Computational Analysis of Laminar Structure of the Human Cortex Based on Local Neuron Features

Andrija Štajduhar
Croatian Institute for Brain Research
University of Zagreb, School of Medicine

Tomislav Lipić
Laboratory for Machine Learning and Knowledge Representation
Ruder Bošković Institute

Goran Sedmak
Croatian Institute for Brain Research
University of Zagreb, School of Medicine

Sven Lončarić
Faculty of Electrical Engineering and Computing
University of Zagreb

Miloš Judaš
Croatian Institute for Brain Research
University of Zagreb, School of Medicine

May 6, 2019

Abstract

In this paper, we present a novel method for analysis and segmentation of laminar structure of the cortex based on tissue characteristics whose change across the gray matter facilitates distinction between cortical layers. We develop and analyze features of individual neurons to investigate changes in architectonic differentiation and present a novel high-performance, automated tree-ensemble method trained on data manually labeled by three human investigators. From the location and basic measures of neurons, more complex features are developed and used in machine learning models for automatic segmentation of cortical layers. Tree ensembles are used on data manually labeled by three human experts. The most accurate classification results were obtained by training three models separately and creating another ensemble by combining probability outputs for final neuron layer classification. Measurement of importances of developed neuron features on both global model level and individual prediction level are obtained.

Keywords Cerebral cortex · cytoarchitecture · machine learning · computational histology

1 Introduction

Interest in analysis of laminar structure is driven by evidence of relationship between features of cytoarchitectonics and cortical functions. Computer-aided methods provide means for fast and objective investigation of the cortical structure and enable researchers to gain better understanding of anatomical and functional organization of the brain, as well as studying many neurological and psychiatric diseases that cause subtle changes in the brain structure. Today, it is thought that variations in neuronal distribution in the brain and their density determine function, with many evidences linking architectonic variation and brain function. These variations define several distinct layers of the cortex that are arranged parallel to the brain’s folded surface. The structure of these layers, known as the laminar structure, also varies, as well as distribution and size of neurons found within the layers. Based on these variations, we distinguish a series of spatially discrete areas of the brain. The subtleties in this fine structure of the brain underlying it’s function can be characterized in great detail by studying the organization of cells across the cortex, which is usually known as cytoarchitecture or

*Corresponding author - University of Zagreb, School of Medicine, Croatian Institute for Brain Research, Šalata 12, Zagreb 10000, Croatia, andrija.stajduhar@hiim.hr
cytoarchitectonics. Most investigations in this field today are mostly done manually which requires significant amount of researchers’ time, introduces observer-dependent bias and reduces the reproducibility of the research. More objective, data-driven methods are required that are also fast, precise and consistent.

1.1 Related work

Since the first methods for automated or semi-automated analysis of cortical layers were developed, central idea in almost all methods was the use of sampling along transverse lines drawn either manually or semi-automatically across the cortex, perpendicular to the laminar structure and spanning the full width of the cortex. At first, changes of optical density, or gray-level index (GLI) was measured by these profiles, which is a crude estimate of changes in neuronal density.

Recent methods for automated laminar analysis developed after the year 2000 have moved on from only using GLI in analysis of cortical profiles. In year 2002, authors develop statistical features that characterize each GLI profile and its discreet derivative, such as the first four moments about the mean, treating the profile as a frequency distribution, for the total of 10 features for each profile. Authors also bring these features in relation to cytoarchitectonics and discuss different distance measures used to compare profile vectors derived from different brain areas. An important step in development of profile features was made eight years later by the introduction of automatic methods for estimation of cell counts, thus providing realistic information about neuron density. Authors provide the number and distribution of neurons within several projection columns in rat cortex, reporting an estimate of neuron number per layers. This data is further used to derive the action potential output of projection columns specific for each layer, clearly showing direct link to analysis of the brain function through analysis of its cytoarchitecture.

The first article that uses features and statistics of individual neurons appeared in 2017. In 2018, authors use automatic segmentation of cells in 7µm thick Nissl-stained sections of the mouse brain and develop 11 shape descriptor features for each cell. After manually setting the upper and lower bounds for thresholding, a binary image is produced, and blobs are classified as glial cells or neurons, which are further subdivided into pyramidal and non-pyramidal. Using profiles across the mouse cortex for each cell type, authors distinguish five cortical layers, merging layers II and III into a single layer. However, only a few hundreds of cells are found on a small patch of tissue labeled by one expert. The method nevertheless offers a novel approach by using shape descriptors of individual cells for cell classification. A year after, use of GLI profiling on a large in 3D dataset was presented. Authors develop automated analysis of the laminar structure of the BigBrain by analysis of GLI peaks to distinguish between the layers. A relation between cytoarchitectonics an in vivo MRI imaging is discussed.

Most recently, in the year 2018, a first approach that does not use profiles across the cortex was proposed. A combined approach of unsupervised and supervised machine learning was used on an extensive dataset of 2-photon microscopic images of the rat cortex. Neurons are automatically segmented and some neurobiologically sound statistics developed and combined with texture descriptors to produce several sets of features used in learning methods. Importance of neuron density is especially emphasized, and it is pointed out that it alone leads to revelation of laminar hierarchies by the proposed unsupervised method. Authors also address the issue of human bias in manual segmentation of cortical layers and use an unsupervised clustering approach to identify and represent the laminar structure. Supervised learning is used to transfer resulting layer segmentation on different brain regions.

Considerations from this review of historical and recent methods leads to following conclusions: (i) Over the years, the impact of human bias in brain parcellation has been increasingly recognized and many methods sought to overcome this issue by developing objective quantitative measures and usage of statistics to distinguish between different layers and brain areas. (ii) Recent advancements in technology yield massive amounts of high-resolution and multimodal data which are far beyond the capabilities of any kind of manual analysis. (iii) Detailed anatomical and biologically meaningful information can be derived on a large-scale. This allows for the introduction of machine learning methods to the field that was until now dominated by the usage of exclusively image processing techniques. (iv) Neuron density seems to be the most important tissue descriptor, it alone facilitating the revealing of laminar structure, followed by neuronal size and shape. Derivation of higher-level descriptors, which may seem more or less neuroanatomically meaningful, can reveal currently neglected additional structuring principles and provide deeper understanding of laminar structure, leading to more reliable methods for layer segmentation. On the other hand, image texture features seem to have less discriminative power.

All this suggests that there is a trend to infer the laminar organization by studying its organization in ever more detail, going beyond limitations of the frameworks of manually created parcellation in conventional atlases. The research in the field should move away from human interpretation towards data-driven, unsupervised analysis with as little parameters as possible. However, the very recent methods are usually developed on rat or mouse cortex data, and
the results for human cortex are lacking. To present a better insight in varying anatomy of the cortex as revealed by histological imaging, outline of some of its important properties follows.

2 Analysis of laminar organization

According to classical works of cytoarchitectonics, neurons are usually distributed in six horizontally superimposed layers (layers I-VI) that are distinguished in most areas of the cortex [6], [39], [16]. Some neurons may be also found in the white matter (WM). The layers are result of variations in cell density, size and shape of neurons which are specific for each cortical layer. These, and potential variation in other features specific to either each individual neuron or its surroundings are our main interest in this paper.

In manual layer delineation, density and size of neurons are the most important characterizations of the laminar structure. Adjacent layers have different densities, with the exception at the border between layers V and VI, which have similar density, although a minor difference still exists.

In this paper, location and area of neurons are considered as fundamental properties of each neuron. Many other characteristics can be developed once when the location and border of a neuron are known, especially with regard to other neurons in surrounding tissue, for instance using distances to obtain neuronal density or measuring average size of neighboring neurons. Research in this paper was conducted using histological sections of the adult human prefrontal cortex, stained with NeuN immunohistochemistry method, from the Zagreb Brain Collection [15]. Preparations were digitized using Hamamatsu Nanozoomer 2.0 scanner at 40x magnification, which corresponds to \(0.226 \mu m/\text{pixel}\) resolution. In the experiments, 10\(\mu m\) and 20\(\mu m\) thick sections were used. Automatic neuron detection and segmentation methods used on histological images provide the capacity to obtain locations of all neurons in the tissue section, giving information about both locations and area of neurons [38], [40].

![Figure 1: Left: An example of digitized histological preparation of human orbitofrontal cortex (gyrus rectus). Right: Locations of individual neurons which are result of automatic neuron detection are plotted over the original image. Variations in neuron densities follow the laminar structure.](image)

Development of other neuron features based on neuron location and density is very natural. For instance, density as a number of neurons within some distance of a neuron, or by average distance to its \(k\) nearest neighbors. It is important to mention a distinction between two types of features: (i) those of each individual neuron, whether its own like location and area or derived as a statistic of features of neurons in some neighborhood, and (ii) image features obtained by conducting measurements in some chosen patch of an image, which more resembles to texture properties.
Values of the three features were measured for each of 229733 neurons found in a histological section. By plotting a histogram of their values one can observe that these features express a multimodal distribution. To separate the distributions, minimization of intraclass variance \([29]\) was used. Based on anatomical descriptions and visualized kernel density estimate shown in Fig. 2, three populations of similar densities are assumed (layers II and IV, layers III, V and VI, and Layer I and white matter), two populations of similar sizes and average neighbor sizes (layers III, V and VI, and layers I, II, IV and white matter). Neuron feature kernel density estimations with thresholds for minimization of intraclass variance are shown in Fig. 2 showing visualization of separating the neurons by the thresholds across the cortical section.

![Image](image.png)

Figure 2: Cell features exhibit multimodal distributions. Thresholds that separate the distributions were obtained using minimization of intraclass variance \([29]\). Left: Three types of neurons were distinguished by cell density in their surrounding area: very sparse (blue), sparse (red) and dense (green). Middle: Larger (red) and smaller (green) neurons. Right: Average size of neighboring neurons is also a feature that contributes to layer segmentation.

From the neuron segmentation, many other statistics can be developed using available software. For instance, ImageJ’s Particle analysis toolkit \([32]\) provides a set of features for each individual neuron, such as statistics of pixel intensity, shape descriptors and even orientation based on parameters of the fitted ellipse.

### 3 Exploratory data analysis

Considerations in the previous section further confirm the importance of the spatial distribution of neurons in discriminating between cortical layers. Taking into account the conclusions about the importance of unsupervised approach to the analysis of laminar structure, clustering-based methods provide an important insight in spatial distribution of neurons. These algorithms often produce results based on some input parameters, whether they are manually set or chosen automatically. The goal of this section is to explore intrinsic structure of neuron distribution through analysis and application of different clustering algorithms, development of various neuron characterizations and analysis of their capacity to discriminate between the neurons of different layers. A portion of the section has been manually labeled and a dataset of over 17000 neurons from all layers including white matter has been obtained. This dataset is used...
throughout the section to visualize distribution of neurons in different layers and track their changes with regard to application of different methods and statistics.

3.1 Density-based clustering

Density-based cluster analysis takes into account the spatial distance of the objects, thus defining clusters of higher or lower densities. Although there are many algorithms to choose from, not all are suitable for separation of neurons to appropriate layers, or even identification of potential sub-layers. Although the approach in this thesis strives to be data-driven with as little assumptions or parameters as possible, some aspects of the data need to be taken into account when choosing the appropriate clustering algorithm.

Algorithms like K-Means [21], K-Medioids [17], affinity propagation [10] and other similar algorithms assume globular shape of the data and cannot identify elongated clusters [13]. Such algorithms often require a predefined number of clusters or other parameters. Another undesirable property is non-determinism - different initialization setup may yield different results. Agglomerative clustering is a concept in which each object starts as its own cluster, and clusters are merged by some similarity criterion repeatedly, until there is only one cluster remaining. This procedure produces a hierarchy, a binary tree, or a dendrogram in which the clustering can be tracked down to the single leaf. There are different strategies on how to merge and split the clusters. The subtrees can be ranked by distance to when clusters merged, or by using more complex criteria like mean distance between the clusters or some other measure. Based on this, the tree is cut at some level and subtrees, or clusters, are determined based on that cut. An important property of this approach is that the clusters can now take any shape, not only globular. However, it may sometimes not be straightforward to determine the level of cut for selecting the clusters. Concept of density-based spatial clustering in applications with noise (DBSCAN) [9] is similar to agglomerative clustering - it applies single linkage clustering to the results in dendrogram, which distinguishes clusters according to a distance parameter. Again, in practice this parameter may be difficult to choose. The method uses another parameter that determines the minimum density around a point and classifies all the points below the minimum density parameter as noise. Consequently, some points are not being assigned to any cluster. With significant improvement in performance that can be achieved using kd-trees for local region queries, the algorithm can tackle large datasets which can be impractical for algorithms other than simple ones, like for instance K-Means. This algorithm can also find non-linearly separable clusters like cortical layers but does not handle small variations in density very well. It often produces smaller clusters enclosed in larger ones, if the points have slightly lower or greater density than the rest of the cluster, which is often the case within cortical layers. Very sparse clusters are also often split into several smaller clusters. Nevertheless, this is the most suitable approach for neuron clustering so far and gives insight about specifics of clustering applied for that purpose, leading to some practical conclusions.

In the context of neuron distributions in the cortex, several important discriminative criteria of the clustering algorithms need to be considered. The algorithm should have low number or even no parameters that need to be tuned and selected. It should also have the ability to handle clusters of arbitrary shape since the cortex is gyrified and layers follow winding patterns, allow clusters with slight variations in density due to the fact that neuron density is not absolutely uniform within each layer and finally, to be able to process large amounts of coordinates, since there may be hundreds of thousands or even millions of neurons found in the tissue section. A few algorithms found in the literature satisfy these criteria to different extent.

3.1.1 OPTICS

Ordering Points to Identify the Clustering Structure, or OPTICS [2], is a cluster analysis algorithm which does not produce explicit clusters, but instead creates an ordering of points representing their density-based clustering structure. The cluster-ordering obtained this way contains information equivalent to the density-based clustering corresponding to a broad range of parameter settings. It addresses one of DBSCAN’s major weaknesses, the problem of detecting meaningful clusters in data of varying density. The coordinates are (linearly) ordered such that points which are spatially closest become neighbors in the ordering. Additionally, a special distance is stored for each point that represents the density that needs to be accepted for a cluster in order to have both points belong to the same cluster. Similar to DBSCAN, the algorithm requires specifying two parameters: epsilon $\epsilon$, which describes the maximum distance (radius) around the point to consider, and minimum number of points required to form a cluster, $MinPts$. A point $i$ is considered as core point if at least $MinPts$ points are found within its $\epsilon$-neighborhood.

The algorithm creates a two-dimensional reachability plot, which is used to obtain the hierarchical structure of the clusters. The ordering of the points as processed by OPTICS is represented on the x-axis and the reachability distance on the y-axis. Since points belonging to a cluster have a low reachability distance to their nearest neighbor, the clusters show up as valleys in the reachability plot. Deeper valleys in the plot are result of denser clusters.
By changing parameters of the algorithm, there are two similar types of result obtained: (i) only neurons from layers II and IV form clusters, with numbers of smaller clusters found within each of the two layers depending on $\epsilon$ values and all other layers are considered as noise, and (ii) layers II and IV form single cluster each, and many small-sized clusters are formed in other regions, surrounded mostly by noise. Such result is obtained if $MinPts$ is increased for the same $\epsilon$ value in case (i). To obtain meaningful clusters, kernels between $80\mu m$ and $200\mu m$ in radius can be chosen, with minimum number of points between 20 and over 150. Reducing the number of neurons identified as noise may be achieved by choosing maximum distance between the points as $\epsilon$. However, this leads to increased computational complexity since every neighborhood query would return the full data set.

![Figure 3: Left: $\epsilon = 200$, $MinPts = 30$ Center: $\epsilon = 200$, $MinPts = 45$ Center: $\epsilon = 200$, $MinPts = 55$](image)

### 3.1.2 HDBSCAN

An algorithm developed as an improvement of both the DBSCAN and OPTICS algorithms with the goal to allow varying density clusters is Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection (HDBSCAN) [7]. The algorithm starts by transforming the space according to density, exactly as DBSCAN does, and performs single linkage clustering on the transformed space. Instead of taking $\epsilon$ value as a cutoff level for the dendrogram, the dendrogram is condensed by viewing splits that result in a small number of points splitting off. This results in a smaller tree with fewer clusters and can then be used to select the most stable or persistent clusters. This process allows the tree to be cut at varying height, selecting the varying density clusters based on cluster stability. The immediate advantage of this are varying density clusters; the second benefit is elimination of the epsilon parameter as it is no longer needed for choosing a cut of the dendrogram. Instead, a new parameter is introduced. It determines minimum cluster size which is used to determine whether points are to be considered as a noise or splitting to form two new clusters. This parameter is much more interpretable and determines the minimum size of clusters to be considered. Nevertheless, by varying the minimum cluster size, the algorithm still cannot clearly separate neurons within the layer borders. Neurons of Layers II and IV usually grouped in two distinct clusters which, however, often protrude into other layers. Also, layers other than II and IV are identified as noise, with smaller clusters sometimes forming in layer VI, confirming the suggestions of neuroanatomy about layer VI having slightly denser distribution of neurons than layers III and V. Visually meaningful clustering was obtained using minimum cluster size in range between 50 and 300 neurons.

### 3.1.3 FSDP

The basic assumption in clustering by fast search and find of density peaks (FSDP) [31], is that the cluster centers are surrounded by other data points which have lower local density. These centers are also assumed to be at relatively large distance from other similar points with large local densities. Therefore, two quantities are computed for each data point $i$, local density $\rho_i$,

$$\rho_i = \sum_j \chi(d_{ij} - d_c),$$

where $\chi(x) = 0$ if $x < 0$ and $\chi(x) = 1$ otherwise, with $d_c$ being a cutoff distance, and $\delta_i$

$$\delta_i = \min_{j : \rho_j > \rho_i} (d_{ij}).$$

In other words, $\rho_i$ is the number of points whose distance to the point $i$ is less than $d_c$, and $\delta_i$ is the minimum distance between a point $i$ and the point which has higher local density. The points with high $\rho_i$ and $\delta_i$ are then considered as cluster centers and other points are assigned to the same cluster as its nearest neighbor of higher density. Each point
Computational analysis of laminar structure of the human cortex based on local neuron features

Figure 4: HDBSCAN clustering with minimum cluster size of 25, 100 and 250 neurons. Low minimum cluster size yields dispersed and granular clusters. Increasing this value results in clusters being consolidated, with dense clusters expanding into adjacent cortical layers and vanishing smaller clusters. Light blue represents neurons identified as noise.

i is shown on a decision graph by using \((\rho_i, \delta_i)\) as its xy-coordinate, thus obtaining a decision graph. The density peaks can then be identified as outliers in the top right region. In our dataset, points with largest \(\rho\) values emerged, as expected, in layers II and IV. For smaller \(d_c\) distances, density peaks in other layers started to appear. The final clustering, however, cannot identify regions of similar density. Depending on \(d_c\), cluster centers that emerge in areas of very high neuron density have large range of influence, due to the way points are assigned to the clusters, which results in clusters spanning across multiple layers. In other words, if \(d_c\) is large, cluster centers will be found mostly in layers II and IV, and neurons in all other layers will be assigned to clusters with centers in the two dense layers. Reasonable clusters started to emerge with the \(d_c \approx 300\) \(\mu m\). Conversely, if chosen \(d_c\) is small, cluster centers will appear in all layers, with very granular clustering obtained with \(d_c < 100\) \(\mu m\). Although some clusters may consist of neurons from single layer, influence of cluster centers of higher density prevents the formation of clusters with distinct border between the layers. Performance of FSDP algorithm with \(d_c = 250\) \(\mu m\) is shown in Fig. 5.

Figure 5: FSDP clustering. Left: points with high density \((\rho)\) and large distance from another point with higher density \((\delta)\) emerge as outliers in top-right corner of the decision graph. Middle: Cluster centers are formed mostly in the areas of high neuron density. Right: final clustering output shows that the FSDP algorithm is unable to separate areas with different densities and create homogeneous clusters.

Since choosing the cutoff distance \(d_c\) is not straightforward, some algorithms that address this issue have been proposed \([42], [25]\). However, these algorithms often compute kernels over all data points, which is impractical for very large numbers of neurons. Also, it is questionable whether neurons at large distances should influence measurements of each other.

In conclusion, clustering algorithms were unable to provide satisfying results based only on neuron locations. It seems that the neuron distribution in the cortex is such that their intrinsic clustering structure may not be well characterized by a single set of global density parameters. However, the investigations performed have provided important insight into some of its properties. The algorithms were performing best when considering neurons within the radius between 100\(\mu m\) and 300\(\mu m\). In terms of numbers of neurons, the algorithms used between 20 and 500 neurons.
to yield reasonable results. This leads to a conclusion that the changing nature of neuron distribution in the brain is best described when performing measurements in this range.

4 Individual and regional neuron features

The considerations from cluster analysis lead to conclusion that considering only the neuron locations is not enough for establishing clear delineation of cortical layers. Other characterizations were needed, such as the second fundamental property of neurons, their shape. In this section, other characteristics of neurons are derived from fundamental ones, such as density from neuron locations, or size from its shape. Characterizations are further developed based on either individual properties of neurons, or by using some statistics of properties of neurons in its neighborhood. The focus of this section is development and exploration of the properties that characterize individual neurons and their measurement in different layers.

One of the conclusions that may be derived from density-based cluster analysis is that fixed-size, predefined radius or kernel methods do not possess discriminative power strong enough to discover layers in laminar structure. There are other approaches to measuring variations in neuron density around an individual neuron, such as computing some statistics using a predefined number of its neighbors, rather than all neurons in predefined area. Although the two approaches may seem equivalent, analysis of nearest neighbors has some advantages. For instance, a predefined radius may be interpreted differently, depending on the image resolution. Also, if it is known that a method is using fixed number of neighbors for each neuron, efficient data structures like kd-trees [3], [23] may be precomputed, enabling faster processing. Considering the large number of neurons found in a histological section, efficiency may be of critical importance.

4.1 Neuron measures

From its shape and location, several features were computed for each neuron, including its area, perimeter, as well as shape descriptors like circularity and roundness. These features form the basis for automated investigations in brain microanatomy. Features like mean, mode and standard deviation of gray value, minimum and maximum gray level were measured but not used in the analysis as these may be heavily influenced by uneven staining across the section. Moreover, features based on gray level and staining, may exhibit different values in different sections, putting in question comparison of measurements made on these sections. However, they were proven very useful in detection of image elements misidentified as neurons. Neuron locations were obtained using and automated method [38] and segmented using the method from [40], which uses grayscale-guided watershed to separate neurons, rather that distance maps obtained from gray-level thresholding. The two steps provided a binary image of segmented, non-overlapping neuron areas. Combined with the original histological image, the two were used as an input to ImageJ particle analysis pipeline [32] for computation of statistics and shape descriptors of individual blobs which represent neurons in the original image.

By examining the distribution of values obtained in measuring various statistics of neurons’ shape, some important conclusions can be made about their distribution across the cortex, and especially with regard to identification of outliers, or image elements that were incorrectly identified as neurons. Values of all shape statistics were examined and neurons with limit values were examined visually in the histological image. Neurons with values at the tails of the distributions were plotted over the histological image to observe the distribution and occurrence of neurons with specific characteristics across the layers. For instance, neurons with largest area were found in layer III of the cortex, and were followed by neurons of layer V and layer VI. Out of 50 largest neurons, 42(84%) were found in layer III, 5(10%) in layer V and 2(4%) in layer VI. Out of 500 largest neurons, 266(53%) were found in layer III, 136(27%) in layer V and 60(12%) in layer VI. This comparison confirms that the procedure follows neuroanatomical observations from classical literature, as discussed in Section 2.

Measurements of variations in grayscale intensity provided important characteristics for outliers detection, and their values were expressed differentially in the cortical layers. Mean grayscale value was found low for the neurons in layer IV, but after close visual examination most of them were neurons in group formations, thus creating larger contrast in the image. On the other hand, neurons of layer VI had lowest individual neuron grayscale value. Neurons with highest mean grayscale values were mostly found in layer I. Measurements of mode, minimum and maximum grayscale values yielded similar results. Neurons with lowest median were predominantly found in layer VI, also in layer IV, and in the middle of layer III, sometimes referred to as layer IIIb. Those found in layer IV were usually overlapping neuron or small neuron clusters that grouped together appear darker in the image. No conclusion was made on the reason found for neurons of layer VI having such low individual grayscale intensities. Standard deviation of the values was not expressed differentially in layers, but neurons with very low standard deviation were usually found to be artefacts. Skewness and especially kurtosis had strong discriminative power in identification misidentified neurons.
Neurons with values at the tails of the distribution for these two statistics were excluded as outliers. Neuron **circularity** and **roundness** were found the lowest in layer VI - which is known to be made of multipolar neurons with dendrites reaching in many directions. This in another case in which features of automatically segmented neurons reflect the descriptions from neuroanatomy and neuron morphology.

Most of the neurons with values at the either tail of distributions for many basic neuron measures were found to be outliers, image elements misidentified as neurons, which made approximately 1.81% of automatically detected neurons. It is important to identify and exclude such outliers, as they may influence measurements taken on neurons in the area in close proximity of an outlier. The outliers were detected by different measures, usually having extreme values in several of them.

### 4.2 Dependence of nearest neighbors to distance and area

An informative measure that integrates number of neighbors and radius of its neighborhood is to observe the distance from a neuron to its $k$-th neighbor. Fig. 6 shows how the distances are increasing in different layers. Maximum and minimum values, as well as mean and standard deviation are also shown. Distance to the $k$-th nearest neighbor increases in varying rates for neurons found in more or less dense layers. Due to the very low density, neurons of layer I and especially white matter have in average larger distances to the $k$-th nearest neighbor. Conversely, neurons in layers II and IV will find their neighbors in close vicinity, resulting in a curve with lower slope. A property also distinguished between the layers is the mean size of neighboring neurons. The curves for each layer are shown with shared $y$-axis in Fig. 7.

![Figure 6: Distances to the $k$-th nearest neighbor for neurons in different layers. Mean (solid red line), median (dashed red line) and standard deviation around the mean (blue lines) exhibit different slope between the layers, depending on the neuron density in the neighborhoods around neurons. Top, middle and bottom row represent distances to 10, 100 and 1000 nearest neighbors, respectively.](image)

![Figure 7: Top: mean distances to the 100th nearest neighbor. Curve slope is determined by neuron density - steeper slope indicates larger distances between the neurons. Bottom: mean area of the neurons up to the 200th nearest neighbor.](image)
4.3 Measures of neighboring neurons

Observing properties of a defined number of nearest neighbors of a particular neurons has several advantages over observations made in fixed radius, both in interpretability and computational efficiency, as mentioned in the previous section. For the distances to a neurons’ $k$-th nearest neighbor, mean, skewness, kurtosis, bin entropy and other statistics were computed for different values of $k \in \{50, 100, 250, 500, 1000\}$. To increase computation efficiency, a kd-tree was computed which enabled fast queries on a chosen number of neighbors. For brevity, only some of the features whose changes between layers facilitate their segmentation are shown. Visualization of distribution of features according to layers and with respect to different number of neighbors used to compute statistics is shown in Figure 8.

Figure 8: Average neuron area for neurons of each layer and for $k \in \{100, 250, 500, 1000\}$ nearest neighbors

Statistics of other measures were also computed. However, not all are useful for distinguishing neurons within the layers. For instance, neurons with measures related to pixel intensities such as standard deviation of grayscale values or their integrated density (sum of values) were not observed to be distributed with some particular trend but are rather dispersed across all layers. The exception to this is mean or median values, having lowest values mostly in layer VI. Still, the importance of this measure was not shown to be very high. Measures regarding neuron shape like area, circularity or perimeter provided more discriminative power. This is not unexpected, since the findings in classical neuroanatomical research relies to large extent on shape and size of the neurons. In conclusion, measures that were chosen for development of further features were neuron area, perimeter, circularity and median. Densities within a fixed radius $dc$ of $25\mu m$, $50\mu m$, $125\mu m$, $250\mu m$ and $500\mu m$ were also computed for each neuron and added to this set of features.
4.4 Convex hull

Another approach that combines area around a neuron and a number of its neighbors is obtaining a convex hull of its \( k \)-neighbors. Some simple statistics used to describe the hull of each neuron were used, like hull area, perimeter, average nearest distance for neurons found in the hull and standard deviation of nearest distances. The rationale behind this approach is the assumption that neuron density is reflected through the area contained within the hull, or hull’s perimeter.

![Figure 10: Convex hulls of 100, 250, 500 and 1000 nearest neighbors of neurons in different layers.](image)

The spatial distribution of neurons may be quantitatively measured using nearest neighbor index (NNI). It is a measure of point distribution which quantitatively describes whether points follow usually subjective patterns of regular, clustered or random distribution. This measure is often used in geography and biology in studying dispersion of settlements or organization of trees in an area. For instance, NNI can be used to determine whether a forest has been planted by humans, in which case the pattern of distribution will be more regular, or has grown naturally, having a random spatial distribution. NNI measures the distance between each point and its nearest neighbor’s location. All the nearest neighbor distances are averaged, and if the average distance is less than the average for a random distribution, the distribution of the features being analyzed is considered clustered. If the average distance is greater than a random distribution, the features are considered regularly dispersed. The index is expressed as the ratio of the mean observed distance divided by the expected distance, which is based on a random distribution with the same number of points covering the same total area,

\[
NNI_i = \frac{1}{n} \frac{\sum_{j=1}^{n} d(i, j)}{0.5 \sqrt{HullArea(i)/n}}.
\]

The values of NNI are obtained in range of \( 0 \sim 2.15 \), with the interpretation of observed values around 1 as randomly distributed points. Points with higher NNI are interpreted as more uniformly dispersed while points with lower NNI as more clustered. Neurons in all layers with the exception of layer I and white matter tend more towards uniformly dispersed distribution, especially neurons of layer IV.

4.5 Second-order features

Statistics computed for neurons become neuron features themselves. All the statistics developed in the previous subsection can be used in the similar way as original neuron measures were used, as new descriptors of each neuron, and statistics of these may be computed. However, these second-order features did not bring significant advancements in discovery of features important for layer segmentation. The effect of taking a statistic of some other statistic often blurs the details created by original statistic. Also, these features may be hard to interpret from neuroanatomical perspective.

4.6 Oriented measurements

Depending on it’s position in the cortex, a neuron may be placed more towards middle or more towards the edge of its layer. Features computed around a neuron, whether for a fixed number of its neighbors or in some specified range, may be confounded by reaching into adjacent layer and using neurons with different properties for computation of statistics. To identify this case, measurements may be bounded and taken only from neurons found within the range of angle. Directing measurements at different angles around a neuron, it is possible to compare measurements made in these slices. Such features measured in several directions provide the capacity to identify border neurons and measure changes of neuronal properties. One of the simplest methods to compare measurements made in the slices is to compute their differences.
Figure 11: Measurements of a neuron’s neighborhood may be taken at angle in different directions. These are especially important for neurons at the border of layers, whose neighborhood measurement may vary in different directions, resulting in local heterogeneity of neuron features.

4.6.1 Diversity measures

Slices may be further compared considering their number and the number of neurons found within each slice. They may be regarded as measurement units reaching from a single neuron, each unit representing a population of neighboring neurons found in a given direction from the central neuron. The relationship of different populations within an area has been extensively studied in the frame of biological diversity of species, landscapes and other [22], [27]. Considering the neurons in a slice as members of a single species, and the $k$ neighbors of a neuron as the population of all species in their habitat, one may employ several biodiversity measures to evaluate the relationship between the species. In this context, the number of slices is the number of different species, or richness, and relative abundance of the different species in an area as evenness. These simple measures are commonly combined in order to utilize both aspects of diversity, which are often used and interpreted in association with other aspects of natural habitat [22]. The two most often used such measures are Shannon index [36] and Simpson index [37], who became popular through frequent use in quantitative studies of animal and plant species diversity.

The Shannon index was originally proposed in Information theory to quantify the entropy of appearance of characters in strings of text. In the context of biodiversity, it gives a quantitative measure of the uncertainty in predicting the species of an individual chosen randomly from the population. The formula for calculating the index may be written in several ways

$$\text{Shannon} = - \sum_{i=1}^{R} p_i \ln p_i = - \sum_{i=1}^{R} \ln p_i^p = \ln \left( \frac{1}{\prod_{i=1}^{R} p_i^{p_i}} \right),$$

(2)

where $R$ is the number of different species, or slices, and $p_i$ is the proportion of species of the $i$th type in the population, or proportion of neurons in $i$th slice to the number of the neurons in $k$-neighborhood. If all slices have equal number of neurons, $p_i$ values equal 1/$R$, and the Shannon index takes the maximum value of $\ln R$. If the numbers are unequal, the weighted geometric mean of the $p_i$ values is larger, which results in the index having smaller values. The index equals zero if the neurons from only one slice are present, since there is no uncertainty in predicting the slice they are in.

The Simpson index measures the probability that the two individuals randomly chosen (with replacement) from the total population will be of the same species. It is defined as

$$\text{Simpson} = \sum_{i=1}^{R} p_i^2,$$

(3)

where $R$ and $p_i$ are the same as in $2$. This equation is equivalent to the weighted arithmetic mean of the proportional abundances $p_i$, with abundances used as the weights. The index gives information about the relation of number of types and the presence of the dominant type. Mean proportional abundance of the slices increases with decreasing number of slices and the increasing abundance of the slice with largest number of neurons, the index obtains small values in regions of high diversity like neurons on borders between the layers, thin layers, and especially layer I neurons. The
index is large in homogeneous areas like the middle of layer III, where slices reaching from a neuron remain in the area of the layer. Values of the Simpson index are bound between $1/R$, which is reached when the number of neurons in all slices is equal, and unity, reached when only one slice is populated.

4.7 Distance to areas with homogeneous density

Highest-level features of neurons, developed in this section, are related to measurements of distance from a neuron to closest neuron with certain values of previously developed features. For instance, as seen in Fig. 2, neurons of layer I and white matter can be distinguished by having small neuron density in their vicinity. By this, sparse regions of the section are identified. In similar fashion are identified the dense regions containing layers II and IV. Moreover, the sparse region may further be split using the hull area feature developed in previous section - neurons in white matter will have large hull area, in contrast to the neurons of layer I, whose hull is bound between the border of the tissue and the dense layer II. By this, layer I and white matter neurons are identified. It is not a definitive and final classification of all neurons in these layers. Some neurons may have been omitted, for it is important to be restrictive when identifying neurons in these regions, since misidentification of a single neuron of, for instance, layer III for neuron of layer I has a significant impact on the whole region of the digitized slice. Neurons of white matter may further be split by their distance to the sparse regions and hull area into sulcal and gyral neurons, since there is an apparent bimodal distribution of these features. Distances to these areas are used as high-level neuron features. By computing distances to layer I and white matter, cortical thickness and depth of each neuron can be computed, shown in Fig. 12 which is an important measure to obtain.

Figure 12: Developed high-level features based on local density and convex hull radius are used to obtain tissue features without drawing profiles and sampling perpendicular to the cortex. Left: sparse areas are separated into layer I and white matter. Middle: cortical depth. Right: cortical thickness.

Dense region may be split into layer II and layer for using the distance to layer I feature. Neurons of layer II will have smaller distance to layer I than those of layer IV, since there is the wide layer III separating the two dense layers. Finally, taking into account results from clustering analysis and the experience with the neuron features developed so far, standard deviation and mean of 250 nearest neighbors’ distance to sparse and dense regions were added to the feature set for each neuron in the section. Some results from development of these features are shown in Fig. 13.

5 Mapping neurons to cortical layers using machine learning

The extensive feature set developed in the previous section provides informative descriptors of cortical organization. However, no single feature was found that would be able to provide clear segmentation of cortical layers. Although some features were clearly more expressed in certain layers than the others, it is not straightforward what exactly is changing between the layers as well as the impact and interconnection of different features. This leads to an assumption that there are information contained in the developed features that can be analyzed, combined and used to produce precise classification of neurons with regard to their location within the cortical layers.
Developed high-level features based on several statistics are used for estimates of layer positions. Left: map of distances to sparse regions. Middle: Sparse regions are split into layer I and white matter regions, in which sulcal and gyral neurons are identified. Right: final output of the 4-region segmentation.

The goal is to learn how to use the developed features to classify neurons and produce segmentation of cortical layers. In this section, a classical supervised machine learning method is described, guided by manual segmentation of layers are used to investigate feature importances, create models that predict the neurons’ layers and finally produce the segmentation of cortical layers on the whole histological section.

5.1 Dataset of manually labeled cortical layers

A portion of the histological preparation of human prefrontal cortex was given to three human experts who manually delineated borders between the layers in the cortex. The apparent inconsistencies and mutual disagreement between the experts on where to draw the boundaries shows the presence of experts’ bias. The agreement of all three raters may be as low as 8.6%, for the thin layer IV. The experts’ agreement on the neurons found in the layers is shown in Fig. 14.

The experts disagreed on the boundaries of all layers, with the exception of the very apparent layer I/layer II boundary. Cortical layers as delineated by the experts are shown in Fig. 15. As it is often the case with research in biological sciences, especially when dealing with human specimens, the amount of data available for use as baseline in this research was very limited, which is an important factor when choosing appropriate learning methods.

5.2 Machine learning models

Reviewing the literature on machine learning [4], [19], decision trees were chosen as the method for prediction and interpretation of cortical lamination for its several advantages, being a state of the art for the prediction problems with tabular style input data such as neuron features. Decision trees require no data preparation and neuron features developed in previous section can be directly used as input for learning. They also need not be scaled or normalized,
Computational analysis of laminar structure of the human cortex based on local neuron features

**Figure 15:** Cortical layers as delineated by the experts. Significant disagreement on the boundaries of the layers can be observed, as well as on the boundary between the cortex and the white matter.

since decision trees do not require data normalization, as is the case for many other methods such as clustering methods or neural networks. Another important characteristic is good performance without large training data and application on large datasets using standard computing resources performed in reasonable time.

Besides being a successful and widely-used, decision trees are simple to use, interpret and understand. They can easily be visualized, the rules used for splitting nodes can be examined, input features analyzed using built-in feature selection and importance measurements, and the results can easily be interpreted and communicated even to non-experts [14]. This was a desirable property, since this research is not only oriented towards parcellation of cortical layers but also to understanding what tissue features cause the parcellation.

Decision trees mirror human decision making more closely than other approaches [14], which is especially useful when modeling human activities, such as manual delineation of cortical layers, a straight example of decision-making process based on combination of several information about neurons’ characteristics.

### 5.2.1 Tree classifiers

Learning method used for initial training and prediction of neurons’ layers was decision tree classifier. Experiments were performed using scikit-learn [30], a Python library. Using stratified k-fold, the tree classifier was able to achieve 85% cross validation score. However, not all neuron features have proven useful and informative for the classifier, many of them having feature importance measure of 0. Moreover, reducing the number of features used has shown increase in model’s performance, which was able to achieve cross validation score of approximately 90%. Validation curves for the two cases are shown in Fig. 16.

**Figure 16:** Validation curve for a single decision tree. The model was able to achieve better performance using a smaller number of selected high-level features.

Experiments with different sets of features have shown that although some combinations of features do achieve high accuracy on training data, that does not guarantee that the model will perform well on the whole histological section. The introduction of features based on distance to sparse or dense regions has significantly improved model’s ability to separate regions of the sections into parcels following the laminar layout of the cortex. However, a simple tree learner was not able to comprehend the variations of the neuronal features well enough to be able to generalize to the whole histological section.
To produce more precise segmentation of cortical layers through accurate prediction of neurons’ layers, decision trees were combined using gradient boosting method, a well-known and widely used approach \cite{5,24}. Specifically, XGBoost (extreme gradient boosting) \cite{8}, a popular and efficient open-source implementation of the gradient-boosted trees algorithm in Python was used. This method was able to achieve accuracy of 1 on training data, with different subsets of features which, again, not all provided good generalization to the whole sections. The best results were obtained when combining the output of all three raters. Gradient boosting model was trained for each rater using softmax objective. The output probabilities were then summed, and the final prediction was made. Results of this approach are shown in \cite{17} together with the probability map of the model’s output. Classes of neurons are predicted, and neurons grouped in a way that follows laminar pattern of the cortex. This was achieved using mostly high-level features like oriented measurements, distances from areas of homogeneous densities, and also deviation and mean of these distances. Investigation of the impact of features on prediction of class was performed on both global (whole model) and local (individual neuron) level.

The ensemble system provided a confidence map (darker color signifies higher confidence) and predicted layers of the cortex.

5.3 Analysis of individual feature attribution

For deeper understanding of both the model and the data that is being used for making predictions, interpretation of the results beyond accuracy and overall performance is needed. To that end, of great importance is to understand the impact of each feature that is used for making predictions in models such as ensembles by evaluating the importance values that are attributed to each input feature.

A recent approach for measuring feature attribution, the SHAP measure, proved optimal for estimation of neuron features that facilitate neuron classification and cortical segmentation. It was preferred over classical measures such as gain or cover as it is the only measure that satisfies both consistency and accuracy. A method is said to be consistent if the importance estimate of a feature cannot decrease if there is a change of the model such that the model relies more on that feature. The accuracy property states that the sum of all the feature importances should equal the total importance of the model \cite{20}. Other considered methods were weight, cover, permutation and Sabaas \cite{34}. Importances of the features used for training the ensemble model in the previous section are shown in Fig. 18.
Computational analysis of laminar structure of the human cortex based on local neuron features

**Figure 18:** Importance of neuron features in SHAP analysis.

An important aspect of this approach is the measurement and visualization of features that contributed to making a prediction on a single instance of the data. Fig. 19 shows which neuron features of a single neuron of layer VI contributed to increase from the base SHAP value and making the prediction. The image also shows the importance of features that decreased the output value for prediction of the same neuron as a white matter neuron.

**Figure 19:** Contribution of different features for making a prediction of a neuron’s layer. Top: neuron features of a single neuron of layer VI that contributed to increase from the base SHAP value and making the prediction. Bottom: the importance of features that decreased the output value for prediction of the same neuron as a white matter neuron.

### 6 Conclusion

This paper presents a novel framework for analysis of laminar structure of the cortex. From the location and basic measures of neurons, more complex features were developed for use in machine learning models. Tree ensembles, as today’s one of the most powerful and interpretable models were used on data manually labeled by three human experts. The most accurate classification results were obtained by training three XGBoost models separately on three datasets provided by the experts and creating another ensemble from the three, combining their probability outputs for final neuron class prediction.

Measurement of importances of developed features for on both global model level and individual prediction level was performed using SHAP values, an only possible consistent and accurate method for assessment of individual feature
Computational analysis of laminar structure of the human cortex based on local neuron features

Features of the highest level were proven to have the largest importance. This is because they integrate low level feature with anatomical observations, and have a wide reach across the histological section. Features based on oriented measurements, or slices, had considerable feature importance, being able to identify neurons on the border of layers and help towards more accurate classification. It was shown that building on lower-level features by computing some statistics or combining them in some other way yields features that have greater discriminative power and thus greater measured importance. This is probably because of their capacity to overcome local variations in neuron features and taking, for instance, mean of those features. In contrast, using very local features or measures of a neuron such as area, there is no increase in accuracy of prediction of the neuron layer, this features often having importance of 0. This is probably why different methods for local pattern analysis and classical image feature extraction methods are not very successful in cortical layer analysis. An range of variability radius was established, giving an estimate of the size of the neuron’s neighborhood in which measurements should be made in order for it to be large enough to overcome local variations in neuron distribution and recognize its location within the cortical structure on one hand, and narrow enough so that measurements are not confounded by reaching too far into adjacent layers.

During experiments with different models and combinations of datasets, it was noticed that parts of layer III and layer VI are sometimes divided into sublayers that follow the direction of laminar structure, although being very short and not extending through a significant portion of the slice. Further investigations using the developed methodology may provide more detailed insight into sublayering of cortical structure.

Acknowledgments.

This publication was supported by the European Union through the European Regional Development Fund, Operational Programme Competitiveness and Cohesion, grant agreement No. KK.01.1.01.0007, CoRE - Neuro; and the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative.

References

[1] K. Amunts, C. Lepage, L. Borgeat, H. Mohlberg, T. Dickscheid, M.-É. Rousseau, S. Bludau, P.-L. Bazin, L. B. Lewis, A.-M. Oros-Peusquens, et al. Bigbrain: an ultrahigh-resolution 3d human brain model. Science, 340(6139):1472–1475, 2013.
[2] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander. Optics: ordering points to identify the clustering structure. In ACM Sigmod record, volume 28, pages 49–60. ACM, 1999.
[3] J. L. Bentley. Multidimensional binary search trees used for associative searching. Communications of the ACM, 18(9):509–517, 1975.
[4] M. Bramer. Principles of data mining, volume 180. Springer, 2007.
[5] L. Breiman. Arcing the edge. Technical report, Technical Report 486, Statistics Department, University of California at Berkeley, 1997.
[6] K. Brodmann. Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. Barth, 1909.
[7] R. J. Campello, D. Moulavi, A. Zimek, and J. Sander. Hierarchical density estimates for data clustering, visualization, and outlier detection. ACM Transactions on Knowledge Discovery from Data (TKDD), 10(1):5, 2015.
[8] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pages 785–794. ACM, 2016.
[9] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd, volume 96, pages 226–231, 1996.
[10] B. J. Frey and D. Dueck. Clustering by passing messages between data points. science, 315(5814):972–976, 2007.
[11] A. Hopf. Registration of the myeloarchitecture of the human frontal lobe with an extinction method. Journal fur Hirnforschung, 10(3):259, 1968.
[12] A. Hudspeth, J. Ruark, and J. Kelly. Cytoarchitectonic mapping by microdensitometry. Proceedings of the National Academy of Sciences, 73(8):2928–2931, 1976.
[13] A. K. Jain. Data clustering: 50 years beyond k-means. Pattern recognition letters, 31(8):651–666, 2010.
[14] G. James, D. Witten, T. Hastie, and R. Tibshirani. An introduction to statistical learning, volume 112. Springer, 2013.
Computational analysis of laminar structure of the human cortex based on local neuron features

[15] M. Judaš, G. Šimić, Z. Petanjek, N. Jovanov-Milošević, M. Pletikos, L. Vasung, M. Vukšić, and I. Kostović. The zagreb collection of human brains: a unique, versatile, but underexploited resource for the neuroscience community. *Annals of the New York Academy of Sciences*, 1225(1), 2011.

[16] J. H. Kaas. The functional organization of somatosensory cortex in primates. *Annals of Anatomy-Anatomischer Anzeiger*, 175(6):509–518, 1993.

[17] L. Kaufman and P. J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons, 2009.

[18] D. Li, M. Zavaglia, G. Wang, Y. Hu, H. Xie, R. Werner, J.-S. Guan, and C. C. Hilgetag. Discrimination of the hierarchical structure of cortical layers in 2-photon microscopy data by combined unsupervised and supervised machine learning. *bioRxiv*, page 427955, 2018.

[19] S. Liu, X. Wang, M. Liu, and J. Zhu. Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics*, 1(1):48–56, 2017.

[20] S. M. Lundberg, G. G. Erion, and S.-I. Lee. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*, 2018.

[21] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[22] A. E. Magurran. *Measuring biological diversity*. John Wiley & Sons, 2013.

[23] S. Maneewongvatana and D. M. Mount. It’s okay to be skinny, if your friends are fat. In *Center for Geometric Computing 4th Annual Workshop on Computational Geometry*, volume 2, pages 1–8, 1999.

[24] L. Mason, J. Baxter, P. L. Bartlett, and M. R. Frean. Boosting algorithms as gradient descent. In *Advances in neural information processing systems*, pages 512–518, 2000.

[25] R. Mehmood, G. Zhang, R. Bie, H. Dawood, and H. Ahmad. Clustering by fast search and find of density peaks via heat diffusion. *Neurocomputing*, 208:210–217, 2016.

[26] H. S. Meyer, V. C. Wimmer, M. Oberlaender, C. P. De Kock, B. Sakmann, and M. Helmstaedter. Number and laminar distribution of neurons in a thalamocortical projection column of rat vibrissal cortex. *Cerebral cortex*, 20(10):2277–2286, 2010.

[27] H. Nagendra. Opposite trends in response for the shannon and simpson indices of landscape diversity. *Applied geography*, 22(2):175–186, 2002.

[28] S. Nosova, L. Snopova, and V. Turlapov. Automatic detection of neurons, astrocytes, and layers for nissl-stained mouse cortex. 2017.

[29] N. Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[31] A. Rodriguez and A. Laio. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–1496, 2014.

[32] C. T. Rueden, J. Schindelin, M. C. Hiner, B. E. DeZonia, A. E. Walter, E. T. Arena, and K. W. Eliceiri. ImageJ2: ImageJ for the next generation of scientific image data. *BMC bioinformatics*, 18(1):529, 2017.

[33] M. Ryzen. A microphotometric method of cell enumeration within the cerebral cortex of man. *Journal of Comparative Neurology*, 104(2):233–245, 1956.

[34] A. Saabas. Interpreting random forests, 2014.

[35] A. Schleicher, K. Amunts, S. Geyer, P. Morosan, and K. Zilles. Observer-independent method for microstructural parcellation of cerebral cortex: A quantitative approach to cytoarchitectonics. *NeuroImage*, 9(1):165 – 177, 1999.

[36] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[37] E. H. Simpson. Measurement of diversity. *Nature*, 1949.

[38] A. Štajduhar, D. Džaja, M. Judaš, and S. Lončarić. Automatic detection of neurons in neun-stained histological images of human brain. *Physica A: Statistical Mechanics and its Applications*, 519:237–246, 2019.

[39] C. F. von Economo and G. N. Koskinas. *Die cytoarchitektonik der hirnrinde des erwachsenen menschen*. J. Springer, 1925.
[40] A. Štajduhar, C. Lepage, M. Judaš, S. Lončarić, and A. C. Evans. 3d localization of neurons in bright-field histological images. In ELMAR (ELMAR), 2018 60th International Symposium, pages 75–78. IEEE, 2018.

[41] K. Wagstyl, C. Lepage, S. Bludau, K. Zilles, P. C. Fletcher, K. Amunts, and A. C. Evans. Mapping cortical laminar structure in the 3d bigbrain. Cerebral Cortex, 28(7):2551–2562, 2018.

[42] S. Wang, D. Wang, C. Li, and Y. Li. Comment on "clustering by fast search and find of density peaks". arXiv preprint arXiv:1501.04267, 2015.

[43] K. Zilles, A. Schleicher, N. Palomero-Gallagher, and K. Amunts. Quantitative analysis of cyto-and receptor architecture of the human brain. In Brain Mapping: The Methods (Second Edition), pages 573–602. Elsevier, 2002.