

# SCIENTIFIC REPORTS



OPEN

## Discrimination of the hierarchical structure of cortical layers in 2-photon microscopy data by combined unsupervised and supervised machine learning

Dong Li<sup>1</sup>, Melissa Zavaglia<sup>1,5</sup>, Guangyu Wang<sup>2,3</sup>, Hong Xie<sup>4</sup>, Yi Hu<sup>3,2</sup>, Rene Werner<sup>1</sup>, Ji-Song Guan<sup>2,3</sup> & Claus C. Hilgetag<sup>1,6</sup>

The laminar organization of the cerebral cortex is a fundamental characteristic of the brain, with essential implications for cortical function. Due to the rapidly growing amount of high-resolution brain imaging data, a great demand arises for automated and flexible methods for discriminating the laminar texture of the cortex. Here, we propose a combined approach of unsupervised and supervised machine learning to discriminate the hierarchical cortical laminar organization in high-resolution 2-photon microscopic neural image data of mouse brain without observer bias, that is, without the prerequisite of manually labeled training data. For local cortical foci, we modify an unsupervised clustering approach to identify and represent the laminar cortical structure. Subsequently, supervised machine learning is applied to transfer the resulting layer labels across different locations and image data, to ensure the existence of a consistent layer label system. By using neurobiologically meaningful features, the discrimination results are shown to be consistent with the layer classification of the classical Brodmann scheme, and provide additional insight into the structure of the cerebral cortex and its hierarchical organization. Thus, our work paves a new way for studying the anatomical organization of the cerebral cortex, and potentially its functional organization.

Although the mammalian cerebral cortex is not particularly thick, its laminar structure is remarkably complex<sup>1–5</sup>. More than a century ago, the classical Brodmann scheme identified six cortical layers, followed more recently by findings of sub-layers within them<sup>6,7</sup>, indicating the existence of a hierarchical (encapsulated) organization.

A hierarchical organization exists on many levels in the brain, for instance with regard to the anatomical connectivity of cortical areas<sup>8,9</sup>, their functional connectivity<sup>10</sup>, or intrinsic time scales across primate cortex<sup>11</sup>. The motivation to study such a hierarchical neural organization ultimately derives from the desire to understand brain function. With respect to the cerebral cortex, a number of studies have shown functionally discriminable properties of cortical layers. For instance, in rat primary somatosensory cortex, the response latency to external stimuli in layer Va is significantly different from layers Vb, IV and VI<sup>12</sup>; and in mouse cortex, memory trace neurons are found primarily in layer II/III rather than layer IV<sup>13</sup>. From a technical point of view, calcium imaging or gene expression imaging in brain tissue achieves a spatial resolution in the micrometer range<sup>14–23</sup>, allowing detailed analysis of, for instance, the mouse cortex (thickness around 1 mm)<sup>24</sup>. The use of modern imaging techniques for further investigation of the laminar hierarchy of cerebral cortex, therefore, promises further insights into brain

<sup>1</sup>Institute of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, 20246, Germany. <sup>2</sup>School of Life Science and Technology, ShanghaiTech University, Shanghai, 201210, China. <sup>3</sup>School of Life Sciences, Tsinghua University, Beijing, 100086, China. <sup>4</sup>Institute of Brain-Intelligence Technology, Zhangjiang Laboratory, Shanghai Research Center for Brain Science and Brain-Inspired Intelligence, Shanghai, 200031, China. <sup>5</sup>Focus Area Health, Jacobs University Bremen, Bremen, 28759, Germany. <sup>6</sup>Department of Health Sciences, Boston University, Boston, MA, 02215, USA. Dong Li, Melissa Zavaglia and Guangyu Wang contributed equally. Correspondence and requests for materials should be addressed to C.C.H. (email: [c.hilgetag@uke.de](mailto:c.hilgetag@uke.de))

function, but corresponding analysis approaches have to be able to deal with, and take advantage of, currently recorded massive high-resolution neural imaging data.

The conventional way of manually labelling cortical layers according to a classical atlas clearly faces problems in this era of massive high-resolution neuroimaging data<sup>25</sup>. Cortical laminar patterns vary from area to area, from species to species, and even from individual to individual within the same species. Whereas the increasing availability of a large amount of data is potentially a great advantage for research, it is too labor and time intensive to label the cortical layers for every location in every animal and research project. Moreover, manual labelling entails an observer-specific bias, particularly in areas without clearly visible traits for differentiating the layers, or when going beyond classically established layers to identify potential sub-layers. In addition, although the classical definition of laminar layers and their hierarchy contributed much to cortical research, by grouping neurons for comparable data analysis and computational modeling, it sometimes appears overly restrictive. For instance, it suggests that, in terms of hierarchical organization, Brodmann's six layers can be considered to reside on the same hierarchy level, and further sub-layers on the levels below. However, particularly with respect to the cytoarchitecture of rodents, it has been found that Brodmann's layers II and III display high anatomical similarity<sup>25</sup>, whereas Brodmann's layer Va significantly differs from other sub-layers of layer V<sup>1,26–30</sup>. Thus, there is a great demand for neuroimaging data processing solutions that can address these challenges, by providing an automatic and objective differentiation, but also flexible definition of cortical laminar hierarchies.

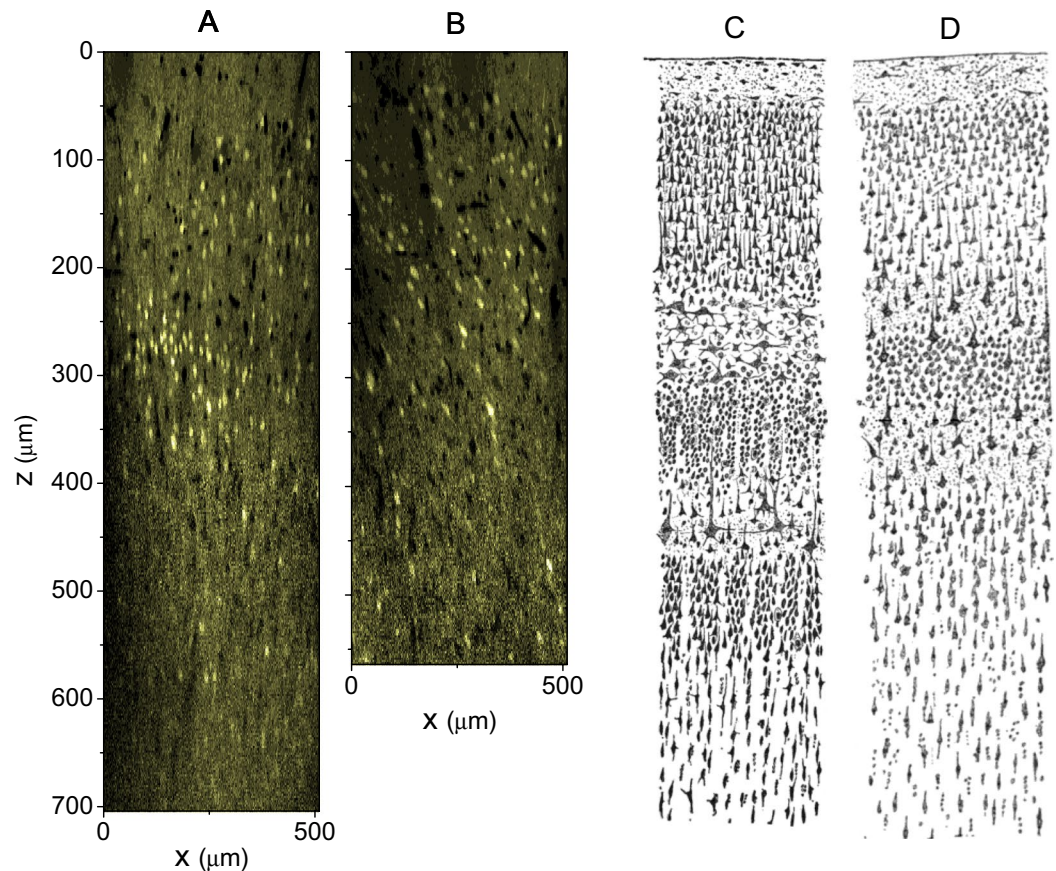
Machine learning is a very promising approach in this regard, and existing findings of studies on conventional layer labelling and laminar properties of cortical neurons, such as morphological<sup>31</sup>, physical<sup>32–36</sup>, or functional<sup>37</sup> properties of the neurons, already provide a range of well-characterized features that can be employed for machine learning-based laminar pattern discrimination.

Some efforts have already been made to automatically and objectively discriminate neurons belonging to different layers by supervised machine learning approaches<sup>38–41</sup>, yielding appealing laminar classifications; yet, the training data sets and methods are still limited by the framework of conventional atlases. The existing methods can, therefore, address the issue of labor intensity, but do not yet contribute to the other challenges. In contrast, unsupervised learning and clustering algorithms<sup>42</sup> allow, in principle, for more flexibility. So far, however, existing work in this context still requires the predefinition of a desired number of layers and this aspect, in turn, limits the required layer and sub-layer definition flexibility. In the present work, we, therefore, propose combining unsupervised and supervised machine learning to discriminate the hierarchical organization of cortical layers. To illustrate the feasibility of the approach, the developed methodology is applied to 2-photon microscopic image data from mouse cortex. For small cortical foci, we use unsupervised clustering to identify the laminar cortical structure. During clustering, we do not demand a particular number of layers per hierarchical level, but automatically set up the level-specific layer structure. Supervised machine learning is then used to bridge the hierarchical clustering results obtained at different cortical locations and animals, that is, to ensure existence of a consistent layer labelling system across cortical locations. Manual labelling is, therefore, not required for the discrimination process. We, nevertheless, compare the classification results with existing laminar cortical schemes as part of an extensive evaluation to demonstrate plausibility of our layer discrimination results. In particular, using neuro-biologically meaningful features, we illustrate that our discrimination results are consistent with the classical Brodmann scheme of cortical layers, and in addition provide useful information on the hierarchical laminar organization in different cortical areas.

## Materials and Methods

**Imaging data description.** *Data acquisition.* The used data was high-resolution 2-photon imaging mouse data previously described by Xie *et al.*<sup>13</sup>. Specifically, the mouse strain is BAC-EGR-1-EGFP (Tg(Egr1-EGFP)GO90Gsat/Mmucd from the Gensat project, distributed from Jackson Laboratories. Animal care was in accordance with the institutional guidelines of Tsinghua University, and the whole experimental protocol was also approved by Tsinghua University. 3–5 months old mice received cranial window implantation as previously described<sup>13</sup>; recording began one month later. To implant the cranial window, the animal was immobilized in custom-built stage-mounted ear bars and a nosepiece, similar to a stereotaxic apparatus. A 1.5 cm incision was made between the ears, and the scalp was reflected to expose the skull. One circular craniotomy (6–7 mm diameter) was made using a high-speed drill and a dissecting microscope for gross visualization. A glass-made coverslip was attached to the skull. For surgeries and observations, mice were anesthetized with 1.5% isoflurane. EGFP fluorescent intensity (FI) was imaged with an Olympus Fluoview 1200 MPE with pre-chirp optics and a fast AOM mounted on an Olympus BX61WI upright microscope, coupled with a 2 mm working distance, 25x water immersion lens (numerical aperture 1.05).

*Data characteristics.* For each mouse, usually 10 to 20 cortical locations were monitored. Each location was 510  $\mu\text{m}$   $\times$  510  $\mu\text{m}$  broad and scanned until a depth of between 300  $\mu\text{m}$  and 800  $\mu\text{m}$  (see, for example, Fig. 1), containing a few to over ten thousand neurons. The image resolution in *z* direction was always 2  $\mu\text{m}$ ; in-plane resolution was 0.996  $\mu\text{m}$   $\times$  0.996  $\mu\text{m}$ . Each mouse was scanned once per day, for between a few and up to tens of days. In this study, we randomly chose sixty locations from eight mice, including 10 in motor areas, 7 in Posterior parietal (PTLp) areas, 18 in retrosplenial cortex (RSC) areas, 7 in primary somatosensory (SSp) areas, 6 in anterior medial visual (VISam) areas, 8 in primary visual (VISp) areas, 2 in unspecified visual (VIS) areas, 1 location on the boundary between Motor and SSp and 1 location around the junction of PTLp, SSp and RSC. Each location was named in such a way that it was distinguishable. The area types were manually annotated by one experimental expert (G.W.) and approved by all other experimental experts among the authors (Y.H., H.X., and J.-S.G.). Technically, the physical locations of the datasets and the visual perception of the spatial distribution of neuron sizes and densities along the *z* direction of the image data were taken into consideration for annotation. For subsequent analyses, the neuron positions in the images were automatically detected as described in detail by

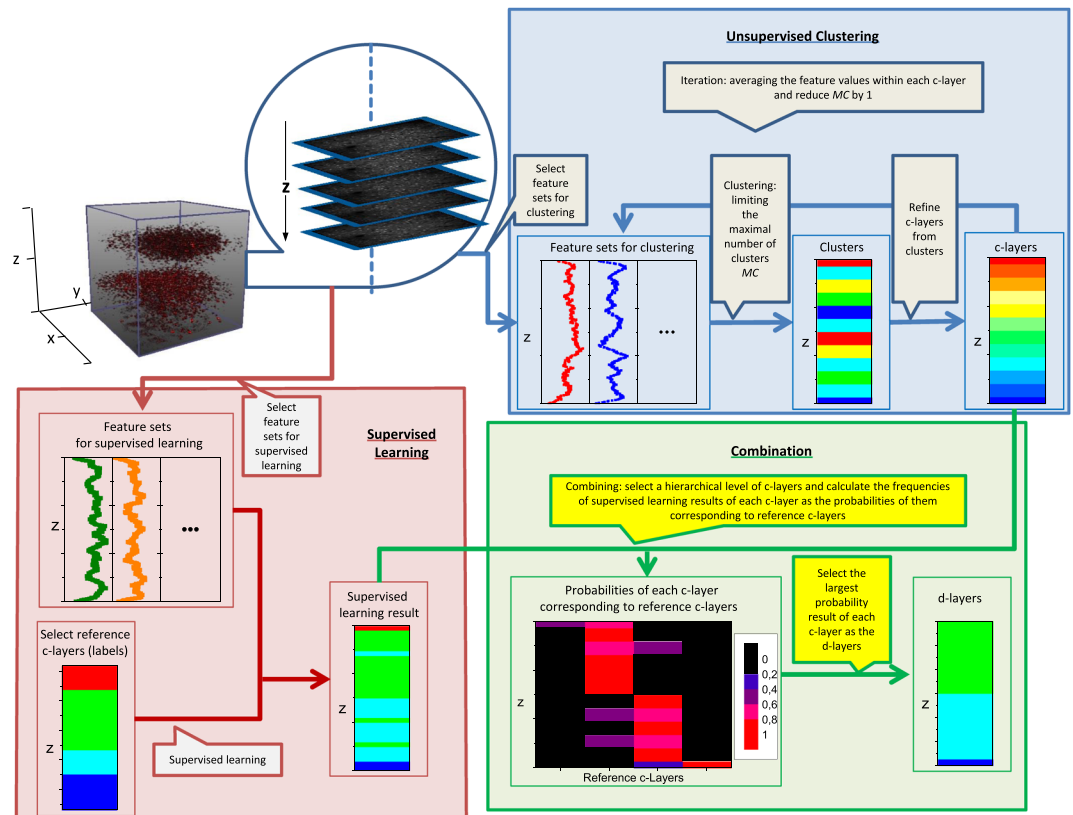


**Figure 1.** Example slices of x-z planes of (A) VISp area and (B) motor area of mouse brain in our study. For comparison purposes, we adapted the drawings of cortical lamination by Santiago Ramon y Cajal of (C) Nissl-stained visual cortex and (D) Nissl-stained motor cortex of a human adult<sup>51</sup>.

Xie *et al.*<sup>13</sup>. The main part of the underlying algorithm was a convolutional neuronal network (CNN) model that was trained in supervised manner based on manual segregation example data (format: 2-dimensional patches of size  $17 \times 17$  pixel). For neuron detection, candidate patches were extracted using a sliding window approach on different scales, applied to the x-y slices of our data. Resulting patches were rescaled to the required size of  $17 \times 17$  pixel and input to the CNN. To avoid multiple detection of a neuron, positive responses within  $6 \times 6 \times 6$  pixel areas of the image data were aggregated and considered a unique neuron, with its position being the centre location of the positive responses. The approach was already demonstrated to result in highly accurate neuron position detection<sup>13</sup>.

**Manual labelling of Brodmann layers.** For all sixty cortical locations, an experimental expert (G.W.) labelled those Brodmann layers (B-layers) that could be unambiguously discriminated by eye in the location-specific image data. The labels were assigned to entire x-y image slices. Depending on the particular locations, the assigned Brodmann layers included (1) the superficial layer, (2) Brodmann layers II/III, (3) Brodmann layer IV, (4) Brodmann layer Va, (5) Brodmann layer Vb, (6) Brodmann layer VI, and (7) deeper parts. Each location was labelled with between four and seven of the Brodmann layers, depending on the area type and imaging specifics. For instance, some areas did not contain a visible layer IV or layer Va, and some image data sets did not cover all layers due to limited scanning depth. The manually assigned Brodmann layers were used for evaluation purposes of the automatic layer discrimination results. To further assess manual labelling uncertainty, a second experimental expert (Y.H.) repeated assignment of the Brodmann layers following the same instructions as the first rater but being blinded to respective results. Figure S1 summarizes the interrater difference distribution of corresponding boundary positions, illustrating an interrater variability and boundary position uncertainty in the order of  $\pm 5$  slices.

**Combined super- and unsupervised layer discrimination: principle.** The overall pipeline of the proposed approach to discriminate the hierarchical laminar structure of the acquired data is shown in Fig. 2. The input observations were 3-dimensional image stacks  $I: (x, y, z) \in \Omega \subset \mathbb{Z}^3 \mapsto I(x, y, z)$  with  $I(x, y, z)$  as image intensity at pixel location  $(x, y, z)$ . Further,  $I_z$  denoted the image slice at depth  $z$  and  $I_{[z_1, z_2]} = I|_{\Omega_{[z_1, z_2]}}$  the restriction of  $I$  to  $\Omega_{[z_1, z_2]} := \{(x, y, z) \in \Omega \mid z_1 \leq z < z_2\}$ , that was, all slices between  $z_1$  and  $z_2$ . The overall aim was to discriminate layers  $I_{[1, z_1]}$ ,  $I_{[z_1, z_2]}$ , etc. and to represent their hierarchical organization in an objective and meaningful



**Figure 2.** Proposed pipeline for the discrimination of hierarchical laminar cortical organization. The top-right panel shows the unsupervised clustering part, whereas the bottom-left panel depicts the supervised learning part. The bottom-right panel illustrates how to combine the unsupervised and supervised learning results.

manner for all studied cortical locations. To this end, we developed a machine learning-based approach inspired by the human's recognition process of cortical layers, consisting of two main parts. Within each cortical location, the discrimination of the laminar hierarchical structure was first obtained via clustering (i.e., unsupervised machine learning). The clustering results for the individual locations were then combined using supervised learning in order to transfer layer labels on particular hierarchical levels that were obtained for specific reference locations to the other locations. Thus, the supervised learning step ensured consistent layer labelling across different locations.

**Terminology.** The terminology *layer* referred to several but different concepts in this study. On an abstract level, layer simply meant a collection of z-slices  $I_{[z_1, z_2]}$ . In this work, the layers that were manually labelled according to Brodmann's atlas were referred to as Brodmann layers (*B-layers*). During unsupervised hierarchical clustering, we further obtained a certain number of layers on each hierarchical level, which were called *c-layer* (*c* is short for clustering), followed by an Arabian number that indicated the vertical order from superficial to deep. When *c-layers* were used as training data in the supervised learning part, they would be denoted as *reference c-layers*. Those *c-layers*, to which the layer labels have to be transferred to, were denoted as *target c-layers*. The results of this final discrimination would be called *d-layers* (*d* for discriminated), followed by an Arabian number, indicating its corresponding position in the reference *c-layer* dataset.

**Part 1: Hierarchical clustering for unsupervised layer definition.** For each cortical location, an iterative procedure was applied to compute the hierarchical structure of the data. Briefly speaking, we iteratively performed feature-based clustering of the z-slices of  $I$ , refined the clustering results in order to define iteration-specific *c-layers* and successively decreased the maximum possible number of clusters, with the maximum possible number of clusters subsequently denoted as  $MC$  (see the unsupervised learning panel in Fig. 2). The entire iteration process resulted in a global hierarchical tree, with the individual hierarchical levels representing the clustering results after the different iterations.

In detail, for each iteration, we first applied a classical hierarchical tree clustering. For the first iteration,  $MC$  was initialized by  $MC_0$ .  $MC_0$  can, in principle, be chosen arbitrarily, but not be too small. In our work, we chose  $MC_0 = 10$ . Each clustering iteration then resulted in a certain number of clusters, with the clusters distributing along the z-direction of the image data set. Yet, depending on the considered features, different isolated *c-layers* might be assigned to the same cluster (e.g. layer  $I_{[z_1, z_2]}$  to cluster 1,  $I_{[z_2, z_3]}$  to cluster 2, and  $I_{[z_3, z_4]}$  again to cluster 1 with  $z_1 < z_2 < z_3 < z_4$ ). As this was not the desired output, every isolated part of a cluster was defined as a specific *c-layer*, which meant that one cluster might be divided into more than one *c-layer* and the overall number of

c-layers was larger or equal to the number of clusters  $MC$  (see again Fig. 2). After defining the c-layers for a specific iteration, the feature values for the individual z slices were replaced by the average values of the features corresponding to all the z-slices of the c-layer encompassing the considered z-slice. With these updated feature values and  $MC$  decreased by one, the next clustering iteration was initiated. This procedure was repeated until  $MC$  had a value of 'one'. Thus, after each iteration, a series of new c-layers was computed that emerged from merging the c-layers of the previous iteration, i.e. the last lower hierarchical level.

From an implementation point of view, clustering was performed using the *linkage* algorithm of the MATLAB Statistics and Machine Learning Toolbox, with the *Euclidean* distance measure in feature space and the minimum variance algorithm (*ward*) as parameters.

**Part 2: Across-location-consistent labelling of c-layers by supervised learning.** To provide c-layer labels and a layer discrimination that was consistent for the different locations, we proposed combining the unsupervised clustering results with a supervised learning procedure (see the supervised learning panel and combination panel in Fig. 2). To this end, we first chose a reference dataset with rich laminar pattern, so as to cover the laminar characteristics and across-layer differences of as many layers as possible. Here, the reference data sets and respective reference c-layers were therefore selected from either VISp or SSp areas, that was, areas known to have well differentiable laminar layers<sup>2,5</sup>. The hierarchical levels on which reference and target c-layers were chosen for combination were in principle arbitrary, but the level should be similar for reference and target data set. In this study, for evaluation purposes, we always selected the 8<sup>th</sup> hierarchical level for the reference location and the 7<sup>th</sup> level in the target locations, as on those levels, the c-layer numbers were most similar to the number of expert-delineated B-layers.

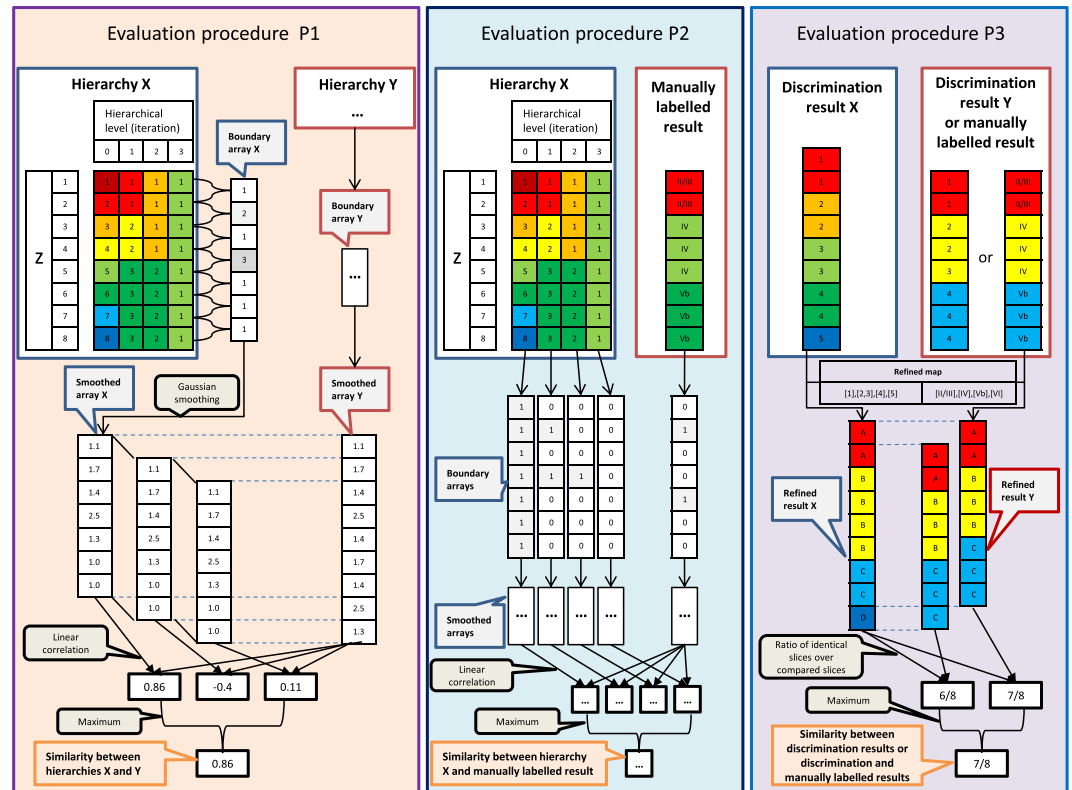
To establish correspondence between the target and the reference c-layers, we first trained an ensemble of support vector machines (SVMs) based on the z-slices and slice-wise computed features of the reference data set. In total,  $m$  SVMs (in this work:  $m = 200$  for each reference data set and each feature set) were trained, based on randomly selected slices of the reference data set and respective features.

The SVMs were then, individually for each target c-layer, applied to determine the corresponding reference c-layer. Let the reference c-layer labels be denoted by index  $r = 1, \dots, r_{\max}$  and the target c-layers by  $t = 1, \dots, t_{\max}$ . Each target c-layer  $t$  further consisted of  $n_t$  image slices  $\tilde{I}_1, \dots, \tilde{I}_{n_t}$  and respective features. Applied to each target c-layer slice, the SVM ensemble resulted in  $m$  slice-specific reference c-layer votes; repeating SVM application for all slices  $\tilde{I}_1, \dots, \tilde{I}_{n_t}$  yielded in total  $m \cdot n_t$  votes. Let  $n_r$  be the number of votes for reference c-layer  $r$ , the probability  $p(r) = n_r / (n_t \cdot m)$  was interpreted as similarity measure between target c-layer  $t$  and reference c-layer  $r$ . The reference c-layer with maximum similarity was then assigned to the target c-layer and referred to as d-layer.

The supervised machine learning part was also implemented in MATLAB, using the package LIBSVM<sup>43</sup> and linear kernel SVMs with cost parameter = 1.

**Definition of used features and feature sets.** The following features were considered in this work and computed for the individual image slices  $I_z$ ,  $z = 1, \dots, z_{\max}$  of the 60 areas and corresponding image data sets described in section *Image data description*.

- **Neuron density:** Neuron density was the number of neurons per slice, normalized by the area of the slice (i.e.,  $510 \mu\text{m} \times 510 \mu\text{m}$ ). Neuron density was evaluated based on the automatic neuron detection described in the *Imaging data description* section.
- **Neuron size and neuron shape:** Precise computation and representation of neuron size and neuron shape required accurate segregation of the cell bodies, which is challenging by itself at the moment<sup>44,45</sup> (see also, for instance, the 2018 Data Science Bowl challenge from the Kaggle competition)<sup>46</sup> and beyond the scope of this paper. Instead, we only used rough measures that reflect mean neuron size and 2D shape for the individual slices  $I_z$ . As a result of the automated neuron detection, the locations of the neuron centres were known for the individual slices, and each neuron centre location was associated with a local maximum of the image intensity. To determine neuron size, the neuron soma boundary was estimated based on intensity, using half the image intensity value at the neuron centre relative to the background image intensity as a threshold; the sought size was defined as the number of the enclosed pixels, multiplied by the pixel spacing. To avoid negative influence of outlier values, the *robust fitting* function in MATLAB was used to finally calculate the average size of the neuron somata across all days of scanning for the particular location and image slice. To represent neuron shape, we computed a shape measure that reflected the difference of the shape of the 2D neuron soma area and a circle. Based on the soma delineation described above, the maximum soma diameter  $d$  was determined. The shape measure was then defined by  $(\pi d^2)/(4A)$ , with  $A$  denoting the soma size. Mean neuron shape measures were again computed as averages over all scanning days and neurons.
- **Neuron type:** Accurate discrimination of various types of neurons is currently also a topical challenge<sup>40,47</sup> and, once again, not a direct target of this work. Instead, we only classified the detected neurons based on their shape and size (see definition above) into two groups, corresponding potentially to granule and pyramidal neurons. The classification was performed using a two-class Gaussian mixture model.
- **Image texture features:** To compare the performance of the neuron-related features for layer discrimination to other typical image-based features without direct relation to a neuro-biological meaning, we also computed and analysed the intensity co-occurrence matrices corresponding to the image slices. The following four image texture measures were considered (using the MATLAB function *graycoprops*): *Contrast* (average intensity difference between neighbouring pixels), *Correlation* (average correlation of intensity values of neighbouring pixels), *Energy* (also known as uniformity of an image), and *Homogeneity*.



**Figure 3.** Illustration of applied evaluation procedures. The left panel shows the pipeline to evaluate the similarity between two hierarchical clustering results X and Y. X and Y could, for example, be two hierarchical trees obtained for one image data set but with different feature sets, or two hierarchy trees of data acquired at different locations. The middle panel illustrates the pipeline to evaluate similarity between a hierarchical clustering result and the corresponding manually labelled layers. The right panel shows the pipeline to evaluate two final discrimination results or a respective discrimination result and the corresponding manually assigned layer labels. Again, the two discrimination results can either indicate results computed for a single location but with different feature sets, or results from two different locations. For the latter case, the image data potentially contain a different number of slices.

**Considered feature sets.** In total, we defined six sets of features to demonstrate and compare their performance for layer discrimination: Feature set F1 (only neuron density; i.e. one feature), F2 (neuron density and mean neuron size; two features), F3 (neuron density, the proportions of the two neuron classes, and the mean neuron sizes of the two classes; in total five features), F4 (F3, plus the mean size of all neurons; six features), F5 (neuron density, mean neuron size, and the four image texture features; in total six features), and F6 (only the four image texture features).

**Evaluation strategies.** In line with the structure of the proposed layer discrimination approach, the evaluation consisted of two parts: to evaluate the performed unsupervised clustering and to evaluate the supervised learning-based labelling of the target data  $c$ -layers. Methodologically, this leads to the tasks of measuring (P1) similarity between two computed laminar hierarchies, (P2) measuring similarity between a hierarchical clustering result and a (manually labelled) layer discrimination, and (P3) the direct comparison of two layer discrimination results, either manually or automatically obtained. The applied procedures are shown in Fig. 3 and detailed below.

**Evaluation of unsupervised hierarchical clustering.** To evaluate the results of the unsupervised hierarchical clustering, procedures (P1) and (P2) were used.

Procedure (P1) was applied to study the influence of the introduced feature sets on the hierarchy trees obtained by clustering. As shown in Fig. 3 (left panel), we first introduced a layer boundary function  $\Delta L^F$  that, for a given hierarchical tree after clustering based on a feature set  $F$ , assigns the particular tree level (traversing the tree in bottom-up direction) to each pair of consecutive slices ( $I_z, I_{z+1}$ ) of the image data and area at hand, on which the slices for the first time were assigned to the same cluster. For example, in Hierarchy X in the left panel of Fig. 3, the first and the second slice merged at the first hierarchical level and, consequently,  $\Delta L(I_1, I_2) = 1$ . The resulting layer boundary array

$$\Delta L^F(I) = \left( \Delta L^F(I_1, I_2), \Delta L^F(I_2, I_3), \dots, \Delta L^F(I_{z_{max-1}}, I_{z_{max}}) \right)^T \in \mathbb{Z}^{z_{max}-1}$$

was Gaussian-smoothed, taking into account potential boundary position uncertainties (here: applying a span of 5 elements). This smoothed boundary array  $\widetilde{\Delta L}^F(I)$  was the basis for subsequent evaluation steps. In particular, for each of the 60 areas, we evaluated the similarity  $\kappa$  of the boundary arrays  $\widetilde{\Delta L}^F(I)$  obtained for two different feature sets by calculating the linear correlation coefficient of the two boundary arrays.

Procedure (P2) was applied to measure the similarity between a hierarchical clustering result for a specific area and the manual layer labelling of the area. The manual labelling could be interpreted as a single-level hierarchy; thus, the boundary array  $\Delta L^{man}(I)$  given by the manual labelling only contained the elements 0 and 1 (see Fig. 3, middle panel). To, nevertheless, derive a similarity measure between  $\Delta L^{man}(I)$  and a feature set-specific hierarchical clustering result, we calculated a set of *binary* boundary arrays, one for each level of the hierarchical tree. If  $l$  denoted the hierarchical level considered, the corresponding binary boundary array  $\Delta L_l^F$  was defined by

$$\Delta L_l^F(I_z, I_{z+1}) = \begin{cases} 1, & \Delta L^F(I_z, I_{z+1}) > l \\ 0, & \text{else} \end{cases}$$

Similar to procedure (P1), after smoothing the binary boundary arrays, we calculated the linear correlation between  $\widetilde{\Delta L}^{man}(I)$  and the individual  $\widetilde{\Delta L}_l^F(I)$  for all levels  $l$ . Finally, the maximum correlation coefficient  $\kappa_m$  of the coefficients obtained for the individual levels  $l$  was used to quantify the similarity between the laminar hierarchy and the manual labelling.

*Evaluation of reference-based layer discrimination and label transfer.* The technical procedure (P3) to evaluate the supervised learning-based transfer of layer labels from a reference to target data sets is described in the right panel of Fig. 3. Different to (P1) and (P2), we now assumed a specific layer discrimination and labelling result for an image data set  $I$  to be given that was to be compared to a ground truth layer discrimination of  $I$ . With  $M^X: \mathbb{Z} \rightarrow L_X$  mapping an image slice  $I_z$  to a layer label  $M^X(z) =: M_z^X$  of an ordered label set  $L_X$  that was specific to the discrimination result  $X$  and a similar function  $M^Y: \mathbb{Z} \rightarrow L_Y$  for the ground truth layer discrimination, the goal was to define a similarity measure between  $M^X$  and  $M^Y$ . As a first step, the potentially different label sets  $L_X$  and  $L_Y$  were mapped to a common label set  $L_{XY}$ . For example, in the right panel of Fig. 3,  $L_X$  represented the d-layer numbers 1, 2, etc, while  $L_Y$  denoted Brodmann layers, indicated by Roman numerals. The different labels were mapped to the joint label set  $L_{XZ}$  and the respective labels A, B, etc. With  $\hat{M}^X: \mathbb{Z} \rightarrow L_{XY}$  as  $X$ -specific mapping of  $I_z$  into  $L_{XY}$  (and analogously for the manually labelled ground truth  $Y$ ), a similarity array was defined by

$$dM(I, X, Y) = \left( dM^{X,Y}(I_1), \dots, dM^{X,Y}(I_{z_{max}}) \right)^T \in \mathbb{Z}^{z_{max}}$$

with

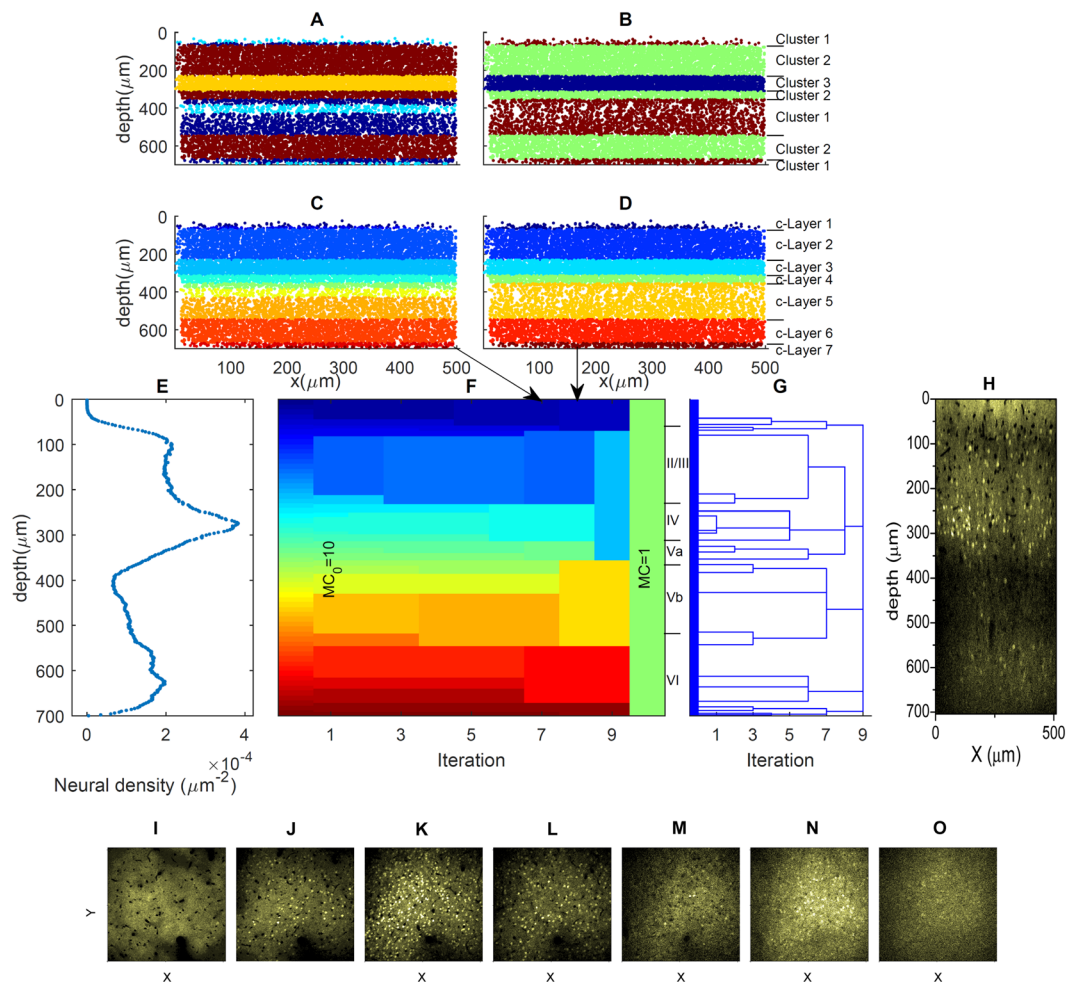
$$dM^{X,Y}(I_z) = \begin{cases} 1, & \hat{M}_z^X = \hat{M}_z^Y \\ 0, & \hat{M}_z^X \neq \hat{M}_z^Y \end{cases}$$

that served as the basis of the performed evaluation. In detail, the respective evaluation covered three aspects:

**Analysis of reference data and reference c-layers:** In a first step, we studied the properties and the quality of the reference data and reference c-layers, which served as the basis of the label transfer to different target data sets. As motivated above, the considered reference data were the 8 VISp and 7 SSsp areas and the c-layers extracted after the 8<sup>th</sup> iteration of unsupervised clustering using the F1 feature set. To alleviate interpretability of subsequent evaluation results (i.e. to ensure that d-layers assigned to target data sets could directly be compared to the target data set B-layers), we visually inspected the consistency of the reference c-layers and the manually assigned reference data B-layers. If necessary, reference c-layers were merged and apparent inconsistencies corrected. In Fig. 4F (data set M248L1), it could, for instance, be seen that c-layers 6 and 7 did not merge at the selected hierarchy level (the 8<sup>th</sup> level), and both layers corresponded to the manually assigned layer 6 (i.e. B-layer VI); the two c-layers were therefore merged. The final mapping between the VISp reference c-layers and the manually assigned labels is summarized in Table 1; analogous information for SSsp can be found in Table S1.

Based on the refined reference c-layers and the refined B-layers labels, two quality measures,  $n_B$  and  $f_{c=B}$ , were introduced and evaluated for the individual reference data sets.  $n_B$  denoted the total number of the manually labelled layers that the data set covered, which should be close to seven. Yet, due to limited scanning depth, the reference data sets did not necessarily cover all possible layers (see section *Image data description*), resulting in a smaller number, and limiting its usability as a reference data set.  $f_{c=B}$  represented the number of slices where both the assigned refined c-layer and the assigned B-layer pointed to the same cluster, divided by the total number of slices of the data set. Referring to the aforementioned definitions,  $f_{c=B}$  equalled the averaged value of the entries of the similarity array  $dM(I, c, B)$ ; here,  $c$  indicated reference c-layer-based and  $B$  the B-layer-based discrimination of the data set  $I$ . Thus,  $f_{c=B}$  reflected the performance of the unsupervised clustering of  $I$ , with  $f_{c=B} = 1$  as ideal performance.

**Comparison of d- and B-layers (all locations):** To further evaluate the performance of supervised label transferring and the influence of the feature sets thereon, we transferred the refined reference c-layers to all locations



**Figure 4.** Example of unsupervised clustering (VISp area M248L1). Panels (A,B) show the distribution of the clusters as direct output of the original clustering algorithm and panels (C,D) the respective layers after 7<sup>th</sup> and 8<sup>th</sup> iterations. Panels (F,G) represent the whole iteration process and the hierarchical tree obtained with neuron density as unique feature (panel E). The manually labelled layers are indicated to the right of panel (F). Panel (H) shows a corresponding x–z slice, and bottom panels seven x–y slice examples at depth  $z = 40 \mu\text{m}$  (I),  $z = 160 \mu\text{m}$  (J),  $z = 280 \mu\text{m}$  (K),  $z = 320 \mu\text{m}$  (L),  $z = 460 \mu\text{m}$  (M),  $z = 620 \mu\text{m}$  (N), and  $z = 690 \mu\text{m}$  (O), corresponding to the seven c-layers on the 8<sup>th</sup> hierarchical level of panel (F).

and data sets. Thus, for each of the 15 reference data sets and 6 feature sets, 60 target datasets were considered and corresponding d-layers determined. This final d-layer definition was evaluated by slice-wise comparison of the d-layer- and the manually assigned B-layer-based discrimination, that was, by means of the measure  $f_{d=B}$  that represented the fraction of the target data set slices with agreement of its associated d- and B-layers.

Re-labelling of reference data: As a third aspect, we constrained the aforementioned comparison to the reference datasets themselves serving as target datasets, but still using the different feature sets F1–F6. To measure the performance, we introduced the index  $f_{d=c}^{\text{ref}}$  that represented the ratio of the number of reference dataset slices for which the original refined c-layers and the assigned d-layer were identical, and the total number of slices, i.e. the mean value of  $dM(I, c, d)$ . Ideally, if all slices within one reference c-layer had uniform but unique feature expressions,  $f_{d=c}^{\text{ref}}$  equals 1. Yet, this was not necessarily the case because, for instance, the unsupervised clustering approach comprised spatial connectivity constraints, while the supervised learning did not. The index  $f_{d=c}^{\text{ref}}$  could, therefore, be understood as a measure to evaluate the relevance of individual features and appropriateness of the considered feature sets for supervised learning-based label transfer.

*Analysis of cross-location similarity of discrimination results.* As final part of the evaluation, we analysed the similarity of layer discrimination results between any pair of locations (i.e., in total  $60 \times 60$  pairs) for unsupervised clustering as well as supervised learning and the different feature sets. From a neuro-biological perspective, image data from locations of the same area type could be expected to have more similar hierarchical laminar patterns than images from different area types; this should be reflected by automatic layer discrimination and appropriate features sets. For unsupervised clustering, we again calculated hierarchical laminar similarity using procedure (P1) and the linear correlation coefficient  $\kappa$ . For supervised learning, the d-layers were computed as



Reference location name	Reference c-layer structure	Manually labelled layer structure
M226L5*	[1, 2], [3], [4, 5]	[1, 2], [3], [5]
M248L1	[1], [2], [3], [4], [5], [6, 7]	[1], [2], [3], [4], [5], [6]
M248R3**	[1], [2, 3], [4], [5]	[1], [2, 3], [4], [5]
M262L2	[1], [2], [3], [4], [5], [6, 7], [8, 9]	[1], [2], [3], [4], [5], [6], [7]
M262L3	[1], [2], [3], [4], [5], [6], [7]	[1], [2], [3], [4], [5], [6], [7]
M262R3	[1], [2], [3], [4], [5]	[1], [2], [3], [4], [5]
M339R5	[1], [2], [3], [4, 5], [6, 7]	[1], [2], [3], [5], [6]
M339R6	[1], [2], [3], [4]	[1], [2], [3], [5]

**Table 1.** Refined maps between Reference c-layers from VISp areas (F1 feature set and the 8<sup>th</sup> level on the hierarchy) and the manually labelled layers. Layer numbers grouped in parentheses were layers that were considered as one single layer during evaluation; see respective explanations for details. \*Visual inspection reveals that the reference c-layer 1 looked more like a combination of Brodmann's superficial layer and layer II, and c-layer 2 like the layer III. Therefore, c-layers 1 and 2 were merged, and analogously the manually labelled layers 1 and 2 were merged. \*\*The transition locations between, on the one hand, reference c-layers 2 and 3 and, on the other hand, the manually labelled layers 2 and 3 were far apart (see Fig. S2). To ensure labelling consistency, we merged layers 2 and 3 in both cases.

described above. d-layer-based discrimination results computed for two (different) image data sets were then compared by means of the index  $f_{d_1=d_2}$  that resembled the introduced measure  $f_{d=B}$  but using the d-layers of the first image (denoted by  $d_1$ ) and the d-layers of the second image (indicated by  $d_2$ ). Yet, the image data sets acquired at different locations contained in parts a different number of slices. If so, we applied a sliding-window approach as illustrated in Fig. 3, left and right panels: The smaller boundary array was compared with all possible chunks of the longer array; the maximum similarity index was finally regarded as the desired measure of similarity.

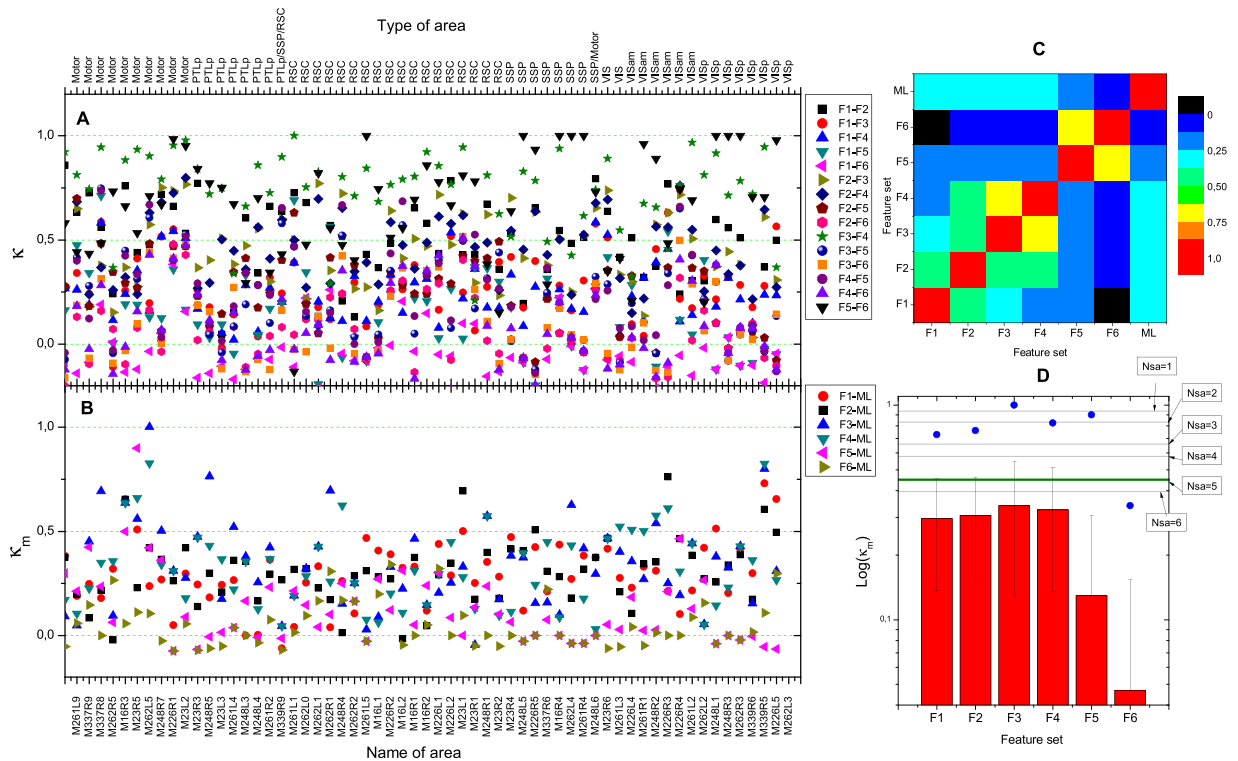
## Results

**Unsupervised clustering.** Figure 4 shows an unsupervised clustering result example for a VISp location, using the simplest neuro-biological feature set F1, that was, only neuron density. The feature values for the different z-slices are shown in panel E of the figure. Panels A and B illustrate the distribution of the clusters and panels C and D the c-layers after the 7<sup>th</sup> and 8<sup>th</sup> iteration of the clustering algorithm and the corresponding hierarchical laminar levels, respectively. Panels F and G further represent an overview of the entire iterative clustering process. Specifically, panel F depicts the process in terms of the iteration-specific c-layer distribution, while panel G represents the corresponding global hierarchical tree. Comparing panels F and G to the manual labelling result (shown between panel F and G), it can be seen that our hierarchical tree and especially levels 7 and 8 were in accordance with Brodmann's classification scheme. Yet, corresponding B-layers merged at different hierarchical levels of our tree. For instance, B-layers IV and Va merged at the second last level, whereas B-layers Vb and VI merged at the last level. In addition, the hierarchical tree also indicated the existence of sub-layers that eventually merged with different 7<sup>th</sup> or 8<sup>th</sup> level c-layers compared to B-layers. Figure S2 contains, for instance, a thin c-layer before the 7<sup>th</sup> level, which, according to manual labelling, should belong to B-layers II/III, but merged with more superficial c-layers at the 7<sup>th</sup> level. These examples illustrate the potential of the proposed clustering approach to allow deeper insight into the layer structure and its hierarchical organization when compared to the classical approach of manual labelling.

The influence of the feature sets on the clustering and the similarity of the resulting hierarchies are summarized in Fig. 5, using a VISp area data set. It can be seen that feature sets that shared features resulted in more similar trees (Fig. 5A,C). Further, regarding  $\kappa_m$ , the feature sets F1, F2, F3 and F4 yielded much more similar results compared to the manual labelling than feature sets F5 or F6 ( $p < 0.001$ , t-test with Bonferroni correction); see Fig. 5B,D. Compared to F1, the similarity to the manual labelling was, however, only slightly increased for F2, F3, or F4, with the differences not being significant ( $p > 0.2$ ); see Fig. 5D. This, in turn, illustrates that already neuron density alone appears to contain the main information that is able to guide the manual labelling (see also the last *Results* section).

**Supervised learning and layer transfer.** Similar to the unsupervised clustering results, assignment of d-layers was sensitive to the selection of the applied feature set. In addition, the d-layers depended on the selected reference data set and the reference c-layers, respectively. Yet, the example shown in Fig. 6 illustrates that, if appropriate, neuro-biologically meaningful feature sets were used, the assigned d-layers were similar to the manual labelling results. In the specific case, the VISp data set of Fig. 4 served as reference data set; the respective reference c-layers were to be transferred to a RSC location data set. The unsupervised learning results (using feature set F1) for the RSC data set are shown in panel B of Fig. 6, and the supervised learning results (using feature set F3; results shown for all 200 SVMs) in panel A. The final d-layers are presented in panel D and well agreed with the B-layers indicated to the right of the d-layers.

Different to unsupervised clustering, in the supervised learning part, the single feature *neuron density* was not sufficient to ensure high similarity between d-layers and manual labels. This finding was reflected in the evaluation of the quality measures  $n_B$ ,  $f_{c=B}$ ,  $f_{d=c}^{\text{ref}}$ , and  $f_{d=B}$ , with the results summarized in Fig. 7 (VISp reference data sets) and S3 (SSp reference data sets). In particular, panels A to H in Fig. 7 and panels A to G in Fig. S3 represent

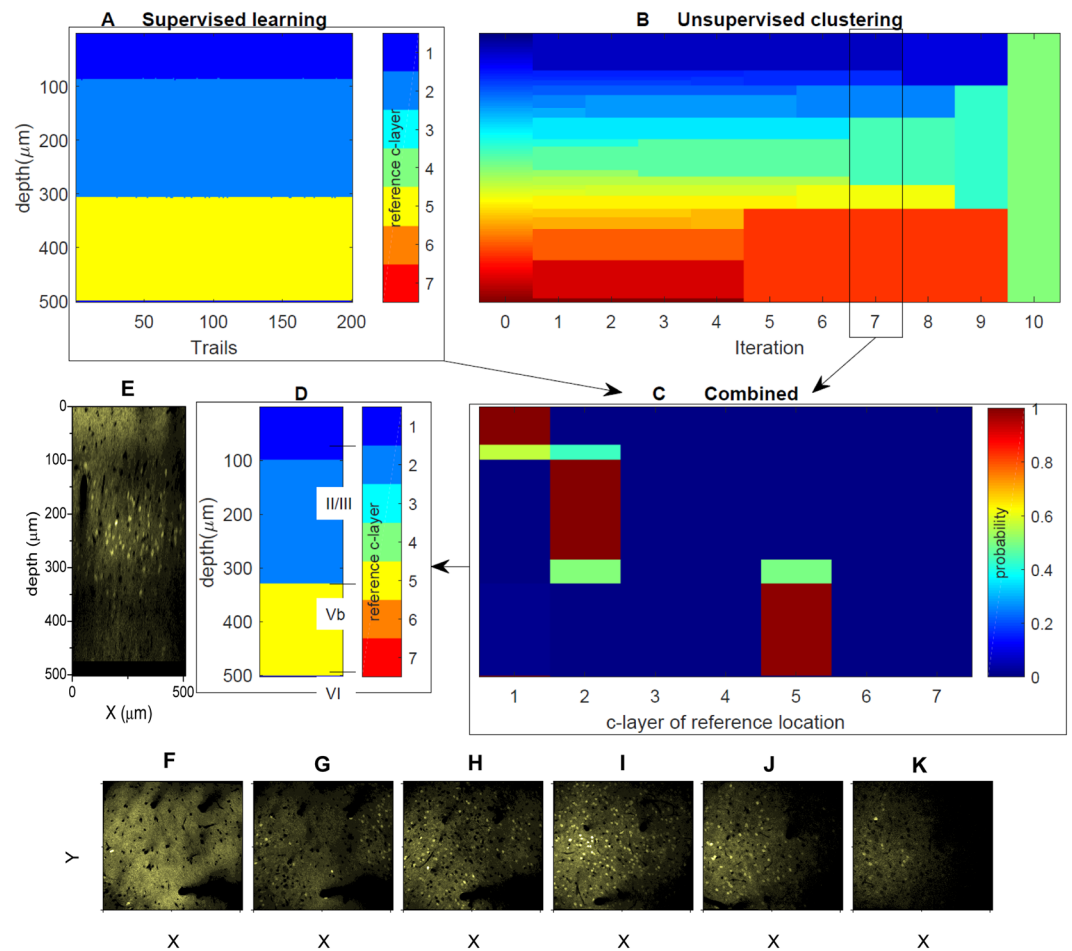


**Figure 5.** Evaluation of unsupervised clustering. **(A)** The similarity  $\kappa$  of the results obtained for any pair of feature sets and all 60 data sets studied, and **(B)** the similarity  $\kappa_m$  of the feature set-specific results and the manually labelled layers (indicated with ML). The bottom label of panel (B) shows the data set names and the top label of panel (A) their area types. **(C)** The first six columns and first six rows show the mean values of  $\kappa$  over all 60 locations, and the seventh column and the seven rows the mean value of  $\kappa_m$  over all 60 locations. The element at (ML, ML) is manually set to 1. **(D)** The detailed statistics of  $\text{Log}(\kappa_m)$ . The histograms show the mean values, error bars the standard deviations, and the blue dots the maximum values. The minimal values are close to zero or even negative, and are, therefore, not shown. Pairwise comparison of F1, F2, F3, and F4 values is not significant ( $p > 0.2$ ), whereas pairwise comparison of F1, F2, F3, or F4 to F5 or F6 is significant ( $p < 0.001$ ). To illustrate the meaning of the obtained  $\kappa_m$  values, we randomly shifted the manually assigned layer boundaries, with the boundary shift values chosen in the range of  $[-Nsa, Nsa]$  slices. We repeated runs with  $Nsa = 1, 2, \dots, 6$ . For each  $Nsa$  value, each of the 60 locations was re-labelled and the artificial labelling result compared to the original manual labelling. Corresponding mean values are plotted in panel (D). The olive line depicts the respective correlation coefficient when the layers manually labelled by the second expert reader were used as test result. It can be seen that a  $Nsa$  value of 5 approximately resembles interobserver variability of layer labelling; data points at and above this line can be considered to indicate ideal performance.

the feature set-specific performance in terms of the index  $f_{d=B}$ , i.e. the fraction of slices with agreement of the d-layers and the manually assigned B-layers, for the individual VISp and SSp reference data sets. It could be seen that F3 and F4 resulted in higher  $f_{d=B}$  values than F1 for all reference data sets, with the differences being significant ( $p < 0.01$ ) for, e.g., six of the eight VISp reference data sets. However, F4 did not result in a significantly better performance than F3 ( $p > 0.4$  in all eight cases), even though F4 included the mean size of all neurons as additional feature. Similar observations held also true for  $f_{d=C}^{ref}$ .

**Neuro-biological vs. image texture features.** Figures 5, 7 and S3 further reveal that the neuro-biologically inspired features and features sets F1 to F4 clearly outperformed the direct image-based texture features and feature sets F5 and F6 in terms of  $\kappa_m$  and  $f_{d=B}$ . Thus, every feature and feature set led to a hierarchical tree, but the results were not necessarily neuro-biologically meaningful and comparable to classical layer discrimination patterns.

This observation is further supported by Fig. 8, which represents cross-location similarity of d-layer-based discrimination results. Here, feature set F1 was used for clustering, F3 for supervised learning and an SSp area was chosen as reference data set. The resulting pattern revealed high similarity between image data from locations of the same area type, and thereby resembled the expected results and the corresponding pattern obtained by the manual labels (see Supplemental Fig. S4), although corresponding observations after unsupervised clustering and hierarchy comparison were not that obvious (Fig. S5). Comparison of Figs 8 and S6 further demonstrate that the results were also sensitive to the selection of the reference c-layers, but, more importantly, application of the direct image-based features did not lead to a (neuro-biologically) meaningful pattern at all (Fig. S7).

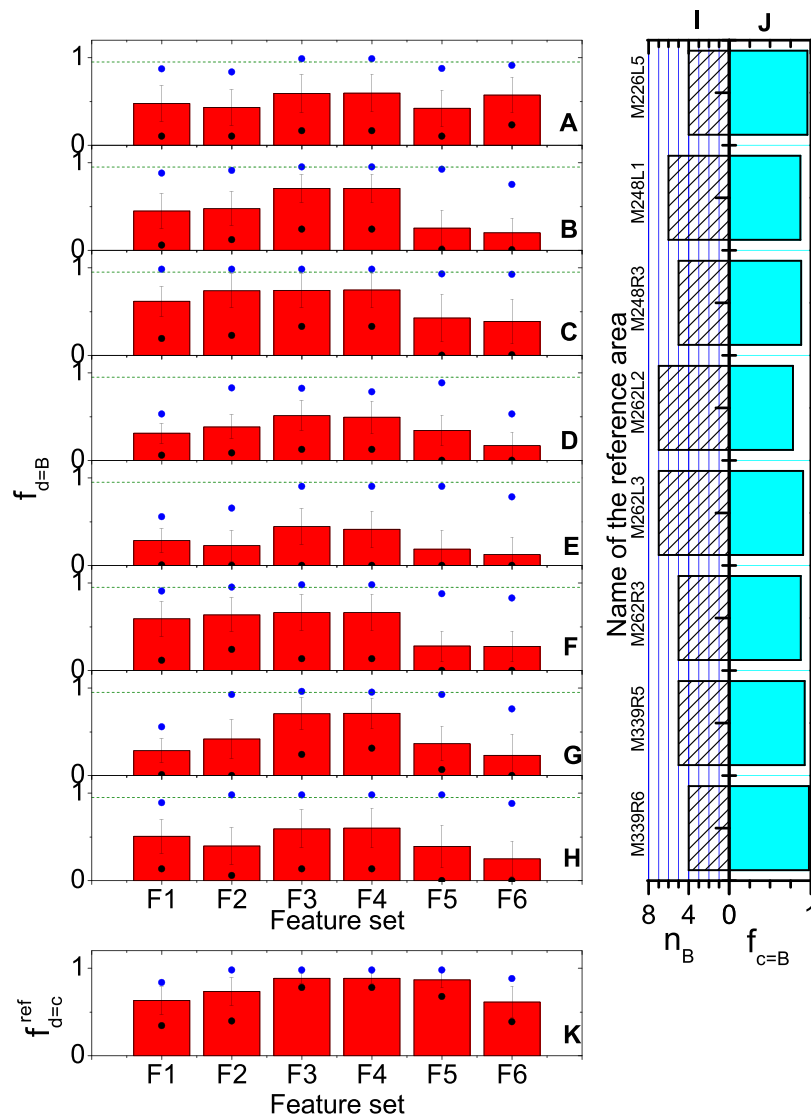


**Figure 6.** Example for transferring reference c-layers to a target data set. Panels (A,B) represent the results of the supervised and unsupervised learning of the target data set. Panels (C,D) show the probability of the assignment of the reference c-layers to the target c-layers as determined by an SVM ensemble and the final discrimination result. Unsupervised clustering was based on feature set F1, and the supervised learning on F3. The dataset under discrimination is the RSC area M262L0, and the reference data for supervised learning are the c-layers of the 8<sup>th</sup> hierarchical level of VIsP area M248L1, that is, the one shown in Fig. 4. For comparison purposes, the manually labelled layers are indicated to the right of panel (D). Panel (E) shows a corresponding representative x–z plane, and the bottom panels six x–y plane slice examples at  $z = 50 \mu\text{m}$  (F),  $z = 90 \mu\text{m}$  (G),  $z = 140 \mu\text{m}$  (H),  $z = 220 \mu\text{m}$  (I),  $z = 300 \mu\text{m}$  (J), and  $z = 400 \mu\text{m}$  (K), corresponding to the six c-layers on the 7<sup>th</sup> hierarchical level in panel B. Both panels (E,F) correspond to the first d-layer in panel (D), whereas panels (G–I) correspond to the second d-layer, and panel J corresponds to the third d-layer.

## Discussion

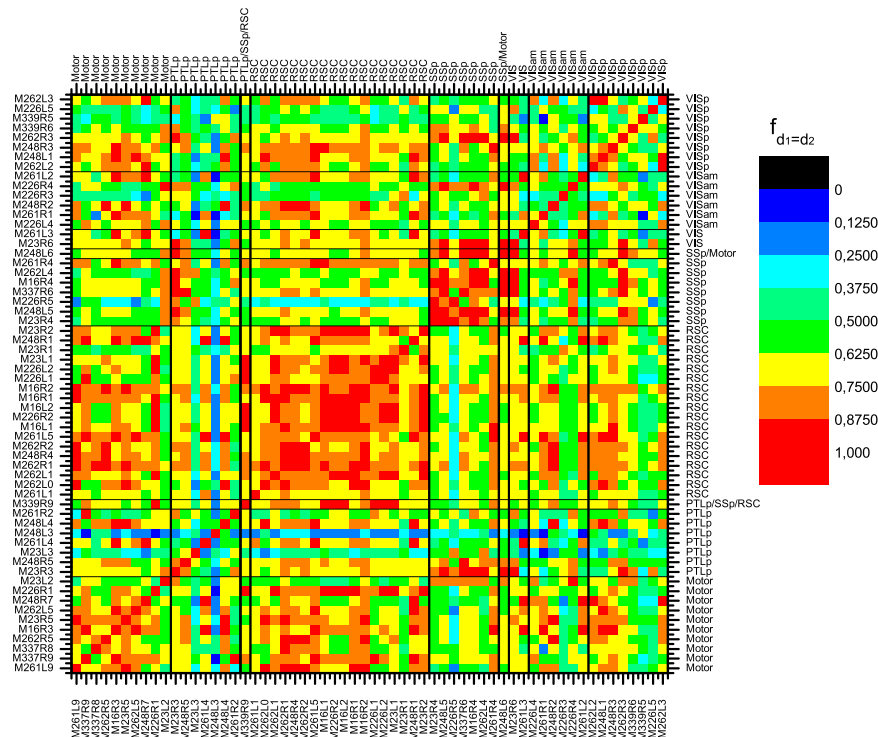
Machine learning methods that are able to discriminate cortical laminar patterns and their hierarchy are of high demand in the field of neuroscience<sup>38–41</sup>. In this study, we demonstrate that combining unsupervised clustering and supervised learning is well suited for such tasks. The main ideas are that iterated unsupervised clustering allows appropriately describing the laminar hierarchical structure for individual cortical locations, and supervised learning employing “rich structure” reference locations enables to bridge those hierarchical patterns between different locations. Last but not least, our results also illustrate that selecting neurobiologically meaningful feature sets is an important prerequisite for obtaining neurobiologically meaningful layer discrimination results. Subsequently, some more specific advantages as well as interesting aspects and potential shortcomings of our approach are discussed in more detail.

**No need for manual labelling.** One significant advantage of our layer discrimination procedure is that there is no need to manually identify and label the layers. Therefore, the method is less limited by subjective perception and related potential observer bias. Simultaneously, the procedure provides extended information on the hierarchy of laminar cortical organization and opens up more flexible ways to study cortical structure–function relations. Naturally, human interaction is still required to select the hierarchical level and reference locations and layers, depending on the particular research interest, but this is inevitable and not considered a shortcoming of the method itself.



**Figure 7.** Supervised learning evaluation results for the VISp reference data and c-layers: Although the exact numbers are partly sensitive to the quality of the training data, the neuro-biologically inspired feature sets F1–F4 eventually result in labels that more closely resemble the manual ones than those obtained using the image texture features F5 and F6. (A–H)  $f_{d=B}$  statistics for the different reference locations (panels) and different feature sets (horizontal axis). The histograms show the mean values over the 60 locations, the error bar the standard deviations, and blue and black dots the maximum and minimum values. Comparison of differences between F1 and F3 or F1 and F4 are, except for M262R3 and M339R6, significant ( $p < 0.01$ ). In contrast, differences between F3 and F4 are not significant for all cases ( $p > 0.4$ ). For comparison purposes, the olive lines represent respective statistics when the datasets manually labelled by the second reader are used as the test data. Panels (I,J) contain additional information about the reference c-layers. Panel (I) shows the number of manually discriminated and labelled layers of the reference locations, namely  $n_B$ . (J) shows the similarity between the refined reference c-layer structures and the refined manually labelled layers structures. (K) The statistics of the self-identical measurement (=re-labelling) of the reference locations, where all the symbols have the same statistical meaning as in panels (A–H). All reference c-layers were selected on the hierarchical level after the 8<sup>th</sup> clustering iteration. Target c-layers were selected on the last hierarchical level that exhibited a layer number not smaller than the number of reference c-layers.

**Slice- versus (individual) neuron-based layer discrimination.** In this work, we use vertical image slices as objects of interest to perform the discrimination task, which is different from alternative approaches that detect and then assign individual neurons to different layers<sup>40</sup>. In comparison, assigning neurons to layers usually requires quite a number of elaborated features, whereas in this work we eventually used five or fewer features (F1 has one feature; F3 has five features) to obtain promising results. In addition, the number of neurons is naturally larger than that of the image slices; in our case, the number of neurons in a single image data set is on the scale of ten thousand, whereas the data set comprises only a few hundred slices. Therefore, advantages of using slices include (depending on the complexity of considered neuron features), for example less computation time and simpler feature selection.



**Figure 8.** Similarity of discrimination results for any pair of datasets, using feature set F3 for supervised learning and M16R4 as reference data set. Reference c-layers were selected after the 8<sup>th</sup> clustering iteration, and all target c-layers were selected on the last hierarchical level with the number of layers not being smaller than the number of the reference c-layers. Similarity of discrimination results for areas of the same type as well as for areas commonly considered structurally similar (e.g. Motor and RSC, which do not have B-layer IV) was in general high.

In contrast, a shortcoming is that in the horizontal direction (i.e. within slices), the entire slices are assumed to be homogeneous regions that can be unambiguously assigned to a specific layer. Alternatively, smaller regions or patches of the slices with more homogeneous characteristics could be selected for clustering.

A related aspect refers to a more philosophical issue. As is well known, there exist transition zones between the cortical layers, which cannot unambiguously be assigned to specific layers. During neuron-based laminar discrimination, transition layers could in principle be defined as regions where upper and lower layer neurons are entangled together. Nonetheless, such compartments would usually be assigned to either the upper or the lower layer, based on some implementation-specific criteria<sup>40</sup>. In our approach, transition layers automatically show up as separate layers at more detailed hierarchical levels, but may merge with either the upper or the lower neighbouring layer at higher hierarchical levels. As an example, in Figs 4, 6 and S2, such layers can be easily identified.

**Impact of feature selection: What means “meaningful” layer discrimination?** It needs to be noted that, especially for the unsupervised clustering approach, it is not easy to evaluate the appropriateness of a feature set. Any (combination of) features will generate a hierarchical tree. In our case, we evaluate the appropriateness of features mainly by their ability to generate a final single-level layer discrimination that resembles manually labelled (Brodmann) layers. Human observers (and most probably also Brodmann himself one century ago)<sup>48,49</sup> discriminate cortical layers mainly visually based on neuron density and neuronal morphology. Thus, it was expected that corresponding features performed best in terms of respective quality measures. Yet, other features and feature sets do not necessarily have to be less meaningful, but can potentially reveal additional, currently neglected structuring principles. Finally, to judge appropriateness and importance of feature sets and discrimination results, it remains to be seen and investigated whether there exists, for example, a functional correspondence of the uncovered hierarchy and layer discrimination.

Still, based on similarity to manual labelling results according to Brodmann’s classical laminar classification scheme, we tentatively conclude that F1 (neuron density), F2 (neuron density and mean neuron size), F3 (neuron density, the proportions of the two neuron classes, and the mean neuron sizes of the two classes), and F4 (F3 plus the mean size of all neurons) give promising results, and, if computation cost is considered critical, neuron density alone already leads to reliable laminar hierarchies by the proposed unsupervised clustering approach. For the described supervised learning part, we, nevertheless, demonstrated that F3 and F4 significantly improve performance compared to neuron density alone. In contrast, the image texture features and F6 do not lead to satisfying results at all.

Taking into account that our applied methods of neuron size and shape detection and neuron classification are not very precise at the moment (as the development of such algorithms was not our main goal and beyond the scope of this work), we can reasonably expect that developing better methods will further improve the presented discrimination results.

**Future directions.** In the current work, we demonstrated the proposed combination of unsupervised and supervised learning to successfully discriminate laminar hierarchies in 2-photon microscopy image data of the mouse cortex. Being the first implementation and application of our approach, it naturally offers potential for refinement and extension both in terms of methodology and applications.

One aspect is again related to the performed slice-based layer discrimination and the underlying assumption of feature homogeneity throughout the slice and existence of layer boundaries only between the slices (if they are not considered transition zones). We consider this assumption to be appropriate with respect to the given image data mainly due to the relatively smooth mouse brain and cortex. Yet, to extend this approach to other species like primates, the potentially curved surface and layer boundaries within the analysed 3-dimensional image areas have to be taken into account and the hierarchical clustering has to be extended to respective 2- or 3-dimensional regions of interest.

In addition, since established evaluation methods like the Rand index<sup>50</sup> were, according to our understanding, not directly applicable in our specific context, due to, e.g., the hierarchical organization we were interested in, we developed and introduced novel and in our opinion intuitive evaluation methods and measures that could be easily adjusted to our subtasks (e.g. comparison of a hierarchical clustering results to a fixed-level discrimination results). We are looking forward to observing whether these measures will be taken up by other working groups and if a need to further adjust them will be identified by the community.

Despite respective hierarchical organization of laminar patterns, much work still focusses on a specific hierarchical level and many people on layer-specific functions. Our proposed approach could also support related research, but additional efforts will be necessary. In particular, the targeted hierarchical level has to be automatically detected for the different locations addressed in the individual studies. In the current work, the 8<sup>th</sup> level of the clustering hierarchy was chosen for all samples to demonstrate feasibility of our algorithm because of its high similarity with the well-known and wide spread Brodmann layer discrimination scheme. Yet, this does not necessarily mean that for individual areas there potentially existed no slightly deviating hierarchy levels that more closely resembled the Brodmann scheme. To auto-detect them, we would have had to define more specifically the properties and features of the Brodmann areas, which means, our expectation regarding the desired hierarchy level, which was beyond the scope of the presented work. This, in turn, leads to the interesting question which combination of the approached (neuro-biologically inspired) features would eventually best suited for this specific task. And similarly to our application, one has to define unique properties of the desired hierarchical level in sufficient and testable detail for other target applications and research questions before being able to adapt our approach to it.

Continuing with additional future directions in terms of applications, the proposed approach could also be extended to discriminate the considered cortical areas themselves. Our data demonstrated that the obtained discrimination results partly reflected properties of areas types (see again Fig. 8). In turn, cortex parcellation could also be based on the laminar discrimination results. Yet, comprehensive parcellation requires datasets that cover the entire area of interest, and in the given context the whole animal brains, which were not available for the current study.

Finally, we would like to note that considered meaningful features in the current study referred to cytoarchitecture architecture. In turn, one can expect to automatically identify a myelin-based laminar structure if using myeloarchitecture-related features for layer discrimination. The main research focus of our group will, however, be to study functional laminar structure by using functional data.

## Conclusions

Combining unsupervised and supervised machine learning allows discriminating (and finally analysing) the hierarchical laminar cortical organization in 2-photon image data without the need of manual labelling. Unsupervised clustering offers the opportunity to identify the hierarchical laminar organization of image data obtained at the different cortical locations. Regarding respective features for layer discrimination, it turned out that neuron density alone was already sufficient to achieve biologically plausible results during unsupervised clustering, whereas supervised learning for cross-location label transferring requires consideration of additional features. The computed discrimination results agree well with the classical Brodmann scheme of cortical layers and reflect the layer similarity of data acquired at locations of the same type of cortical area. At the same time, they also provide additional, new information on the laminar structure of cortical areas, for instance, by providing insight into the hierarchical structure revealed by the clustering analysis.

## Data Availability

Codes and example data for test are available via <https://github.com/IPMI-ICNS-UKE/ICNS-CoLD>.

## References

- Ahissar, E., Sosnik, R., Bagdasarian, K. & Haidarliu, S. Temporal frequency of whisker movement. II. Laminar organization of cortical representations. *Journal of Neurophysiology* **86**, 354–367 (2001).
- Brodmann, K. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. (Barth, 1909).
- Kandel, E. R., Schwartz, J. H. & Jessel, T. M. *Principles of neural science (Fourth ed.)*. (United State of America: McGraw-Hill, 2000).
- Schilling, K. G. *et al.* The VALIDATE29 MRI Based Multi-Channel Atlas of the Squirrel Monkey Brain. *Neuroinformatics* **15**, 321–331 (2017).
- von Economo, C. F. & Koskinas, G. N. *Die cytoarchitektonik der hirnrinde des erwachsenen menschen*. (J. Springer, 1925).
- Tömböl, T. In *Cellular components of the cerebral cortex* (eds Peters, A. & Jones, E. G.) 479–519 (Plenum Press: New York, 1984).
- Wise, S. P. & Jones, E. G. Cells of origin and terminal distribution of descending projections of the rat somatic sensory cortex. *Journal of Comparative Neurology* **175**, 129–157 (1977).
- Hilgetag, C. C., O'Neill, M. A. & Young, M. P. Indeterminate organization of the visual system. *Science* **271**, 776–776, <https://doi.org/10.1126/science.271.5250.776> (1996).
- Hilgetag, C. C., O'Neill, M. A. & Young, M. P. Hierarchical organization of macaque and cat cortical sensory systems explored with a novel network processor. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **355**, 71–89 (2000).
- Zhou, C., Zemanová, L., Zamora, G., Hilgetag, C. C. & Kurths, J. Hierarchical Organization Unveiled by Functional Connectivity in Complex Brain Networks. *Physical Review Letters* **97**, 238103, <https://doi.org/10.1103/PhysRevLett.97.238103> (2006).

11. Murray, J. D. *et al.* A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience* **17**, 1661, <https://doi.org/10.1038/nrn.3862> (2014).
12. de Kock, C. P. J., Bruno, R. M., Spors, H. & Sakmann, B. Layer- and cell-type-specific suprathreshold stimulus representation in rat primary somatosensory cortex. *The Journal of Physiology* **581**, 139–154 (2007).
13. Xie, H. *et al.* *In vivo* imaging of immediate early gene expression reveals layer-specific memory traces in the mammalian brain. *Proceedings of the National Academy of Sciences* **111**, 2788–2793, <https://doi.org/10.1073/pnas.1316808111> (2014).
14. Barth, A. L. Visualizing circuits and systems using transgenic reporters of neural activity. *Current Opinion in Neurobiology* **17**, 567–571 (2007).
15. Barth, A. L., Gerkin, R. C. & Dean, K. L. Alteration of neuronal firing properties after *in vivo* experience in a FosGFP transgenic mouse. *Journal of Neuroscience* **24**, 6466–6475 (2004).
16. Birkner, A., Tischbirek, C. H. & Konnerth, A. Improved deep two-photon calcium imaging *in vivo*. *Cell calcium* **64**, 29–35 (2017).
17. O'Donovan, M. J., Ho, S., Sholomenko, G. & Yee, W. Real-time imaging of neurons retrogradely and anterogradely labelled with calcium-sensitive dyes. *J Neurosci Methods* **46**, 91–106 (1993).
18. Ricard, C. *et al.* Two-photon probes for *in vivo* multicolor microscopy of the structure and signals of brain cells. *Brain Structure and Function*, 1–33 (2018).
19. Stosiek, C., Garaschuk, O., Holthoff, K. & Konnerth, A. *In vivo* two-photon calcium imaging of neuronal networks. *Proceedings of the National Academy of Sciences* **100**, 7319–7324 (2003).
20. Svoboda, K., Denk, W., Kleinfeld, D. & Tank, D. W. *In vivo* dendritic calcium dynamics in neocortical pyramidal neurons. *Nature* **385**, 161, <https://doi.org/10.1038/385161a0> (1997).
21. Thunemann, M. *et al.* Deep 2-photon imaging and artifact-free optogenetics through transparent graphene microelectrode arrays. *Nature Communications* **9**, 2035, <https://doi.org/10.1038/s41467-018-04457-5> (2018).
22. Wachowiak, M. & Cohen, L. B. Representation of odorants by receptor neuron input to the mouse olfactory bulb. *Neuron* **32**, 723–735 (2001).
23. Wang, K. H. *et al.* *In vivo* two-photon imaging reveals a role of arc in enhancing orientation specificity in visual cortex. *Cell* **126**, 389–402 (2006).
24. Lerch, J. P. *et al.* Cortical thickness measured from MRI in the YAC128 mouse model of Huntington's disease. *Neuroimage* **41**, 243–251 (2008).
25. Ahissar, E. & Staiger, J. S1 laminar specialization *Scholarpedia* **5**, 7457 (2010).
26. Frick, A., Feldmeyer, D., Helmstaedter, M. & Sakmann, B. Monosynaptic connections between pairs of L5A pyramidal neurons in columns of juvenile rat somatosensory cortex. *Cerebral cortex* **18**, 397–406 (2007).
27. Lefort, S., Tómm, C., Sarria, J.-C. F. & Petersen, C. C. H. The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron* **61**, 301–316 (2009).
28. Manns, I. D., Sakmann, B. & Brecht, M. Sub- and suprathreshold receptive field properties of pyramidal neurones in layers 5A and 5B of rat somatosensory barrel cortex. *The Journal of physiology* **556**, 601–622 (2004).
29. Molnár, Z. & Cheung, A. F. P. Towards the classification of subpopulations of layer V pyramidal projection neurons. *Neuroscience Research* **55**, 105–115 (2006).
30. Schubert, D., Kötter, R., Luhmann, H. J. & Staiger, J. F. Morphology, electrophysiology and functional input connectivity of pyramidal neurons characterizes a genuine layer Va in the primary somatosensory cortex. *Cerebral cortex* **16**, 223–236 (2005).
31. Oberlaender, M. *et al.* Cell type-specific three-dimensional structure of thalamocortical circuits in a column of rat vibrissal cortex. *Cerebral cortex* **22**, 2375–2391 (2011).
32. Opitz, A. *et al.* P098 Cortical layer-specific distribution of TES electric fields. *Clinical Neurophysiology* **128**, e59–e60 (2017).
33. Tang, A. *et al.* P135 Intermittent theta burst increases action potential threshold and spike firing in cortical layer 5 pyramidal neurons. *Clinical Neurophysiology* **128**, e82 (2017).
34. Saikali, S. *et al.* A three-dimensional digital segmented and deformable brain atlas of the domestic pig. *Journal of neuroscience methods* **192**, 102–109 (2010).
35. Maas, L. C. *et al.* Early laminar organization of the human cerebrum demonstrated with diffusion tensor imaging in extremely premature infants. *Neuroimage* **22**, 1134–1140 (2004).
36. Chen, G., Wang, F., Gore, J. C. & Roe, A. W. Identification of cortical lamination in awake monkeys by high resolution magnetic resonance imaging. *Neuroimage* **59**, 3441–3449 (2012).
37. Mahmud, M. *et al.* An automated method for detection of layer activation order in information processing pathway of rat barrel cortex under mechanical whisker stimulation. *Journal of neuroscience methods* **196**, 141–150 (2011).
38. Stajduhar, A., Judas, M. & Loncaric, S. *Automatic segmentation of cortical layers in histological images of human brain* (Poster No. H014, 10<sup>th</sup> July, 11<sup>th</sup> FENS Forum, Berlin, Germany, 2018).
39. Nosova, S., Snopova, L. & Turlapov, V. Automatic detection of neurons, astrocytes, and layers for NISSL-stained mouse cortex. *Journal of WSCG* **25**, 143–150 (2017).
40. Somasundar, V., Padmanabhan, R., Roysam, R., Barnes, C. A. & Lister, J. P. *Semi-automated layer classification tool for defining cortical architecture* (Poster No. 712.15, Society for Neuroscience, Washington, DC, USA, 2017).
41. Wagstyl, K. *et al.* Mapping Cortical Laminar Structure in the 3D BigBrain. *Cerebral Cortex* **28**, 2551–2562 (2018).
42. Bastiani, M. *et al.* Automatic segmentation of human cortical layer-complexes and architectural areas using *ex vivo* diffusion MRI and its validation. *Frontiers in neuroscience* **10**, 487 (2016).
43. Chang, C.-C. & Lin, C.-J. LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* **2**, 27 (2011).
44. Kim, T. H. *et al.* Long-term optical access to an estimated one million neurons in the live mouse cortex. *Cell reports* **17**, 3385–3394 (2016).
45. Guerrero-Pena, F. A. *et al.* Multiclass Weighted Loss for Instance Segmentation of Cluttered Cells. *arXiv preprint arXiv:1802.07465* (2018).
46. [www.kaggle.com/c/data-science-bowl-2018](https://www.kaggle.com/c/data-science-bowl-2018).
47. García-Cabezas, M. Á., John, Y. J., Barbas, H. & Zikopoulos, B. Distinction of neurons, glia and endothelial cells in the cerebral cortex: an algorithm based on cytological features. *Frontiers in neuroanatomy* **10**, 107 (2016).
48. Zilles, K. & Amunts, K. Centenary of Brodmann's map—conception and fate. *Nature Reviews Neuroscience* **11**, nrn2776 (2010).
49. Amunts, K. & Zilles, K. Architectonic mapping of the human brain beyond Brodmann. *Neuron* **88**, 1086–1107 (2015).
50. Hubert, L. & Arabie, P. Comparing partitions. *Journal of classification* **2**, 193–218 (1985).
51. Standing, S. *Gray's anatomy e-book: the anatomical basis of clinical practice*. (Elsevier Health Sciences, 2015).

## Acknowledgements

This work is funded by German Research Foundation (DFG) and the National Natural Science Foundation of China in the project Cross-modal Learning, DFG TRR-169/NSFC (61621136008) - A2 to C.C.H. and J.S.G., DFG SPP2041 as well as HBP/SP2 SGA2, DFG SFB -936 - A1, Z3 to CCH, NSFC (31671104) to J.S.G. and DFG SFB -1328 - A2 to R.W. The authors thank Changsong Zhou for helpful discussions and Farid I. Kandil for his suggestions on the denotation of the different types of layers.

### Author Contributions

J.-S.G. and C.C.H. designed the research; D.L., M.Z. and R.W. developed the discrimination methodology; G.W., H.X. and Y.H. worked on the experiments and pre-processed the image data; D.L., R.W., J.-S.G. and C.C.H. wrote the paper; all authors approved the paper.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-43432-y>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019