in Neuroscience*, *6*(49), 1–11.

Hosp, J.A., Strüber, M., Yanagawa, Y., Obata, K., Vida, I., Jonas, P., Bartos, M. (2014). Morphophysiological criteria divide dentate gyrus interneurons into classes. *Hippocampus*, *24*(2), 189–203.

Kandel, E.R., Schwartz, J.H., Jessell, T.M., et al. (2000). *Principles of neural science* Vol. 4. New York: McGrawhill.

Lefebvre, J.L., Sanes, J.R., Kay, J.N. (2015). Development of dendritic form and function. *Annual Review of Cell and Developmental Biology*, *31*, 741–777.

López-Cruz, P. L., Larrañaga, P., DeFelipe, J., Bielza, C. (2014). Bayesian network modeling of the consensus between experts: an application to neuron classification. *International Journal of Approximate Reasoning*, *55*(1), 3–22.

Lu, Y., Carin, L., Coifman, R., Shain, W., Roysam, B. (2015). Quantitative arbor analytics: unsupervised harmonic co-clustering of populations of brain cell arbors based on l-measure. *Neuroinformatics*, *13*(1), 47–63.

Magee, J.C. (2000). Dendritic integration of excitatory synaptic input. *Nature Reviews Neuroscience*, *1*(3), 181.

Mainen, Z.F., & Sejnowski, T.J. (1996). In uence of dendritic structure on firing pattern in model neocortical neurons. *Nature*, *382*(6589), 363.

McGarry, L.M., Packer, A.M., Fino, E., Nikolenko, V., Sippy, T., Yuste, R. (2010). Quantitative classification of somatostatin-positive neocortical interneurons identifies three interneuron subtypes. *Frontiers in Neural Circuits*, *4*, 12.

Mihaljević, B., Benavides-Piccione, R., Bielza, C., DeFelipe, J., Larrañaga, P. (2015). Bayesian network classifiers for categorizing cortical gabaergic interneurons. *Neuroinformatics*, *13*(2), 193–208.

Mottini, A., Descombes, X., Besse, F. (2014). Axonal tree classification using an elastic shape analysis based distance. In *2014 IEEE 11th International symposium on biomedical imaging (ISBI)* (pp. 850–853). IEEE.

Parekh, R., & Ascoli, G.A. (2013). Neuronal morphology goes digital: a research hub for cellular and system neuroscience. *Neuron*, *77*(6), 1017–1038.

Rall, W. (1967). Distinguishing theoretical synaptic potentials computed for different soma-dendritic distributions of synaptic input. *Journal of Neurophysiology*, *30*(5), 1138–1168.

Ruz, I.D., & Schultz, S.R. (2014). Localising and classifying neurons from high density mea recordings. *Journal of Neuroscience Methods*, *233*, 115–128.

Santana, R., McGarry, L., Bielza, C., Larrañaga, P., Yuste, R. (2013). Classification of neocortical interneurons using affinity propagation. *Frontier in Neural Circuits*, *7*, 185.

Scorcioni, R., Polavaram, S., Ascoli, G.A. (2008). L-measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nature Protocols*, *3*(5), 866–876.

Sharpee, T.O. (2014). Toward functional classification of neuronal types. *Neuron*, *83*(6), 1329–1334.

Sümbül, U., Zlateski, A., Vishwanathan, A., Masland, R.H., Seung, H.S. (2014a). Automated computation of arbor densities: a step toward identifying neuronal cell types. *Frontiers in Neuroanatomy*, *8*, 139.

Sümbül, U., Song, S., McCulloch, K., Becker, M., Lin, B., Sanes, J.R., Masland, R.H., Seung, H.S. (2014b). A genetic and computational approach to structurally classify neuronal types. *Nature Communications*, 5.

Teeter, C.M., & Stevens, C.F. (2011). A general principle of neural arbor branch density. *Current Biology*, *21*(24), 2105–2108.

Torben-Nielsen, B. (2014). An efficient and extendable python library to analyze neuronal morphologies. *Neuroinformatics*, *12*(4), 619.

Uji, Y., Yamamura, H., et al. (1995). Morphological classification of retinal ganglion cells in mice. *Journal of Comparative Neurology*, *356*(3), 368–386.

Uylings, H.B., & van Pelt, J. (2002). Measures for quantifying dendritic arborizations. *Network: Computation in Neural Systems*, *13*(3), 397–414.

Uylings, H.B., Van Pelt, J., Verwer, R.W., McConnell, P. (1989). Statistical analysis of neuronal populations. In *Computer techniques in neuroanatomy* (pp. 241–264). Springer.

Zhao, T., & Plaza, S.M. (2014). Automatic neuron type identification by neurite localization in the drosophila medulla. arXiv:1409.1892.