Cluster Analysis of Cortical Pyramidal Neurons Using SOM

Andreas Schierwagen[1,⋆], Thomas Villmann[2], Alan Alpár[3], and Ulrich Gärtner[3]

[1] Institute for Computer Science, University of Leipzig, 04109 Leipzig, Germany
schierwa@informatik.uni-leipzig.de
[2] Department of Mathematics/Physics/Computer Sciences,
University of Applied Sciences Mittweida, 09648 Mittweida, Germany
[3] Department of Neuroanatomy, Paul Flechsig Institut for Brain Research,
University of Leipzig, 04109 Leipzig, Germany

Abstract. A cluster analysis using SOM has been performed on morphological data derived from pyramidal neurons of the somatosensory cortex of normal and transgenic mice.

Keywords: Cluster analysis, Kohonen's SOM, transgenic mouse, somatosensory cortex, pyramidal neurons, dendritic morphology.

1 Introduction

In the neurosciences, brain organization is studied at different levels both in ontogenetic and phylogenetic development. At the cellular level, neurons in different brain regions of one species and in the same brain region of different species are compared with respect to their structural and functional properties. We find a great variety in neuronal shapes and cell types, as well as a large variability within neuron classes (Fig. 1, left). Neuronal structure is characterized by elongated processes (neurites); among them, two kinds can be differentiated, the often-branching dendrites and the axon. From a functional point of view, axons and dendrites are conduits for electrical and chemical signals. The shapes of neurites determine both the routes for signal transmission within the nervous system and the way in which electrical signals are processed and transmitted [1].

The formation of a neuron's dendritic and axonal branching patterns is partly determined by genetic factors and partly by interactions with the surrounding tissue. The invention of transgenic mice mutations has provided important means for understanding gene function. In these mutants, gene overexpression may affect several organs and tissues, including the cells and networks of the brain.

In the following, we describe the method of morphological quantification (neuromorphometry) and its use for classification of pyramidal neurons. More specifically, neurons in the same brain region (somatosensory cortex) of two 'species' of mice (wildtype and transgenic type) have been compared with respect to their shape properties. Shape classification was performed with an unsupervised learning method, i.e. Kohonen's self-organizing maps.

⋆ Corresponding author.

F. Schwenker and N. El Gayar (Eds.): ANNPR 2010, LNAI 5998, pp. 120–130, 2010.

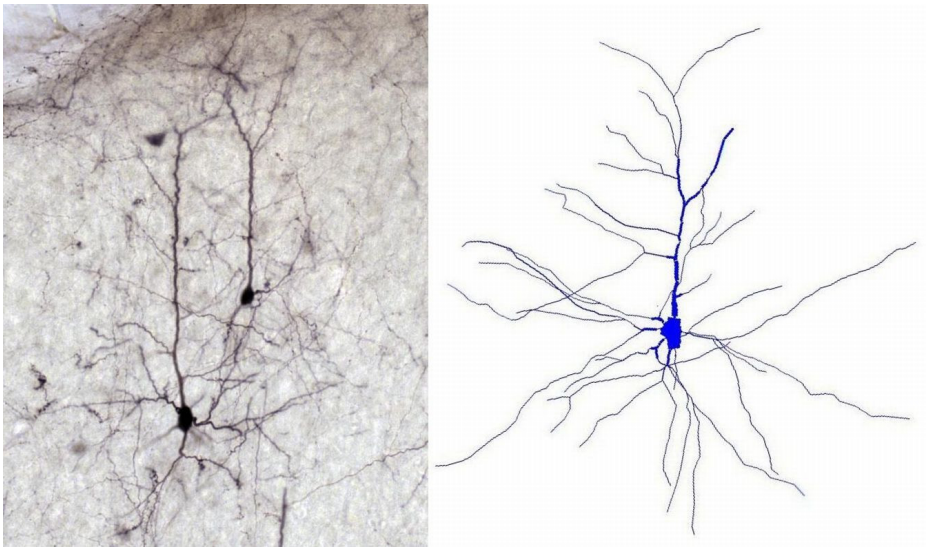

Fig. 1. Pyramidal neurons from layers II/III of *synRas* mouse cortex. Left: light micrograph of retrogradely labelled neurons, right: pyramidal neuron rendered with CVAPP.

2 Materials and Methods

2.1 Experimental Basis and Reconstructions

We used data derived from experiments with three *synRas* mice aged nine months as well as with three wildtype mice of the same age. The details of experimental procedures have been described in [2]. The main deliverables were two sets of retrogradely labelled pyramidal cells (28 cells from wildtype and 28 cells from transgenic mice) which were reconstructed (Fig. 1) using Neurolucida^{TM} (MicroBrightField, Inc.). The system allowed accurate tracing of the cell processes in all three dimensions and continuous adjustment of the dendritic diameter with a circular cursor. A motorized stage with position encoders enabled the navigation through the section in the xyz axes and the accurate acquisition of the spatial coordinates of the measured structure. All visible dendrites were traced without marking eventual truncation of smaller dendritic sections. This may have led to certain underestimation of the dendritic tree, especially in *synRas* mice with a larger dendritic tree. To gain both optimal transparency for optimal tracing facilites and at the same time a possibly complete neuronal reconstruction, sections of 160 μm thickness were used. Thicker sections allowed only ambiguous tracing of thinner dendritic branches. Shrinkage correction was carried out in the z axis, but not in the xy plane, because shrinkage was negligible in these dimensions [2].

2.2 Editing and Conversion of Morphology Files

The morphology files created in this way were processed with CVAPP [3], a cell viewing, editing and format converting program for morphology files (Fig. 1, right). In particular, CVAPP has been used to edit and convert the Neurolucida ASCII files to the SWC format describing the structure of a neuron in the simplest possible way. In this format, each line encodes the properties of a single neuronal compartment. The format of a line in a SWC file is as follows: $nTxyzRP$. In turn, these numbers mean: (1) an integer label (increasing by one from one line to the next) that identifies the compartment, (2) an integer that represents the type of neuronal segment (0-undefined, 1-soma, 2-axon, 3-dendrite, 4-apical dendrite, etc.); (3)-(5) xyz coordinates of compartment, (6) radius of compartment, (7) parent compartment (defined as -1 for the initial compartment).

2.3 Terminology and Shape Characteristics

Neurons are 3D objects, and the location of their cell bodies within the nerve tissue, as well as the number, spatial extent, branching complexity and 3D embedding of their axonal and dendritic arborizations, are prominent shape characteristics that may differ significantly between cell types. Morphological measurements of these characteristics (neuromorphometry) can be described as follows [4,5].

The dendrites of a neuron are representeded as trees of segments arising from the cell body (soma). A segment is defined as a portion of dendrite extending between two branching points (intermediate segments), or between a node (branching point) and a tip (terminal segments). Dendritic segments are approximated by cylindric sections of length l and diameter d. The distance from the soma to a point on the dendritic tree measured along the course of the segments lying inbetween is the path length which is generally greater than the Euclidean distance between the corresponding points. Measurements of dendritic trees can be differentiated into metrical and topological ones. Measurements include the length and diameter of the segments, path lengths, Euclidean distances of terminal tips from the cell soma and branching angles. A different class of measures is concerned with the spatial embedding in 3D space and focuses on, e.g., the spatial extension, spatial density, spatial orientation, and space filling (fractality) of the structure [6].

Dendritic trees may be categorized by topological type depending on the patterning of segments, independently of metrical and orientation features. The tree is reduced to a skeleton structure of points (branching or terminal points) and segments between these points. Such a skeleton forms a specific tree out of a finite set of possible different topological tree types. The tree-asymmetry index provides a discriminative measure based on asymmetries of pairs of subtrees at bifurcations. Other parameters used are *order* and *degree* (Fig. 2).

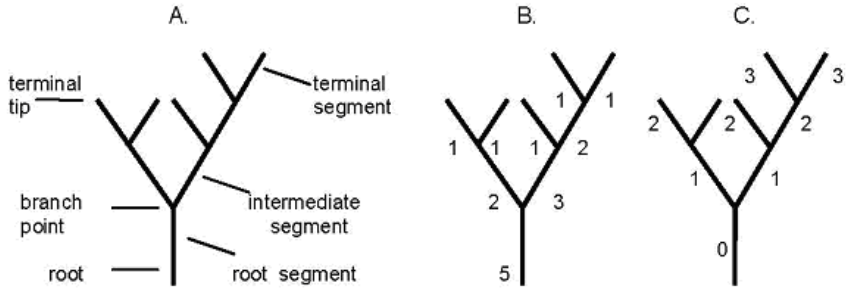

Fig. 2. Representation of dendritic topology. The tree is depicted by a set of connected segments (A) which are labeled by the degree of their subtrees (B) and their centrifugal order (C).

The *order* γ represents the topological distance from the soma. Its value is an integer incremented at every bifurcation (centrifugal order). A value of $\gamma = 0$ is assigned to the primary segments, i.e. those emerging directly from the soma. The *degree* n represents the number of tips of a subtree (or partition) stemming from a segment. In a binary tree, it is related to m, the number of segments of the partition, by $m = 2n - 1$. The dendritic systems of pyramidal cells are divided into basal and apical parts. Each of the subdivisions is considered separately (see Section 3).

All morphologic measurements were extracted from the anatomical files by using the L-Measure software [7]. This software allows the automated extraction of multiple sets of neuroanatomical parameters from reconstructed neurons. Our neuron data included no (or only partial) reconstructions of the axon; thus only the soma and the dendritic trees (basal and apical separately) were considered. L-Measure allows to derive more than 40 neuroanatomical parameters; a number of them, however, are correlated with each other. Thus, we selected 19 independent morphometric parameters (see Table 1) for each of which mean value, standard deviation, minimum and maximum were calculated.

2.4 Feature Space Representation

The shape of a given neuron in three-dimensional space can be characterized by the vector $\boldsymbol{x} = [x_1, x_2, ..., x_N]$ where x_i stands for any measurement. If $\boldsymbol{x}$ allows the original shape to be reconstructed to a specified degree of accuracy, $\boldsymbol{x}$ provides a (more or less) complete representation of the original shape. In the present case, the dimension of the feature space is $19 \times 4 = 76$.

The vector $\boldsymbol{x}$ can be displayed as point in the respective N-dimensional feature space. Neurons belonging to two distinct neural classes can be characterized by clouds of points in the feature space. Provided the measures are sound, vectors defined by similar neurons will be located close to each other in the feature space, while those corresponding to different neurons will tend to be distant from each other. This means, the Euclidean distance in the feature space is related to the

Table 1. Definitions of morphological parameters used in this study

Bifurcations	Number of dendritic bifurcations (nodes)
Segments	Number of dendritic segments
Trees	Number of dendritic trees arising from the soma
Tips	Number of terminal segments per dendritic trees (tree degree)
Segment order	Number of bifurcations between soma and terminal tips
Partition asymmetry	Partition asymmetry of a subtree, defined as $A_p = \frac{\|r-l\|}{r+l-2}$, with r and l the degrees of the two sub-subtrees
Diameter	Diameter of a segment
Segment length	Length of a dendritic segment
Daughter ratio	Diameter ratio of the two daughter segments at a bifurcation
Parent–child ratio	Diameter ratio between parent and daughter diameter at a bifurcations
Last parent diameter	Diameter of parent segment of the last bifurcation before tips
Rall's power	Best fitting parameter n for Rall's formula: $d_p^n = d_a^n + d_b^n$, where d_p , d_a, d_b, are the parent diameter and the two daughter diameter
Segment taper	Taper rate per segment, calculated as the difference between final and initial diameter divided by the initial diameter at each segment
Unit taper rate	Taper rate per unit length, calculated as the difference between final and initial diameter divided by the segment length
Euclidean distance	Euclidean distance between soma and segment
Path distance	Path distance between soma and segment
Contraction	Ratio between the Euclidean distance and the distance along the path
Local bifurcation angle	Angle between the two daughters at each bifurcation
Remote bifurcation angle	Angle between the tips of the two daughters of each bifurcation

dissimilarity between the shape of the respective cells. It is therefore obvious to approach the problem of classifying nerve cells through a procedure of clustering the respective vectors in the feature space.

2.5 Cluster Analysis

The data classification was realized using the Kohonen Self-Organizing Map (SOM) algorithm [8] to map the complex database on the two-dimensional plane visualizing the *synRas* activation effect, and to designate the relevant variables contributing to the model.

The SOM algorithm is an unsupervised learning procedure which can be summarized as follows. The data are assumed to be n-dimensional real data vectors

$\mathbf{v} \in V$ which are mapped onto a set $\mathbf{W} = \{\mathbf{w_r}\}_{\mathbf{r}\in A}$ of prototypes. Thereby, A is an usually two-dimensional discrete lattice and $N = \#A$ is the number of prototypes. SOMs, here taken in the variant of T.Heskes [9], minimize the cost function

$$E = \int P(\mathbf{v}) \sum_{\mathbf{r}} \delta_{\mathbf{r}}^{\mathbf{s}(\mathbf{v})} \sum_{\mathbf{r}'} h_\sigma(\mathbf{r}, \mathbf{r}')(\ \mathbf{v} - \mathbf{w}_{\mathbf{r}'})^2 d\mathbf{v} \tag{1}$$

with $\delta_{\mathbf{r}}^{\mathbf{s}(\mathbf{v})}$ being the Kronecker-symbol and $P(V)$ is the data distribution. The neighborhood function

$$h_\sigma(\mathbf{r}, \mathbf{r}') = \exp\left(\frac{-\|\mathbf{r} - \mathbf{r}'\|}{2\sigma^2}\right) \tag{2}$$

describes the learning cooperativeness between the lattice nodes. Learning of the prototypes is realized as a stochastic gradient descent on the cost function E for decreasing neighborhood range σ with respect to the prototypes $\mathbf{w_r}$:

$$\triangle\, \mathbf{w_r} = -\varepsilon \frac{\partial E}{\partial \mathbf{w_r}} \tag{3}$$

$$= \exp\left(\frac{-\|\mathbf{s}(\mathbf{v}) - \mathbf{r}\|}{2\sigma^2}\right)(\mathbf{v} - \mathbf{w_r}) \tag{4}$$

whereby

$$\mathbf{s}(\mathbf{v}) = \underset{\mathbf{r}\in A}{\operatorname{argmin}} \left[\sum_{\mathbf{r}'} h_\sigma(\mathbf{r}, \mathbf{r}')(\mathbf{v} - \mathbf{w}_{\mathbf{r}'})^2\right]. \tag{5}$$

is the so–called winning node. The mapping rule (5) determines a winner–take–all learning.

SOMs can be taken as a mapping of high-dimensional data onto a low-dimensional lattice [10]. Under certain conditions this mapping is topology preserving, i.e. similar data points are mapped onto neighbored lattice neurons or the same lattice node [11]. Yet, this topographic mapping can not be achieved for any data-lattice-configuration. Growing variants of SOM (GSOM) try to adapt the edge length ratio as well as the dimensionality of the lattice to result a topographic map [12].

Nevertheless, topography has to be judged after SOM-learning by respective quality measures to assure topography. For this purpose, the robust *topographic product* (TP) is frequently applied [13]. It calculates the averaged distortion of distance ratios within the set of prototypes and relates these to the respective node distance ratios within the SOM-lattice. The TP yields approximately zero-values for topology preserving mapping whereas values deviating significantly from zero indicate violations in topographic mapping. For a detailed description we refer to [13]. If the SOM is topology preserving then further investigations like component plane analysis or other can be applied. Component planes picture the value distribution of the prototypes $\mathbf{w_r}$ for a single data dimension. Thereby, the prototypes are arranged according to the position of their assigned lattice nodes $\mathbf{r}$. In this way, the topological ordering of the prototypes can be visualized.

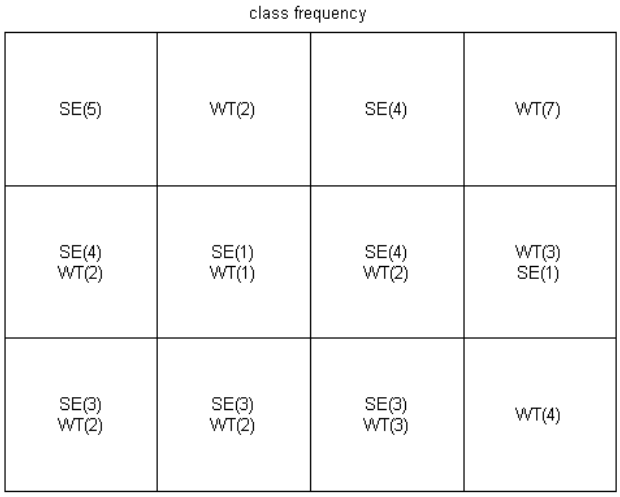

Fig. 3. SOM frequency map for neuron morphometry data. It shows the distribution of the samples, wildtype (WT) and transgenic neurons (SE), over each SOM node.

One method to detect clusters in the lattice space of SOMs is the U-Matrix method [14]. The U-matrix for a two-dimensional rectangular lattice A of size $n_1 \times n_2$ has the size $(2n_1 - 1) \times (2n_2 - 1)$ and is calculated according the following scheme: For each lattice node $\mathbf{r} = (r_1, r_2)$ the matrix element $u_{2r_1,2_{r2}}$ is the mean of the distances of the prototype $\mathbf{w_r}$ to those prototypes $\mathbf{w_{r'}}$ for which $\mathbf{r'}$ is a neighbored node in the rectangular lattice A. The elememt $u_{2r_1-1,2_{r2}}$ is the distance to the prototype $\mathbf{w_{r'}}$ with $\mathbf{r'} = (r_1 - 1, r_2)$. The other neighbored matrix elements of the entry $u_{2r_1,2_{r2}}$ are calculated accordingly. Thus, matrix entries give an indication of cluster boundaries between the respective nodes in case of a topographic mapping. Therefore, topology preservation is strongly demanded for correct interpretation [10].

The SOM analysis was performed in MATLAB with the publicly available SOM Toolbox [15].

3 Results

For the given data set of 56 data points, each containing the $n = 76$ morphological features, GSOM yielded a 4×3 lattice (Fig. 3). As we can see, each SOM node was hit by pyramidal neuron data. Some nodes were hit either only by transgenic (SE) or by wildtype (WT) neuron data, and others by both. The overall classification accuracy after labeling the map by majority vote is about 77% which refers to non-random partitioning but with class overlap.

After learning the TP was calculated to measure the map quality. It yielded a value of 0.0078 indicating a good topology preservation. This topological ordering can also be seen considering the component planes (CP) of the map (Fig. 4). To

make it more accessible, only the first 20 of the 76 parameters were presented as CPs. Shown are the values of only one feature in each map unit (in the original by color-coding). These planes show a clear structure and refer to correlated features if the respective planes look similar in the value distribution.

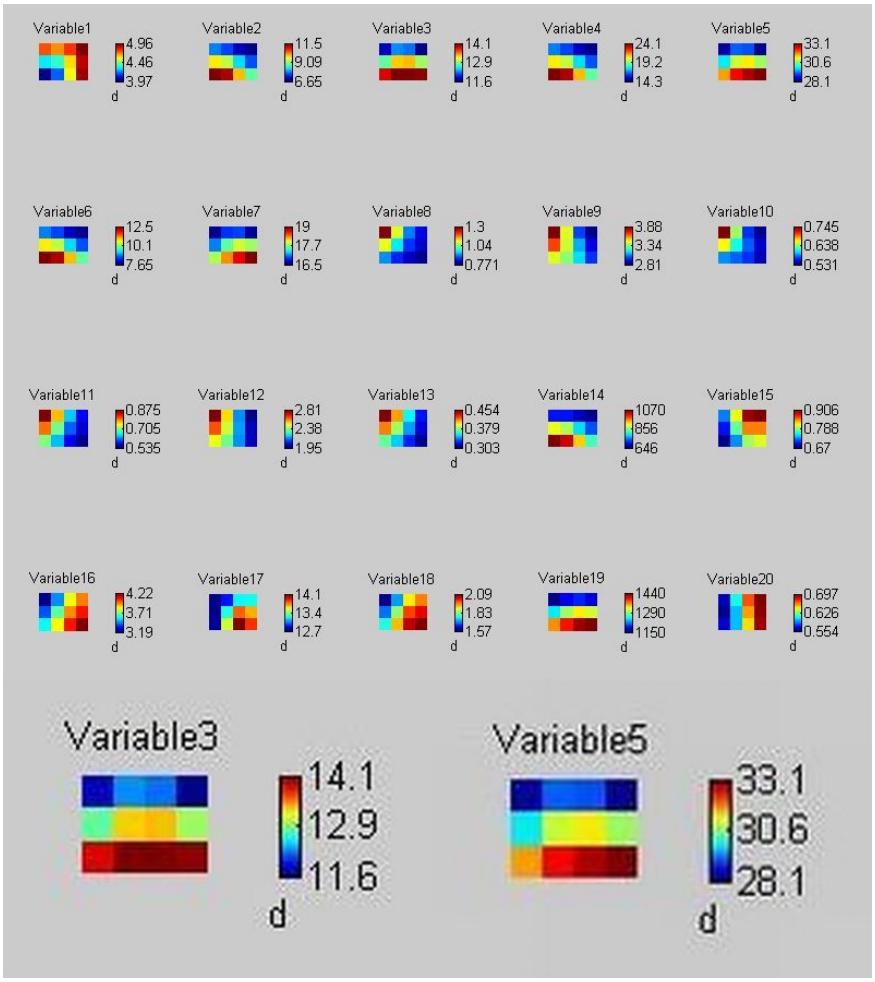

Fig. 4. Component planes of the SOM in Fig. 3

If we take for example the parameters 3 and 5, we see that the CPs of these two features are quite similar, meaning that the distances of these values in the same area of the map correlate. According to the parameter coding list, the two parameters represent the number of bifurcations and the number of branches of the basal dendritic tree. Here a correlation is actually present, because for binary trees a strict relationship holds (see Section 2.3). Comparing the class

distribution (Fig. 3) with the component planes one can further detect features which are distinguishing for the classes. For example, low values in the features 8–13 refer with high probability to wildtype pyramidal neurons.

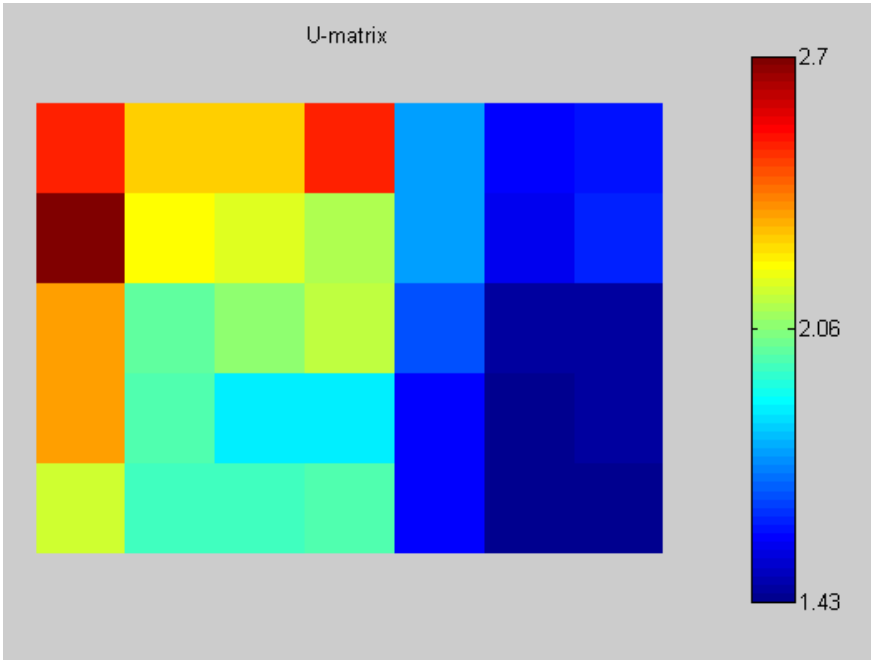

Fig. 5. U-Matrix of generated SOM

The topological ordering allows further investigations by computing the U-Matrix. Fig. 5 shows the U-Matrix of the 28 wildtype and 28 *synRas* pyramidal neurons on a size 4×3 SOM. The U-Matrix indicates that the data space can be divided in regions of different data density. The right part of the map has very high density whereas in the left part the density is moderate or low (left upper corner). Further, we see from the frequency map (Fig. 3) that the data of the wildtype pyramidal neurons are mainly mapped onto the dense part of the map. Hence, the parameter of these data have a low variance whereas the variety for the transgenic pyramidal neurons is higher (upper left corner).

4 Conclusion

In this paper, we described a neuromorphometry protocol and its application for classification of cortical neurons using Kohonen SOM. We employed the visualization capabilities of the SOM method – the U-matrix and the component plane representation – to analyze samples of pyramidal neurons of the somatosensory

cortex of wildtype and transgenic mice with respect to their shape properties. Our primary goal was to explore the question of whether pyramidal neurons from layer II/III of the somatosensory cortex of wildtype and of transgenic mice should be considered as members of two different classes or not.

In previous studies, we obtained ambiguous results which needed clarification. In Ref. [2], we found that the volume of the neocortex of synRas mice is expanded up to 25% as compared to wildtype mice. This is due to the dramatically enlarged volume of the cortical pyramidal cells caused mainly by increased dendritic diameter and tree degree, whereas the number of neurons remains unchanged. Changes are generally less prominent in layers II/III than in layer V. For example, topological analyses revealed significant differences between *synRas* and wildtype mice regarding any parameters considered, i.e., number of intersections, branching points (nodes) and tips (leaves), in both basal and apical dendrites of layer V neurons but not of layers II/III neurons. Thus, while pyramidal neurons from layer V of *synRas* and wildtype mice cortex can be clearly considered as members of different classes, the situation has been ambiguous for layer II/III neurons.

Our study has demonstrated that the SOM can be successfully employed for resolving the shape classification problem of layer II/III neurons. As a result of SOM–learning, pooled data of the two samples were clustered with classification accuracy of 77%. This means, the data were non-random partitioned but with class overlap. Both the topological product of 0.0078 and the component planes of the map indicate a good topology preservation.

In particular, the consideration of the component planes suggests that selected features may be sufficient for classification. For this investigation a more detailed analysis is required. In future work, we plan to use the method of supervised learning vector quantization with relevance learning for classification, which automatically provides a classification dependent feature weighting [16].

Acknowledgements

We thank our student C. Roll for technical contributions to the cluster analysis. This work was supported in part by Deutsche Forschungsgemeinschaft (Grant GA 716/1-1).

References

1. Schierwagen, A.: Mathematical and Computational Modeling of Neurons and Neuronal Ensembles. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) Computer Aided Systems Theory - EUROCAST 2009. LNCS, vol. 5717, pp. 159–166. Springer, Heidelberg (2009)
2. Alpár, A., Palm, K., Schierwagen, A., Arendt, T., Gärtner, U.: Expression of constitutively active p21H-rasVal12 in postmitotic pyramidal neurons results in increased dendritic size and complexity. J. Comp. Neurol. 467, 119–133 (2003)
3. Cannon, R.C.: Structure editing and conversion with cvapp (2000), `http://www.compneuro.org/CDROM/nmorph/usage.html`

4. Van Pelt, J., Schierwagen, A.: Morphological analysis and modeling of neuronal dendrites. Math. Biosciences 188, 147–155 (2004)
5. Schierwagen, A.: Neuronal morphology: Shape characteristics and models. Neurophysiology 40, 366–372 (2008)
6. Schierwagen, A., Costa, L.F., Alpár, A., Gärtner, A.U., Arendt, T.: Neuromorphological Phenotyping in Transgenic Mice: A Multiscale Fractal Analysis. In: Deutsch, A., et al. (eds.) Mathematical Modeling of Biological Systems, vol. II, pp. 191–199. Birkhuser, Boston (2007)
7. Scorcioni, R., Polavaram, S., et al.: L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. Nat. Protocols 3, 866–876 (2008)
8. Kohonen, T.: Self-Organizing Maps. Springer, Heidelberg (1997)
9. Heskes, T.: Energy functions for Self-Organizing Maps. In: Oja, E., Kaski, S. (eds.) Kohonen Maps, pp. 303–316. Elsevier, Amsterdam (1999)
10. Bauer, H.-U., Herrmann, M., Villmann, T.: Neural Maps and Topographic Vector Quantization. Neural Networks 12, 659–676 (1999)
11. Villmann, T., Der, R., Herrmann, M., Martinetz, T.: Topology Preservation in Self–Organizing Feature Maps: Exact Definition and Measurement. IEEE Transactions on Neural Networks 8, 256–266 (1997)
12. Bauer, H.-U., Villmann, T.: Growing a Hypercubical Output Space in a Self–Organizing Feature Map. IEEE Transactions on Neural Networks 8, 218–226 (1997)
13. Bauer, H.-U., Pawelzik, K.: Quantifying the neighborhood preservation of Self-Organizing Feature Maps. IEEE Transactions on Neural Networks 3, 570–579 (1992)
14. Ultsch, A., Siemon, H.P.: Kohonen's self–organizing feature maps for exploratory data analysis. In: Proceedings of ICNN 1990, International Neural Network Conference, pp. 305–308. Kluwer, Dordrecht (1990)
15. Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J.: SOM Toolbox for Matlab 5. Report A57, April 2000. Helsinki University of Technology, Finland (2000)
16. Hammer, B., Villmann, T.: Generalized Relevance Learning Vector Quantization. Neural Networks 15, 1059–1068 (2002)