



Multi-dimensional classification of GABAergic interneurons with Bayesian network-modeled label uncertainty

Bojan Mihaljević^{1*}, Concha Bielza¹, Ruth Benavides-Piccione^{2,3}, Javier DeFelipe^{2,3} and Pedro Larrañaga¹

¹ Computational Intelligence Group, Departamento de Inteligencia Artificial, Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid, Madrid, Spain

² Laboratorio Cajal de Circuitos Corticales, Centro de Tecnología Biomédica, Universidad Politécnica de Madrid, Madrid, Spain

³ Instituto Cajal, Consejo Superior de Investigaciones Científicas, Madrid, Spain

Edited by:

David Hansel, University of Paris, France

Reviewed by:

Volker Steuber, University of Hertfordshire, UK

Benjamin Torben-Nielsen, Okinawa Institute of Science and Technology Graduate University, Japan

*Correspondence:

Bojan Mihaljević, Departamento de Inteligencia Artificial, Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, Boadilla del Monte 28660, Madrid, Spain
e-mail: b.mihaljevic@fi.upm.es

Interneuron classification is an important and long-debated topic in neuroscience. A recent study provided a data set of digitally reconstructed interneurons classified by 42 leading neuroscientists according to a pragmatic classification scheme composed of five categorical variables, namely, of the interneuron type and four features of axonal morphology. From this data set we now learned a model which can classify interneurons, on the basis of their axonal morphometric parameters, into these five descriptive variables simultaneously. Because of differences in opinion among the neuroscientists, especially regarding neuronal type, for many interneurons we lacked a unique, agreed-upon classification, which we could use to guide model learning. Instead, we guided model learning with a probability distribution over the neuronal type and the axonal features, obtained, for each interneuron, from the neuroscientists' classification choices. We conveniently encoded such probability distributions with Bayesian networks, calling them *label Bayesian networks* (LBNs), and developed a method to predict them. This method predicts an LBN by forming a probabilistic consensus among the LBNs of the interneurons most similar to the one being classified. We used 18 axonal morphometric parameters as predictor variables, 13 of which we introduce in this paper as quantitative counterparts to the categorical axonal features. We were able to accurately predict interneuronal LBNs. Furthermore, when extracting crisp (i.e., non-probabilistic) predictions from the predicted LBNs, our method outperformed related work on interneuron classification. Our results indicate that our method is adequate for multi-dimensional classification of interneurons with probabilistic labels. Moreover, the introduced morphometric parameters are good predictors of interneuron type and the four features of axonal morphology and thus may serve as objective counterparts to the subjective, categorical axonal features.

Keywords: probabilistic labels, consensus, distance-weighted k nearest neighbors, multiple annotators, neuronal morphology

1. INTRODUCTION

There are two main neuron subpopulations in the cerebral cortex: excitatory glutamatergic neurons, constituting approximately 80% of all cortical neurons, and inhibitory GABAergic interneurons, representing the remaining 20%. Although less numerous, GABAergic interneurons (for simplicity, interneurons), play multiple critical cortical functions and are highly heterogeneous with regards to their morphological, electrophysiological, and molecular properties (Ascoli et al., 2007). Neuroscientists consider that these differences indicate that various types of interneurons actually exist and that the differences among them are functionally relevant. Although many different classification schemes have been proposed so far (e.g., Fairén et al., 1992; Kawaguchi, 1993; Cauli et al., 1997; Somogyi et al., 1998; Gupta et al., 2000; Maccaferri and Lacaille, 2003), there is no universally accepted catalog of interneuron types (DeFelipe et al., 2013), making it hard to share

and organize data and the knowledge derived from them. Ascoli et al. (2007) have identified a large set of morphological, electrophysiological, and molecular properties which can be used to distinguish among interneuron types. However, gathering such comprehensive data has considerable practical burdens (DeFelipe et al., 2013), making it hard to follow such a classification in practice.

Therefore, DeFelipe et al. (2013) proposed an alternative, pragmatic classification scheme, based on patterns of axonal arborization. The scheme classifies interneurons according to their type and four other features of axonal morphology. It contemplates ten types, most of them well established in literature, such as Martinotti and chandelier, and provides rather precise definitions of their axonal and dendritic morphology. The remaining axonal features are categorical properties such as axon's columnar and laminar reach (i.e., whether it is intra- or trans-columnar;

intra- or trans-laminar)¹. To assess the viability of this classification scheme, that is, whether it is useful for cataloging interneurons, DeFelipe et al. (2013) convened 42 leading neuroscientists to classify 320 interneurons. While the experts easily distinguished some of the neuronal types and the four remaining features, they found some types to be somewhat confusing.

Nonetheless, the data they gathered provides a basis for building an objective, automatized classifier, which would map quantitative neuronal properties to interneuron types and the categories of axonal features. Automatic classification of neurons has been mainly done in an unsupervised fashion (Jain, 2010), seeking to discover groups on the basis of quantitative properties alone (Cauli et al., 2000; Tsiola et al., 2003; Karagiannis et al., 2009; McGarry et al., 2010). However, the availability of expert-provided input on interneuron type membership and their axonal features allows us to learn a model in a supervised fashion (Duda et al., 2000), as done by, e.g., Marin et al. (2002) and Druckmann et al. (2013). When such supervision information is available, supervised learning can yield more accurate models than unsupervised learning (Guerra et al., 2011). In addition, a model obtained in this way can be used to replace experts, as it can, given an interneuron, automatically predict its properties (the type and axonal features).

Using the neuroscientists' classification choices as input for supervised classification is challenging due to the ambiguity in type membership and axonal features of the interneurons. While this ambiguity varied across our data, some interneurons were especially ambiguous: e.g., one was assigned to six different types, with at most 14 (out of 42) experts agreeing on one of these types. Previous efforts to predict the neuronal type and axonal features (DeFelipe et al., 2013; Mihaljević et al., 2014a,b) considered such majority choices as *ground truth*, i.e., as the true type and axonal features, and therefore, for each interneuron, disregarded the opinions of the disagreeing neuroscientists (with the majority for that interneuron). While Mihaljević et al. (2014b) only predicted the neuronal type, DeFelipe et al. (2013) and Mihaljević et al. (2014a) built an independent model for each axonal feature, although these features are complementary.

In this paper, we predict interneuron type and axonal features simultaneously, while accounting for class label ambiguity in a principled way. Namely, for each interneuron, we encode the neuroscientists' input with a joint probability distribution over the five class variables². That is, we consider that each

interneuron has a certain probability of belonging to each possible combination of the five axonal features. Assuming that all experts were equally good at classifying interneurons, these probabilities are given by the relative frequencies of such combinations in the expert-provided input. This way, we take the opinions of all annotator neuroscientists into account. Such probability distributions can be compactly encoded with Bayesian networks (Pearl, 1988; Koller and Friedman, 2009), given sufficient conditional independencies among the variables. We will therefore represent these joint probability distributions over class variables with Bayesian networks and call them *label Bayesian networks* (LBNs). As a first step in the present study, we will obtain LBNs from the experts' input; subsequently, we will train and evaluate our model using LBNs as input.

To the best of our knowledge, this is the first paper tackling multi-dimensional classification (i.e., with multiple class variables; Van Der Gaag and De Waal, 2006; Bielza et al., 2011) with probabilistic labels. Multi-dimensional classification is hard because of dependencies among class variables: ignoring them, by building a separate model for each variable, is suboptimal, while modeling them can result in data scarcity if there are more than a few class variables. Instead of identifying global dependencies among class variables, we predict the LBN of an interneuron by looking at the interneurons most similar to it (i.e., its neighbors in the space of predictor variables), following the lazy-learning *k*-nearest neighbors method (*k*-nn) (Fix and Hodges, 1989). Having found the neighbors of an interneuron, we predict its LBN by forming a consensus Bayesian network (e.g., Matzkevich and Abramson, 1992) among the neighbors' LBNs. In order to give more weight in the consensus distribution to the LBNs of the closer neighbors, we adapt the Bayesian network consensus method developed by Lopez-Cruz et al. (2014).

Note that our method takes LBNs, rather than the expert-provided labels, as input, thus abstracting away the annotators. In a similar real-world scenario, this might be useful for hiding the annotators' labels from the data analyst, for reasons such as confidentiality protection. Furthermore, an LBN could be obtained in multiple ways: by learning from data, eliciting from an expert, or combining expert knowledge and learning from data.

In order to predict the neuronal type and axonal features, we introduce 13 morphometric parameters of the axon to be used as predictor variables. We defined these parameters seeking to capture the concepts represented by the four axonal features (other than neuronal type) and implemented software that computes them from digital reconstructions of neuronal morphology. In addition, we used five other axonal morphometric parameters, computed with NeuroExplorer (Glaser and Glaser, 1990), which were already used as predictors of neuronal type by DeFelipe et al. (2013) and Mihaljević et al. (2014b). In total, we used 18 axonal morphometric parameters as predictor variables.

We found that our method accurately predicted the probability distributions encoded by the LBNs. Also, for comparison

¹Following DeFelipe et al. (2013), we will often refer to the type and the four axonal features simply as *axonal features* (i.e., interneuron type is also encompassed by this term).

²From this point on, we adopt some machine learning terminology. Namely, in machine learning, each (discrete) predicted variable—in our case, these are the five axonal features—is called a *class variable*; this term, therefore, applies to each of the five axonal features, even though only the neuronal type is a class in the usual meaning of this term in neuroscience. *Class labels* are the assignments to the class variables associated with the data points (interneurons); e.g., an interneuron might be labeled as intralaminar. We will sometimes refer to expert neuroscientists as *annotators* because they annotated (labeled) the data with class labels. *Predictor variables* are the independent variables in a model, e.g., we use morphometric properties of an interneuron as predictor variables to predict interneuron type and other axonal features (the latter

are the class variables). We will not use the term “features” as a synonym of “predictor variables,” as commonly done in machine learning, in order to avoid confusion with the term “axonal features.” We will use the terms “axonal features” and “class variables” interchangeably.

with previous work on interneuron classification, we assessed the prediction of the majority class labels, and found that we outperformed (DeFelipe et al., 2013) in per-class majority label accuracy.

The rest of this paper is structured as follows. Section 2 describes the data set, the interneuron nomenclature due to DeFelipe et al. (2013), the morphometric parameters, including the ones we introduce in this paper, and the extraction of LBNs from expert-provided labels; it also describes the proposed method—the distance-weighted consensus of k nearest Bayesian networks—the related methods, the metrics for assessing our method's predictive performance, and, finally, specifies the experimental setting. We provide our results in Section 3, discuss them in Section 4, and conclude in Section 5.

2. MATERIALS AND METHODS

2.1. NEURONAL RECONSTRUCTIONS

We used neuronal reconstructions and expert neuroscientists' terminological choices that were gathered by DeFelipe et al. (2013). Of the 320 interneurons classified in that study, 241 were digitally reconstructed cells (retrieved by DeFelipe et al., 2013 from NeuroMorpho.Org, Ascoli et al., 2007), coming from different areas and layers of the cerebral cortex of the mouse, rat, and monkey. Forty of the reconstructions had one or multiple interrupted (i.e., with non-continuous tracing) axonal processes; when deemed feasible (36 cells), we unified the axonal processes using NeuroLucida (MicroBrightField, Inc., Williston, VT, USA). We omitted the remaining four cells from our study, reducing our data sample to 237 cells.

2.2. AXONAL FEATURE-BASED NOMENCLATURE

DeFelipe et al. (2013) asked 42 expert neuroscientists to classify the above-described interneurons according to the interneuron nomenclature they proposed. The nomenclature consists of six categorical features of axonal arborization. The features' categories are the following:

- Axonal feature 1 (C_1): intralaminar and translaminar
- Axonal feature 2 (C_2): intracolumnar and transcolumnar
- Axonal feature 3 (C_3): centered and displaced
- Axonal feature 4 (C_4): ascending, descending, both, and no
- Axonal feature 5 (C_5): arcade (AR), Cajal-Retzius (CR), chandelier (CH), common basket (CB), common type (CT), horse-tail (HT), large basket (LB), Martinotti (MA), neurogliaform (NG), and other (OT)
- Axonal feature 6 (C_6): characterized and uncharacterized

Cells whose axon is predominantly in soma's cortical layer are intralaminar in C_1 ; the rest are translaminar. Similarly, regarding C_2 , interneurons with the axon predominantly in soma's cortical column are intracolumnar; the rest are transcolumnar. A cell whose dendritic arbor is mainly

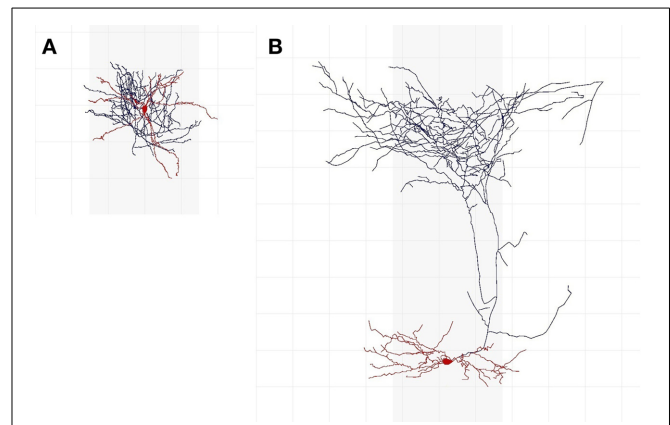


FIGURE 1 | Examples of interneurons of different types and axonal features. (A) Is an intralaminar, intracolumnar, centered, and no cell, according to 37 (out of 42) experts. Most of its axon (shown in blue) is at less than 200 μm from the soma (shown in red); the grid lines are separated by 100 μm and thus appears to be mainly located in soma's cortical layer; it is within soma's cortical column (the gray vertical shadows depict a 300 μm -wide cortical column); and it seems to be centered around the dendritic arbor (also shown in red). It is NG according to 18 experts, CB according to 17 experts, CT according to 3 experts, and OT and AR according to one expert each. **(B)** Is a translaminar, transcolumnar, displaced, and ascending cell according to 39 experts. Its axon reaches around 800 μm above soma (i.e., it seems to extend to another layer); a significant portion of its axon is outside of soma's cortical column; its dendrites are not in the center of the axonal arborization; and its axon is predominantly above the soma. According to 34 experts, this is a MA cell.

located in the center of the axonal arborization is centered (C_3); otherwise it is displaced. C_4 further distinguishes between translaminar (C_1) and displaced (C_3) cells: cells with an axon mainly ascending toward the cortical surface are ascending, cells with an axon mainly descending toward the white matter are descending, whereas those with both ascending and descending arbors are termed both. To those cells that were not translaminar (C_1) and displaced (C_3) we assigned no in C_4 (this category was not contemplated in the original nomenclature). Class C_5 is the interneuron type. A cell is uncharacterized in C_6 if it is not suitable for characterization according to features C_1 – C_5 , due to, e.g., insufficient reconstruction; otherwise, a cell is characterized. An expert who considered that a neuron was uncharacterized did not categorize it according to features C_1 – C_5 . **Figure 1** shows two interneurons characterized according to axonal features C_1 – C_5 .

2.3. PREDICTOR VARIABLES

We used 18 parameters of axonal morphology as predictor variables. Five of these parameters were computed with NeuroExplorer and were already used to predict interneuron types by DeFelipe et al. (2013) and Mihaljević et al. (2014b). In addition, we introduce 13 parameters of axonal morphology, seeking to capture the concepts represented by axonal features C_1 – C_4 . We computed these parameters from 3D interneuron reconstructions files in NeuroLucida's ASCII (*.asc) format.

The five parameters we computed with NeuroExplorer are:

- X_1 : 2D convex hull perimeter (in Z projection).
- X_2 : Axon length.
- X_3 : Axon length at less than 150 μm from the soma.
- X_4 : Axon length at more than 150 and less than 300 μm from the soma.
- X_5 : Axon length at more than 300 μm from the soma.

Parameters X_3 – X_5 are meant to measure axonal arborization with respect to the cortical column. Namely, parameter X_3 approximates arborization length within a (300 μm wide) cortical column (at less than 150 μm from the soma); X_4 approximates the length outside but not far from the column (more than 150 and less than 300 μm from the soma); and X_5 approximates axonal length far from the column (more than 300 μm from the soma). X_1 and X_2 were used by DeFelipe et al. (2013) while Mihaljević et al. (2014a) used X_3 – X_5 as predictor variables.

We introduce the following axonal morphometric parameters:

- X_6 : Axon length within soma's layer.
- X_7 : Axon length outside soma's layer.
- X_8 : Proportion of axon length contained within soma's layer, $\frac{X_6}{X_6+X_7}$.
- X_9 : Axon length within soma's cortical column.
- X_{10} : Axon length outside soma's cortical column.
- X_{11} : Proportion of axon length within soma's cortical column, $\frac{X_9}{X_9+X_{10}}$.
- X_{12} : Distance, in dimensions X and Y , from axon's centroid to the soma.
- X_{13} : Distance from the centroid of the above-the-soma part of the axon to the soma.
- X_{14} : Distance from the centroid of the below-the-soma part of the axon to the soma.
- X_{15} : Proportion of distances X_{13} and X_{14} , $\frac{X_{13}}{X_{13}+X_{14}}$.
- X_{16} : Axon length above the soma.
- X_{17} : Axon length below the soma.
- X_{18} : Proportion of axon length above soma, $\frac{X_{16}}{X_{16}+X_{17}}$.

We computed these parameters following assumptions made by DeFelipe et al. (2013), namely: (a) cortical layer thickness is (roughly) determined by species and cortical area (see following paragraphs for details); and (b) the cortical column is a cylinder whose axis passes through the soma and has a diameter of 300 μm . We measured the distance to soma as the distance to soma's centroid. We computed a centroid of a set of points (e.g., of all the points comprising the reconstructed axon) by averaging those points.

When computing parameters X_6 and X_7 we looked up the approximate layer thickness according to the neuron's species and cortical area. DeFelipe et al. (2013) defined an approximate layer thickness for every species/area/layer combination present in their data, and provided it as additional information for experts who classified the interneurons. This information can be accessed at <http://cajalbbp.cesvima.upm.es/gardenerclassification/>. DeFelipe et al. (2013) specified the approximate thickness in the form of an interval—e.g., stating that layer II/III of the mouse's

visual cortex is 200–300 μm thick—; we used the interval's midpoint (250 μm for the previous example) as an estimate of layer thickness. Also, we assumed that a soma is equidistant from the top and bottom confines of the layer (i.e., a 250 μm thick layer reaches 125 μm above and 125 μm below the soma).

For 16 mouse interneurons, seven of them from the somatosensory and nine from the visual cortex, the cortical layer was not provided. In order to compute variables X_6 and X_7 for these cells, we assumed them to belong to a hypothetical “average layer” for which we assumed a 197 μm thickness in the visual cortex and a 237 μm thickness in the somatosensory cortex. Although only an approximation, we consider this a more informed approximation to the “true” values of these variables than one that could be performed by a distance-computing rule (see Subsections 2.6 and 2.8) if we had left these values unspecified.

2.4. DATA SELECTION AND SUMMARY

Axonal feature C_6 is not a “proper” morphological feature but more of a “filter feature” which indicates whether the remaining axonal features can be reliably identified given a reconstructed interneuron. We therefore omitted C_6 from consideration in this paper. Consequently, we removed from our data set 11 interneurons considered as uncharacterized by a majority (i.e., at least 21) of neuroscientists, considering that these interneurons cannot be reliably classified according to C_1 – C_5 , thereby reducing our data sample to 226 interneurons.

Thus, we have $N = 226$ interneurons, each of them quantified by a vector \mathbf{X} of $m = 18$ real-valued predictor variables (i.e., $\mathbf{x} \in \mathbb{R}^{18}$). We also have $d = 5$ discrete class (i.e., target) variables $\mathbf{C} = (C_1, \dots, C_5)$, with $\mathbf{c} \in \Omega_{C_1} \times \dots \times \Omega_{C_5}$. Each interneuron, $\mathbf{x}^{(j)}$, is associated with a $N_j \times 5$ ($N_j \leq 42$) matrix $\mathcal{C}^{(j)}$ in which each row is an observation of \mathbf{C} due to one annotator neuroscientist, i.e., $\mathcal{C}_{i,a}^{(j)}$ is the label for class variable C_i assigned to interneuron $\mathbf{x}^{(j)}$ by expert neuroscientist a^3 .

Nonetheless, instead of the provided multi-annotator label matrices \mathcal{C} , we require each interneuron to be associated with an LBN in order to apply our method. We obtained LBNs using standard procedures for LBNs from data, and then learned and evaluated our model using these LBNs as input, omitting \mathcal{C} from further consideration (see the next subsection).

2.5. FROM MULTI-ANNOTATOR LABELS TO LABEL BAYESIAN NETWORKS

Prior to applying our method, we learned LBNs from multi-annotator class label matrices \mathcal{C} .

An LBN is a Bayesian network over the class variables \mathbf{C} . A Bayesian network (Pearl, 1988; Koller and Friedman, 2009) \mathcal{B} is a pair $\mathcal{B} = (\mathcal{G}, \Theta)$ where \mathcal{G} , the structure of the network, is a directed acyclic graph whose vertices correspond to the class variables \mathbf{C} and its arcs encode the conditional independencies in the joint distribution over \mathbf{C} , while Θ are the parameters of the

³Only neuroscientists who considered that $\mathbf{x}^{(j)}$ was characterized (axonal feature C_6) labeled $\mathbf{x}^{(j)}$ according to C_1 – C_5 . Therefore, N_j may be less than 42 and varies across interneurons.

conditional probability distributions that the joint distribution is factorized into.

Learning a Bayesian network \mathcal{B} from data consists in two steps: learning network structure, \mathcal{G} (i.e., the conditional independencies it encodes), and, having obtained the structure, learning its parameters (Neapolitan, 2004; Koller and Friedman, 2009). While the second step is generally straightforward, many methods exist for performing the first step. We applied a method belonging to the well-known family of search+score structure learning methods (see Subsection 2.8).

We wanted the learned LBNs to be similar to the actual empirical probabilities observed in the class label matrices, \mathcal{C} . In other words, we wanted the probability distribution factorized by an LBN, $p_{\mathcal{B}^{(j)}}$, to be similar to the empirical distribution, $p_{\mathcal{C}^{(j)}}$ —the relative frequency of each possible state of \mathbf{C} in $\mathcal{C}^{(j)}$. We used this similarity as criterion for selecting the network learning method

(see Subsection 2.8) and measured it with Jensen-Shannon divergence (see Subsection 2.9.1).

Finally, having learned the LBNs, our final data set was $\mathcal{D} = \{(\mathbf{x}^{(j)}, \mathcal{B}^{(j)})\}_{j=1}^N$. **Figure 2** depicts the LBNs for interneurons shown in **Figure 1**, along with the predicted LBNs for those interneurons.

2.6. MULTI-DIMENSIONAL CLASSIFICATION WITH LABEL BAYESIAN NETWORKS

Recall that we have m predictor variables \mathbf{X} , with $\mathbf{x} \in \mathbb{R}^m$, that describe the domain under study, and d discrete class (or target) variables \mathbf{C} , with $\mathbf{c} \in \Omega_{C_1} \times \dots \times \Omega_{C_d}$, that we wish to predict on the basis of observations of \mathbf{X} . We observe a data set, $\mathcal{D} = \{(\mathbf{x}^{(j)}, \mathcal{B}^{(j)})\}_{j=1}^N$, where \mathcal{B} is a label Bayesian network encoding a joint probability distribution over the multi-dimensional class variable \mathbf{C} .

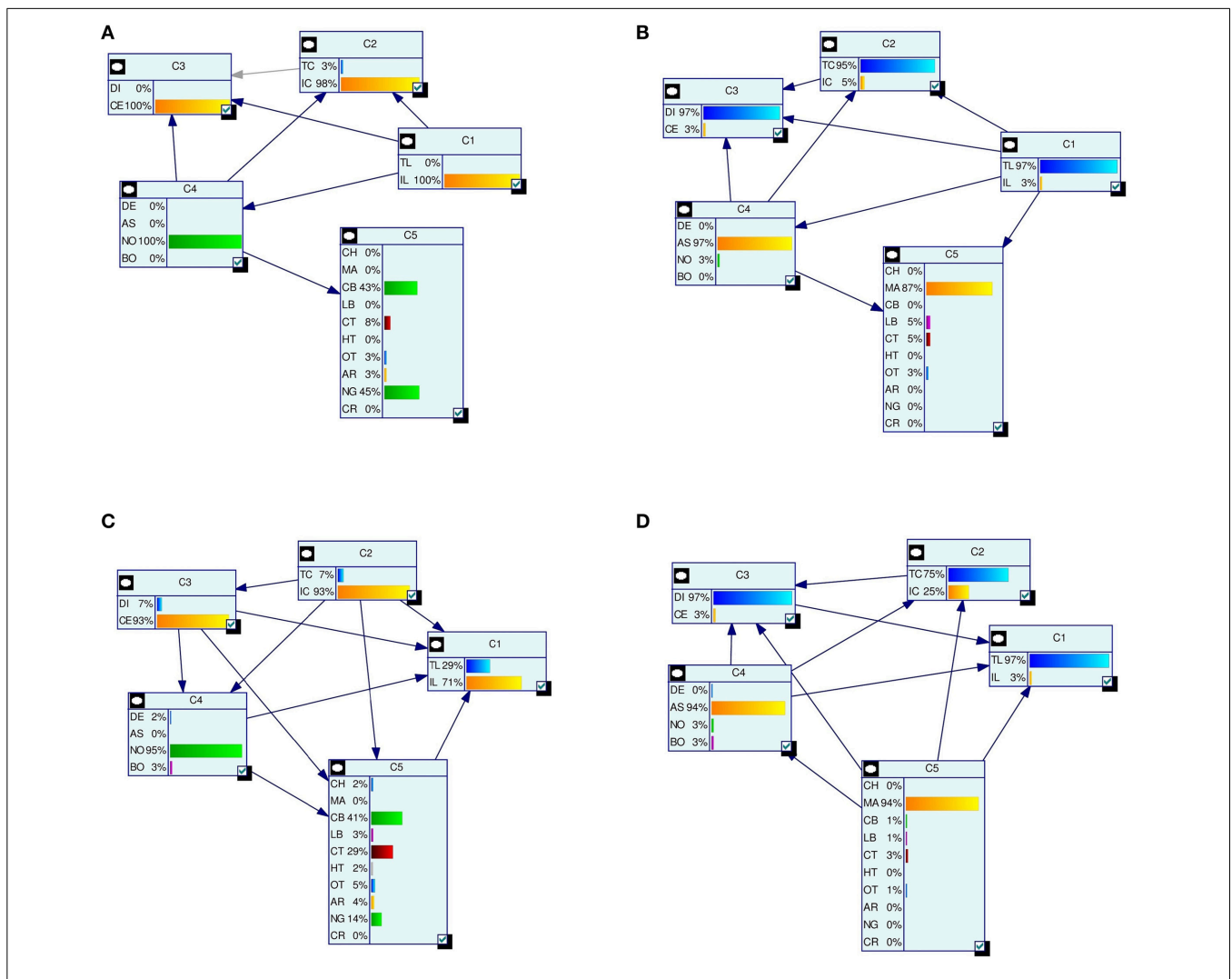
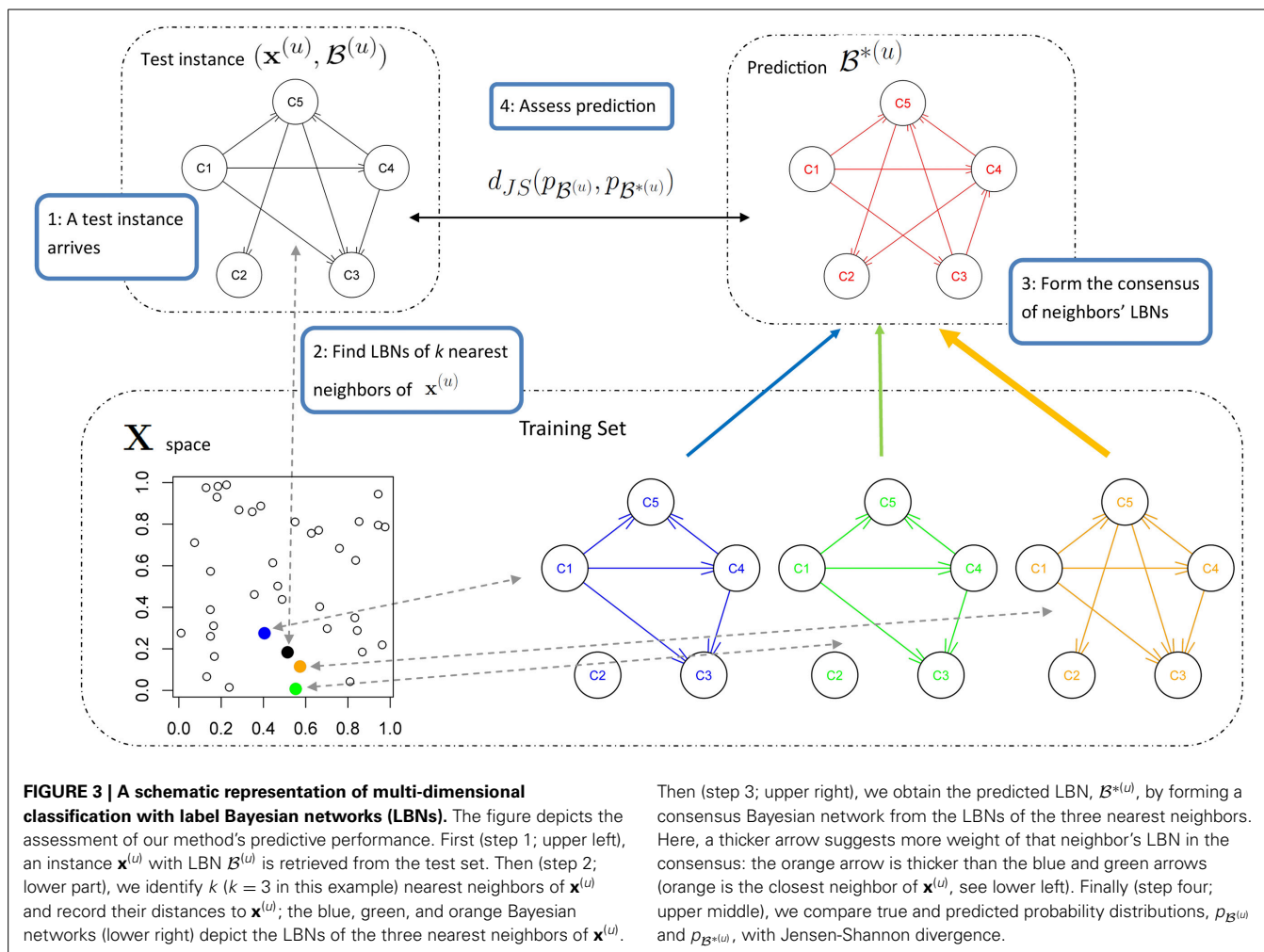


FIGURE 2 | Examples of true (A,B) and predicted (C,D) label Bayesian networks (LBNs) for neurons shown in Figure 1. The leftmost networks (A,C) correspond to interneuron (A) in Figure 1 whereas the right-hand ones (B,D) correspond to neuron (B) in Figure 1. The Bayesian networks are depicted with their nodes (shown as rectangles), arcs, and each node's

marginal probability distribution. The predicted distributions are similar to the true ones for many nodes—e.g., 93 vs. 98% for IC (node C₂) for interneuron (A). Some marginal probabilities do differ, such as that of the NG type for neuron (A)—14% predicted vs. 45% true; a lot of its probability mass was assigned to the more numerous CT type.



We predict the LBN of an unseen instance $\mathbf{x}^{(u)}$ by forming a consensus Bayesian network among the LBNs of its k nearest neighbors ($1 \leq k < N$) in the space of predictor variables. We form the consensus by adapting the method developed by Lopez-Cruz et al. (2014) to weigh the effect of each neighbor's LBN in proportion to that neighbor's relative closeness to $\mathbf{x}^{(u)}$. **Figure 3** summarizes our approach.

k nearest neighbors (k -nn; Fix and Hodges, 1989) is an instance-based (i.e., model-less) classifier, popular in uni-dimensional classification (Duda et al., 2000). It classifies a data instance $\mathbf{x}^{(u)}$ by identifying its k nearest neighbors in the predictor space, according to some distance measure—a common choice is the Euclidean distance—and choosing the majority from the neighbors' labels.

2.6.1. Distance-weighted consensus of Bayesian networks

Combining multiple Bayesian networks into a consensus Bayesian network is a recurring topic of interest. The standard methods for combining the parameters of a joint distribution, disregarding its underlying graphical structure (i.e., the conditional independencies), can yield undesirable results: for example, combining distributions with identical structures may render a consensus distribution with a different structure (Pennock and Wellman,

1999). It is therefore common to first combine network structures (e.g., Matzkevich and Abramson, 1992; Pennock and Wellman, 1999; Del Sagrado and Moral, 2003; Peña, 2011) and combine the parameters afterwards (e.g., Pennock and Wellman, 1999; Etminani et al., 2013). The cited structure-combining methods produce distributions which only contain independencies that are common to all networks, rendering them too complex (i.e., having too many parameters) to be useful in practice.

An alternative is to draw samples from the different Bayesian networks and learn the consensus network from the generated data, using standard methods for learning Bayesian networks from data (Neapolitan, 2004; Koller and Friedman, 2009), as proposed by Lopez-Cruz et al. (2014). Lopez-Cruz et al. (2014) weighted the influence of each Bayesian network on the consensus by sampling from it a number of instances proportional to its weight. We can readily adapt this method to weigh the effect of neighbors' label Bayesian networks in proportion to their closeness to the instance being classified, $\mathbf{x}^{(u)}$, by defining an appropriate weighting function. Before defining the weighting function, let us state the setting more formally.

We want to generate a database \mathcal{D}_u by sampling from k Bayesian networks $\{\mathcal{B}^{(j)}\}_{j=1}^k$ associated to k instances at distances d_1, \dots, d_k from the unseen instance $\mathbf{x}^{(u)}$; from this database, \mathcal{D}_u ,

we will learn the consensus Bayesian network, $\mathcal{B}^{*(u)}$. We want the number of samples in \mathcal{D}_u that are drawn from $\mathcal{B}^{(j)}$ to be proportional to the how close $\mathbf{x}^{(j)}$ is to $\mathbf{x}^{(u)}$. We measure closeness as relative to the remaining $k - 1$ neighboring instances. Thus, if M is the desired size of \mathcal{D}_u and $\mathbf{w} = (w_1, \dots, w_k)$ the weights assigned to the k Bayesian networks (with $\sum_{j=1}^k w_j = 1$ and $w \geq 0$), the number of samples from $\mathcal{B}^{(j)}$, M_j , is then $w_j \times M$. We compute the weights as

$$w_j = \frac{(\sum_{i=1}^k d_i) - d_j}{(k-1)(\sum_{i=1}^k d_i)}.$$

2.7. RELATED METHODS

2.7.1. Multiple annotators

A setting similar to ours, where many annotators provide class labels, occurs when *learning from a crowd* of annotators (Snow et al., 2008; Sorokin and Forsyth, 2008; Raykar et al., 2010; Welinder et al., 2010; Raykar and Yu, 2012). Yet, the crowd may include annotators of different skills and therefore learning a classifier involves estimating the ground truth label from the possibly noisy ones. Methods such as those due to Dawid and Skene (1979); Whitehill et al. (2009); Raykar et al. (2010); Welinder et al. (2010); Raykar and Yu (2012) aim to detect the less reliable annotators and decrease their influence on the ground truth estimate. In our case, however, all annotators are domain experts; furthermore, there is currently no better approximation to ground truth than the opinions of this group of leading experts, as there is no unequivocal or objective way of determining it⁴. We thus consider that every expert's opinion is equally valid and that interneuron type membership and axonal features are uncertain whenever the experts do not completely agree. This allows us to represent interneuron type membership and axonal features of an interneuron with a joint probability distribution over these five class variables.

2.7.2. Probabilistic labels

Probabilistic labels have already been used in machine learning (Ambrose et al., 2001; Grandvallet, 2002; Thiel et al., 2007; Schwenker and Trentin, 2014). Some methods consider these to be imprecise versions of a crisp (i.e., non-probabilistic) ground truth label, which they then try to estimate, while others (Thiel et al., 2007; Schwenker and Trentin, 2014), more in line with our setting, assume that probabilistic labels represent intrinsic ambiguity in class membership and consider them as ground truth. Methods such as k -nn (El Gayar et al., 2006) and support vector machines (Thiel et al., 2007; Scherer et al., 2013) have been adapted to deal with probabilistic labels, while regression-based methods, such as multi-layer perceptrons, can handle them without being adapted (Schwenker and Trentin, 2014). Yet, all of these methods are aimed at predicting a single class variable.

⁴Possibly, the ground truth might be better approximated by consulting more leading neuroscientists. Our group, nonetheless, includes many of the leading experts involved in the interneuron nomenclature effort (Ascoli et al., 2007).

2.7.3. Multi-dimensional classification

Multi-dimensional classification is more general than the related multi-label classification, which has already been considered in neuroscience (Turner et al., 2013). It is hard because the number of possible assignments to the class variables is exponential in their number. Predicting each class variable with an independent model is suboptimal because the variables are, generally, correlated. Modeling many of these dependencies, on the other hand, can lead to data scarcity. Multi-dimensional Bayesian network classifiers (Bielza et al., 2011; Borchani et al., 2013) can balance model complexity and the modeling of dependencies. However, they require crisp class labels in order to be trained and thus cannot be directly applied to our setting.

2.7.4. k -nearest neighbors

k nearest neighbors is a popular instance-based (i.e., model-less) classifier. Among other extensions, the original k -nn classifier has been adapted to weight the effect of a neighbor's class label in proportion to how close that neighbor is to the data instance being classified (e.g., Dudani, 1976; MacLeod et al., 1987; Denoex, 1995; Yazdani et al., 2009). It has also been adapted to deal with non-crisp labels (Jóźwik, 1983; Keller et al., 1985; Denoex, 1995); these non-crisp labels, however, are not probabilistic but possibilistic (encoded with Dempster-Shafer theory) and fuzzy. Besides the handling of non-crisp labels, methods due to Denoex (1995) and Keller et al. (1985) are similar to ours in that they weigh the neighbors' effect on prediction according to their closeness to the data instance being classified. On the other hand, they differ from our method in neither using probabilistic labels nor tackling the prediction of multiple class variables.

2.8. EXPERIMENTAL SETTING

We identified the nearest interneurons by measuring Euclidean distance. Thus, for a pair of interneurons \mathbf{x}^j and \mathbf{x}^o , the distance d_{jo} is given by

$$d_{jo} = \left(\sum_{i=1}^m (x_i^{(j)} - x_i^{(o)})^2 \right)^{\frac{1}{2}}.$$

Prior to computing distances, we standardized all predictor variables X_1, \dots, X_m (i.e., for each X_i , we subtracted its mean and divided by standard deviation).

We drew samples from the neighboring networks using probabilistic logic sampling (Henrion, 1986). We sought to draw enough samples from each distribution so to represent it correctly. We therefore set M , the total number of samples drawn from the k nearest neighbors' distributions (see Section 2.6.1), as $k * 500 * c$, where c was the maximal number of free parameters among the k networks whose consensus is being sought. The number of free parameters of a Bayesian network is the number of parameters that suffice to fully specify the network's probability distribution (recall that a network consists of a structure, \mathcal{G} , and parameters Θ ; see subsection 2.5).

Once we had generated the data set of sample points, we applied a Bayesian network learning algorithm to obtain the consensus probability distribution.

2.8.1. Learning Bayesian networks from data

There were two instances in which we learned Bayesian networks from data: when learning LBNs from expert-provided class label matrices (see Subsection 2.5) and when learning the consensus network from sampled data points (Subsection 2.6.1). We considered three options for the learning procedure and chose the one that we considered most adequate for learning LBNs, according to the criterion described in Subsection 2.5. We then applied this chosen procedure in both instances of network learning.

The Bayesian network learning procedure we used follows the well-known search+score approach. Such a procedure consists of (a) a search procedure for traversing the space of possible network structures and (b) a scoring function. We searched the structure space with the tabu metaheuristic (Glover, 1989, 1990), a local search procedure which employs adaptive memory to improve efficiency and escape local minima, and considered three networks scores: Bayesian Information Criterion (BIC; Schwarz, 1978), K2 (Cooper and Herskovits, 1992) and Bayesian Dirichlet equivalence (BDe; Heckerman et al., 1995). We compared the LBNs produced by the different scores according to how well they approximated the empirical distributions, p_{ϵ} , (see Subsection 2.5) and their complexity (i.e., number of free parameters).

We estimated parameters by maximum likelihood estimation.

2.8.2. Software and assessment

We implemented the computation of the 13 here introduced axonal morphometric parameters from scratch. We performed Bayesian network learning and sampling with the `bnlearn` (Scutari, 2010; Nagarajan et al., 2013) package for the R statistical software environment (R Core Team, 2014).

In traditional uni-dimensional classification, it is common to perform stratified cross-validation, that is, to have similar class proportions in train and test sets. However, such stratification is problematic in the multi-dimensional setting, due to the high number of combinations of class variables. Therefore, instead of stratified cross-validation, we evaluated our model with 20 repetitions of plain (unstratified) 10-fold cross-validation.

2.9. ASSESSING RESULTS

We were primarily interested in predicting LBNs. We assessed this prediction with Jensen-Shannon divergence, a metric which we describe below.

However, for comparison with related work on interneuron classification, we also assessed how well our method predicted crisp (i.e., non-probabilistic) labels. Such an evaluation is negatively biased against our method since we take label ambiguity into account to learn the model while it is evaluated as though a true crisp label existed (i.e., as if there was no ambiguity). Below we describe how we obtained crisp labels and present accuracy metrics for multi-dimensional classification.

2.9.1. Comparing probability distributions

We measured the dissimilarity between two probability distributions, say $p_{\mathcal{B}^{(u)}}$ and $p_{\mathcal{B}^{*(u)}}$, with Jensen-Shannon divergence,

$$d_{JS}(p_{\mathcal{B}^{(u)}}, p_{\mathcal{B}^{*(u)}}) = \frac{1}{2}(d_{KL}(p_{\mathcal{B}^{(u)}}, p_r) + d_{KL}(p_{\mathcal{B}^{*(u)}}, p_r)),$$

where $p_r = \frac{1}{2}(p_{\mathcal{B}^{(u)}} + p_{\mathcal{B}^{*(u)}})$ and $d_{KL}(p_{\mathcal{B}^{(u)}}, p_{\mathcal{B}^{*(u)}})$ is the Kullback-Leibler divergence (Kullback and Leibler, 1951) between $p_{\mathcal{B}^{(u)}}$ and $p_{\mathcal{B}^{*(u)}}$,

$$d_{KL}(p_{\mathcal{B}^{(u)}}, p_{\mathcal{B}^{*(u)}}) = \sum_{\mathbf{c} \in \Omega_{\mathbf{c}}} p_{\mathcal{B}^{(u)}}(\mathbf{c}) \log \left(\frac{p_{\mathcal{B}^{(u)}}(\mathbf{c})}{p_{\mathcal{B}^{*(u)}}(\mathbf{c})} \right).$$

Unlike Kullback-Leibler divergence, Jensen-Shannon divergence is symmetric, it does not require absolute continuity (i.e., that $p_{\mathcal{B}^{*(u)}}(\mathbf{c}) = 0 \implies p_{\mathcal{B}^{(u)}}(\mathbf{c}) = 0$), its square root is a metric, and it is bounded: $0 \leq d_{JS} \leq 1$.

2.9.2. Obtaining crisp labels

In order to assess the prediction of crisp labels, we needed to obtain a “true” crisp class label vector for each interneuron $\mathbf{x}^{(j)}$. We assumed that such “true” labels were given by the choice of the majority of the experts. There were two alternative majority choices: (a) the most commonly selected class label vector, i.e., the most common row in a class labels matrix \mathcal{C} ; and (b) the concatenation of per-class majority labels, i.e., the vector formed by the most common choice for C_1 , the most common choice for C_2 , and so on, until C_5 . We refer to the former as the *joint truth* and to the latter as *marginal truth*; the latter was used in related works on interneuron classification (DeFelipe et al., 2013; Mihaljević et al., 2014a,b) since they predicted the axonal features C_1 – C_5 independently. We compared our predicted crisp labels to both “truths.”

We also needed to extract crisp predictions from a predicted LBNs. The two straightforward methods are analogous to the above-described ones: (a) choosing the *most probable explanation* (MPE), i.e., the most likely joint assignment to \mathbf{C} according to LBN \mathcal{B}^* ; and (b) concatenating the marginally most likely assignments to each of the class variables. For simplicity, we only used the MPE as the predicted crisp class labels vector.

2.9.3. Multi-dimensional classification accuracy metrics

We assessed crisp labels prediction with accuracy metrics for multi-dimensional classification (Bielza et al., 2011):

- The *mean accuracy* over d ($d = 5$ in our case) class variables:

$$\overline{Acc} = \frac{1}{d} \sum_{l=1}^d \frac{1}{N} \sum_{u=1}^N \delta(c_l^{*(u)}, c_l^{(u)}),$$

where $c_l^{*(u)}$ is the predicted value of C_l for u -th instance, $c_l^{(u)}$ is the corresponding true value, and $\delta(a, b) = 1$ when $a = b$ and 0 otherwise.

- The *global accuracy* over d class variables:

$$Acc = \frac{1}{N} \sum_{u=1}^N \delta(\mathbf{c}^{*(u)}, \mathbf{c}^{(u)}).$$

Note that global accuracy is demanding as it only rewards full matches between the predicted vector and the true one. We

also measured uni-dimensional *marginal accuracy* per each class variable,

$$Acc_l = \frac{1}{N} \sum_{u=1}^N \delta(c_l^{*(u)}, c_l^{(u)}).$$

When computing global and mean accuracy, we used the “joint truth” crisp labels. When computing per-class-variable marginal accuracy, we used the “marginal truth” crisp labels vector.

3. RESULTS

3.1. FROM MULTI-ANNOTATOR LABELS TO LABEL BAYESIAN NETWORKS

We first studied whether any network score was particularly adequate for transforming multi-expert labels into LBNs. Different scores yielded networks of different degrees of complexity but were all good at approximating of the empirical probability distribution over the expert-provided labels, p_ϵ (see **Table 1**). We used the score that yielded the best approximation, BDe, in the remainder of this paper. Namely, we used it to (a) transform multi-expert labels into LBNs; and (b) learn a consensus networks from the generated samples.

3.2. PREDICTING LABEL BAYESIAN NETWORKS

We considered four different values of k (the number of nearest neighbors)—namely, 3, 5, 7, and 9—and obtained best results with $k \in \{5, 7\}$. As **Table 2** shows, we predicted the label Bayesian networks relatively accurately, with a Jensen-Shannon divergence of 0.29 for $k \in \{5, 7\}$.

Table 1 | Transforming multi-expert labels into label Bayesian networks using different network scores.

	BIC	K2	BDe
JS divergence	00.10 ± 0.05	00.07 ± 0.04	00.06 ± 00.04
Free parameters	18.22 ± 1.83	31.08 ± 20.58	60.34 ± 31.14

Upper row: average Jensen-Shannon (JS) divergence between the empirical probability distribution over the labels, p_ϵ , and the one encoded by the learned Bayesian network labels, p_B ; lower row: average number of free parameters per learned network. Averaged across entire data set.

Table 2 | Predicting label Bayesian networks and crisp labels.

	JS	Global acc. (%)	Mean acc. (%)
$k = 3$	0.30 ± 0.00	41.29 ± 1.57	79.10 ± 0.74
$k = 5$	0.29 ± 0.00	43.84 ± 1.48	79.52 ± 0.79
$k = 7$	0.29 ± 0.00	43.99 ± 1.26	79.88 ± 0.34
$k = 9$	0.30 ± 0.00	39.46 ± 1.67	78.58 ± 0.52

The leftmost column shows Jensen-Shannon (JS) divergence between predicted label Bayesian networks, p_{B^*} , and label Bayesian networks p_B learned from C . Rightmost columns show global and mean accuracy for predicting joint truth crisp labels vector, i.e., the class label vector most often selected by the experts. Obtained with 20 runs of 10-fold cross-validation.

Figure 2 depicts the true and predicted LBNs for two interneurons, one having barely ambiguous axonal features and another having an ambiguous type; as the figure suggests, the LBN of the former interneuron was accurately predicted, while in that of the latter, the type (C_5) marginal probability was predicted only moderately well.

3.3. PREDICTING CRISP LABELS

We predicted the joint truth (the class label vectors selected by a majority of experts; see Section 2.9.2) relatively accurately—with a mean accuracy of 80% and global accuracy of 44% for $k \in \{5, 7\}$ (see **Table 2**). The latter result means that 44% of the MPEs of the predicted LBNs (B^*) were equivalent to the joint truth vectors.

We also assessed the marginal accuracy for each axonal feature C_1 – C_5 . Here we compared the B^* MPE with the marginal truth, class variable by class variable. We predicted features C_1 – C_4 with over 80% accuracy—up to 88% in case of C_1 —and feature C_5 with 64% accuracy with $k = 7$ (see **Table 3**). Albeit it may seem low, the latter result is better than chance. Namely, DeFelipe et al. (2013) showed that even 40.25% accuracy for C_5 —obtained by a classifier they used—was better than chance. It should also be recalled that the ten neuronal types were often hard to distinguish for expert neuroscientists (DeFelipe et al., 2013). Regarding the prediction of the individual types, accurately predicted ones included the MA and HT types, which were easy to identify for the experts, and the numerous but less clear to the experts types such as CB and LB. The least clear out of the numerous types, CT, was predicted with relatively low accuracy (see **Table 4**).

4. DISCUSSION

Previous studies on interneuron classification (DeFelipe et al., 2013; Mihaljević et al., 2014a,b) used majority crisp labels, estimated for each axonal feature independently, to train and evaluate their models. Mihaljević et al. (2014a) only considered C_5 whereas DeFelipe et al. (2013) and Mihaljević et al. (2014a) predicted axonal features C_1 – C_5 with an independent model for each of them. There were non-methodological differences among these two studies and the present work and therefore any comparison of results ought to be performed with some caution. DeFelipe et al. (2013), for example, used 15 cells more than we did (see Section 2.4), had several of variables’ values corrupted by imperfections in the reconstructions of 36 cells—which we corrected—and used only three values for C_4 —ascending, descending, and both. Furthermore, they used different morphometric predictor parameters (over 2000 of them), and applied a possibly

Table 3 | Accuracy (in %) for each of the five axonal features C_1 – C_5 .

	C_1	C_2	C_3	C_4	C_5
$k = 3$	86.15 ± 1.12	83.17 ± 0.98	86.50 ± 0.88	83.11 ± 0.90	62.69 ± 1.24
$k = 5$	86.49 ± 0.98	83.25 ± 0.79	86.05 ± 0.79	84.18 ± 0.65	63.78 ± 1.11
$k = 7$	88.07 ± 1.01	83.12 ± 0.72	85.29 ± 0.55	84.06 ± 0.74	64.33 ± 1.52
$k = 9$	87.16 ± 1.03	83.06 ± 0.78	85.39 ± 0.78	83.88 ± 0.71	63.79 ± 1.59

Here we compared the marginal true labels to the most probable explanation of the predicted label Bayesian networks. Obtained with 20 runs of 10-fold cross-validation.

Table 4 | Confusion matrix for predicting C_5 with $k = 7$.

	CB	CH	CT	HT	LB	MA	NG	Per-type sensitivity
CB	41	0	10	0	6	3	0	0.68
CH	2	0	1	0	0	0	0	0.00
CT	10	0	25	4	8	11	0	0.43
HT	0	0	3	10	1	0	0	0.71
LB	9	0	3	0	25	3	1	0.61
MA	0	0	2	0	4	36	0	0.86
NG	8	0	0	0	0	0	0	0.00

Here we compared the marginal true label for C_5 (rows) to the C_5 value of most probable explanation of the predicted label Bayesian network (columns). The rightmost column shows per-type sensitivity. Types AR, and CR and OT are omitted since no cell's crisp label was of one of these types. Obtained from a single run of 10-fold cross-validation.

Table 5 | Our best predictive accuracy (in %) vs. best accuracy from DeFelipe et al. (2013), for each of the axonal features C_1 – C_5 .

	C_1	C_2	C_3	C_4	C_5
Present study	88.07	83.25	86.50	84.18	64.33
DeFelipe et al., 2013	85.48	81.33	73.86	60.17	62.24

Our results were obtained with 20 runs of 10-fold cross-validation; those of DeFelipe et al. (2013) were obtained with leave-one-out cross-validation.

more optimistic accuracy estimation technique—leave-one-out estimation. Mihaljević et al. (2014a) considered multiple subsets of the data, formed according to the degree of class label ambiguity of the included cells, and obtained best results with least ambiguous cells (e.g., with 46 cells for C_5). Their best results were thus obtained with a small subset of the 226 cells that we used. When using most of the cells, their results were similar to those of DeFelipe et al. (2013).

Differences aside, in **Table 5** we compare the accuracies from the present study with those from DeFelipe et al. (2013). We outperformed DeFelipe et al. (2013) in predictive accuracy for every axonal feature, even though we used a single model to predict all features simultaneously. We especially outperformed their approach in predicting C_3 and, even more, in predicting C_4 . The latter was likely affected by the use of the additional category no (see subsection 2.2).

Despite the non-methodological differences with the study by DeFelipe et al. (2013), the better accuracies that we achieved might suggest some or all of the following: (a) the introduced morphometric parameters are useful for predicting interneuron type and axonal features; (b) we adequately assigned the value no for cells to which the other values of C_4 did not apply; and (c) our method is adequate for classifying interneurons.

The above results, along with the relatively high global accuracy achieved, 44%, suggest that axonal features C_1 – C_5 are interrelated and that it is useful to attempt predicting them simultaneously.

Finally, several other efforts regarding classification of neurons in general have been performed taking into account

other morphological and/or molecular and/or electrophysiological properties (e.g., Bota and Swanson, 2007; Ascoli et al., 2009; Brown and Hestrin, 2009; Battaglia et al., 2013; Sümbül et al., 2014). These studies indicate that in spite of a large diversity of neuronal types, certain clear correlations exist between the axonal features and dendritic morphologies, and between these anatomical characteristics and some molecular and electrical attributes. Nevertheless, the classification of neurons is still under intense study from different angles, including anatomical, physiological, and molecular criteria, and using a variety of mathematical approaches, such as hierarchical clustering (Cauli et al., 2000; Wang et al., 2002; Tsiola et al., 2003; Benavides-Piccione et al., 2006; Dumitriu et al., 2007; Helmstaedter et al., 2009a,b), k -means (e.g., Karagiannis et al., 2009, affinity propagation (Santana et al., 2013), linear discriminant analysis (Marin et al., 2002; Druckmann et al., 2013), Bayesian network classifiers (Mihaljević et al., 2014a), and semi-supervised model-based clustering (Mihaljević et al., 2014b).

4.1. COMPUTING AXONAL MORPHOMETRIC PARAMETERS

In order to compute some of the newly introduced axonal morphometric parameters—namely, those relative to laminar and cortical distribution—we followed a series of assumptions originating from DeFelipe et al. (2013). These assumptions (simplifications) should be kept in mind when interpreting our results. First, we assumed arbitrary columnar and laminar demarcations. Namely, we considered the diameter of the hypothetical cortical column to be 300 μm (Malach, 1994; Mountcastle, 1998), whereas laminar thickness was estimated for each neuron from its original paper, when such a paper was available, and from relevant literature otherwise. Finally, we assumed that a soma is always located in the center of its layer and cortical column.

5. CONCLUSION

We built a model that can automatically classify an interneuron, on the basis of a set of its axonal morphometric parameters, according to five properties which constitute the pragmatic classification scheme proposed by DeFelipe et al. (2013), namely, the interneuron type and four other categorical axonal features. We guided model construction with a Bayesian network-encoded probability distribution indicating the type and axonal features of each interneuron. We obtained these probability distributions from classification choices provided by a group of leading neuroscientists. We then developed an instance-based supervised classifier which could learn from such multi-dimensional probabilistic input, predicting the output by forming a consensus among a set of Bayesian networks.

We accurately predicted the probabilistic labels over the interneuron type and the four remaining axonal features. Furthermore, we outperformed previous work when predicting crisp (i.e., non-probabilistic) labels. Importantly, and unlike previous work, we predicted the five axonal features simultaneously (i.e., with a single model), which is useful since these features are complementary. Our results suggest that interneuron type and the and four remaining axonal features are related and that it is useful to predict them jointly.

We introduced 13 axonal morphometric parameters which we defined as quantitative counterparts of the four categorical axonal features. Our results suggest that these parameters are useful for predicting the type and the four axonal features. Thus, they might be considered as objective replacements, or surrogates, of the subjective categorical axonal features.

This paper demonstrates a useful application of Bayesian networks in neuroscience, whose potential has been largely unexploited in this field (one exception is functional connectivity analysis from neuroimaging data; see Bielza and Larrañaga, 2014).

It would be interesting to relax the assumption that all neuroscientists who classified our data are equally accurate at classifying all types of interneurons, since some may be more familiar with certain interneuron types than with others, and account for expert competence in our model, similarly to methods for learning from a crowd of annotators such as Raykar et al. (2010) and Welinder et al. (2010).

We also intend to consider new methods for forming a consensus among Bayesian networks.

5.1. DATA SHARING

The data set and the software reproducing our study are available online, at <http://cig.fi.upm.es/bojan/gardener/>.

AUTHOR CONTRIBUTIONS

Bojan Mihaljević, Concha Bielza, and Pedro Larrañaga designed the method and the empirical study. Ruth Benavides-Piccione corrected the faulty interneuron reconstructions. Ruth Benavides-Piccione and Bojan Mihaljević defined the here introduced morphological variables. Bojan Mihaljević performed the data analysis, implementing necessary software, and wrote the paper. Concha Bielza, Ruth Benavides-Piccione, Javier DeFelipe and Pedro Larrañaga critically revised the paper.

ACKNOWLEDGMENTS

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness through the Cajal Blue Brain (C080020-09; the Spanish partner of the Blue Brain initiative from EPFL) and TIN2013-41592-P projects, by the S2013/ICE-2845-CASI-CAM-CM project, and the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 604102 (Human Brain Project). The authors thank Pedro L. López-Cruz and Luis Guerra for useful comments regarding the definition of the here introduced morphometric parameters.

REFERENCES

- Ambrose, C., Denoeux, T., Govaert, G., and Smets, P. (2001). "Learning from an imprecise teacher: probabilistic and evidential approaches," in *10th International Symposium on Applied Stochastic Models and Data Analysis*, Vol. 1 (Compiègne), 101–105.
- Ascoli, G. A., Brown, K. M., Calixto, E., Card, J. P., Galvan, E., Perez-Rosello, T., et al. (2009). Quantitative morphometry of electrophysiologically identified CA3b interneurons reveals robust local geometry and distinct cell classes. *J. Comp. Neurol.* 515, 677–695. doi: 10.1002/cne.22082
- Ascoli, G. A., Donohue, D. E., and Halavi, M. (2007). Neuromorpho.org: a central resource for neuronal morphologies. *J. Neurosci.* 27, 9247–9251. doi: 10.1523/JNEUROSCI.2055-07.2007
- Battaglia, D., Karagiannis, A., Gallopin, T., Gutch, H. W., and Cauli, B. (2013). Beyond the frontiers of neuronal types. *Front. Neural Circuits* 7:13. doi: 10.3389/fncir.2013.00013
- Benavides-Piccione, R., Hamzei-Sichani, F., Ballesteros-Yáñez, I., DeFelipe, J., and Yuste, R. (2006). Dendritic size of pyramidal neurons differs among mouse cortical regions. *Cereb. Cortex* 16, 990–1001. doi: 10.1093/cercor/bhj041
- Bielza, C., and Larrañaga, P. (2014). Bayesian networks in neuroscience: a survey. *Front. Comput. Neurosci.* 8:131. doi: 10.3389/fncom.2014.00131
- Bielza, C., Li, G., and Larrañaga, P. (2011). Multi-dimensional classification with Bayesian networks. *Int. J. Approx. Reason.* 52, 705–727. doi: 10.1016/j.ijar.2011.01.007
- Borchani, H., Bielza, C., Toro, C., and Larrañaga, P. (2013). Predicting human immunodeficiency virus inhibitors using multi-dimensional Bayesian network classifiers. *Artif. Intell. Med.* 57, 219–229. doi: 10.1016/j.artmed.2012.12.005
- Bota, M., and Swanson, L. W. (2007). The neuron classification problem. *Brain Res. Rev.* 56, 79–88. doi: 10.1016/j.brainresrev.2007.05.005
- Brown, S. P., and Hestrin, S. (2009). Cell-type identity: a key to unlocking the function of neocortical circuits. *Curr. Opin. Neurobiol.* 19, 415–421. doi: 10.1016/j.conb.2009.07.011
- Cauli, B., Audinat, E., Lambolez, B., Angulo, M. C., Ropert, N., Tsuzuki, K., et al. (1997). Molecular and physiological diversity of cortical nonpyramidal cells. *J. Neurosci.* 17, 3894–3906.
- Cauli, B., Porter, J. T., Tsuzuki, K., Lambolez, B., Rossier, J., Quenet, B., et al. (2000). Classification of fusiform neocortical interneurons based on unsupervised clustering. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6144–6149. doi: 10.1073/pnas.97.11.6144
- Cooper, G. F., and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Mach. Learn.* 9, 309–347. doi: 10.1007/BF00994110
- Dawid, A. P., and Skene, A. M. (1979). Maximum likelihood estimation of observer error-rates using the EM algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.* 28, 20–28. doi: 10.2307/2346806
- DeFelipe, J., López-Cruz, P. L., Benavides-Piccione, R., Bielza, C., Larrañaga, P., Anderson, S., et al. (2013). New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat. Rev. Neurosci.* 14, 202–216. doi: 10.1038/nrn3444
- Del Sagrado, J., and Moral, S. (2003). Qualitative combination of Bayesian networks. *Int. J. Intell. Syst.* 18, 237–249. doi: 10.1002/int.10086
- Denoeux, T. (1995). A k-nearest neighbor classification rule based on Dempster-Shafer theory. *Syst. Man Cybern. IEEE Trans.* 25, 804–813. doi: 10.1109/21.376493
- Druckmann, S., Hill, S., Schürmann, F., Markram, H., and Segev, I. (2013). A hierarchical structure of cortical interneuron electrical diversity revealed by automated statistical analysis. *Cereb. Cortex* 23, 2994–3006. doi: 10.1093/cercor/bhs290
- Duda, R. O., Hart, P. E., and Stork, D. G. (2000). *Pattern Classification, 2nd Edn.* New York, NY: Wiley-Interscience.
- Dudani, S. A. (1976). The distance-weighted k-nearest-neighbor rule. *IEEE Trans. Syst. Man Cybern.* SMC-6, 325–327. doi: 10.1109/TSMC.1976.5408784
- Dumitriu, D., Cossart, R., Huang, J., and Yuste, R. (2007). Correlation between axonal morphologies and synaptic input kinetics of interneurons from mouse visual cortex. *Cereb. Cortex* 17, 81–91. doi: 10.1093/cercor/bhj126
- El Gayar, N., Schwenker, F., and Palm, G. (2006). "A study of the robustness of knn classifiers trained using soft labels," in *Artificial Neural Networks in Pattern Recognition, Volume 4087 of em Lecture Notes in Computer Science*, eds F. Schwenker and S. Marinai (Berlin, Heidelberg: Springer), 67–80.
- Etmiani, K., Naghibzadeh, M., and Peña, J. M. (2013). DemocraticOP: a democratic way of aggregating Bayesian network parameters. *Int. J. Approx. Reason.* 54, 602–614. doi: 10.1016/j.ijar.2012.12.002
- Fairén, A., Regidor, J., and Kruger, L. (1992). The cerebral cortex of the mouse (a first contribution - the 'acoustic' cortex). *Somatosens. Mot. Res.* 9, 3–36. doi: 10.3109/08990229209144760
- Fix, E., and Hodges, J. L. (1989). Discriminatory analysis. Nonparametric discrimination: consistency properties. *Int. Statist. Rev.* 57, 238–247.
- Glaser, J. R., and Glaser, E. M. (1990). Neuron imaging with NeuroLucida — A PC-based system for image combining microscopy. *Comput. Med. Imaging Graph.* 14, 307–317. doi: 10.1016/0895-6111(90)90105-K
- Glover, F. (1989). Tabu search-part I. *ORSA J. Comput.* 1, 190–206. doi: 10.1287/ijoc.1.3.190
- Glover, F. (1990). Tabu searchpart II. *ORSA J. Comput.* 2, 4–32. doi: 10.1287/ijoc.2.1.4

- Grandvalet, Y. (2002). "Logistic regression for partial labels," in *9th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IMPUS'02)*, Vol. 3 (Annecy), 1935–1941.
- Guerra, L., McGarry, L. M., Robles, V., Bielza, C., Larrañaga, P., and Yuste, R. (2011). Comparison between supervised and unsupervised classifications of neuronal cell types: a case study. *Dev. Neurobiol.* 71, 71–82. doi: 10.1002/dneu.20809
- Gupta, A., Wang, Y., and Markram, H. (2000). Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science* 287, 273–278. doi: 10.1126/science.287.5451.273
- Heckerman, D., Geiger, D., and Chickering, D. M. (1995). Learning Bayesian networks: the combination of knowledge and statistical data. *Mach. Learn.* 20, 197–243. doi: 10.1007/BF00994016
- Helmstaedter, M., Sakmann, B., and Feldmeyer, D. (2009a). L2/3 interneuron groups defined by multiparameter analysis of axonal projection, dendritic geometry, and electrical excitability. *Cereb. Cortex* 19, 951–962. doi: 10.1093/cercor/bhn130
- Helmstaedter, M., Sakmann, B., and Feldmeyer, D. (2009b). The relation between dendritic geometry, electrical excitability, and axonal projections of L2/3 interneurons in rat barrel cortex. *Cereb. Cortex* 19, 938–950. doi: 10.1093/cercor/bhn138
- Henrion, M. (1986). "Propagating uncertainty in Bayesian networks by probabilistic logic sampling," in *Second Annual Conference on Uncertainty in Artificial Intelligence, UAI '86* (Philadelphia, PA), 149–164.
- Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Patt. Recogn. Lett.* 31, 651–666. doi: 10.1016/j.patrec.2009.09.011
- Jóźwik, A. (1983). A learning scheme for a fuzzy k-NN rule. *Patt. Recogn. Lett.* 1, 287–289. doi: 10.1016/0167-8655(83)90064-8
- Karagiannis, A., Gallopin, T., Dávid, C., Battaglia, D., Geoffroy, H., Rossier, J., et al. (2009). Classification of NPY-expressing neocortical interneurons. *J. Neurosci.* 29, 3642–3659. doi: 10.1523/JNEUROSCI.0058-09.2009
- Kawaguchi, Y. (1993). Physiological, morphological, and histochemical characterization of three classes of interneurons in rat neostriatum. *J. Neurosci.* 13, 4908–4923.
- Keller, J. M., Gray, M. R., and Givens, J. A. (1985). A fuzzy k-nearest neighbor algorithm. *IEEE Trans. Syst. Man Cybern.* SMC-15, 580–585. doi: 10.1109/TSMC.1985.6313426
- Koller, D., and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA: MIT press.
- Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* 22, 79–86. doi: 10.1214/aoms/1177729694
- Lopez-Cruz, P., Larrañaga, P., DeFelipe, J., and Bielza, C. (2014). Bayesian network modeling of the consensus between experts: an application to neuron classification. *Int. J. Approx. Reason.* 55, 3–22. doi: 10.1016/j.ijar.2013.03.011
- Maccaferri, G., and Lacaille, J.-C. (2003). Interneuron diversity series: hippocampal interneuron classifications—making things as simple as possible, not simpler. *Trends Neurosci.* 26, 564–571. doi: 10.1016/j.tins.2003.08.002
- MacLeod, J. E., Luk, A., and Titterton, D. (1987). A re-examination of the distance-weighted k-nearest neighbor classification rule. *IEEE Trans. Syst. Man Cybern.* 17, 689–696. doi: 10.1109/TSMC.1987.289362
- Malach, R. (1994). Cortical columns as devices for maximizing neuronal diversity. *Trends Neurosci.* 17, 101–104. doi: 10.1016/0166-2236(94)90113-9
- Marin, E. C., Jefferis, G. S., Komiyama, T., Zhu, H., and Luo, L. (2002). Representation of the glomerular olfactory map in the drosophila brain. *Cell* 109, 243–255. doi: 10.1016/S0092-8674(02)00700-6
- Matzkevič, I., and Abramson, B. (1992). "The topological fusion of Bayes nets," in *Proceedings of the Eighth international conference on Uncertainty in artificial intelligence* (Stanford, CA: Morgan Kaufmann Publishers Inc.), 191–198.
- McGarry, L. M., Packer, A. M., Fino, E., Nikolenko, V., Sippy, T., and Yuste, R. (2010). Quantitative classification of somatostatin-positive neocortical interneurons identifies three interneuron subtypes. *Front. Neural Circuits* 4:12. doi: 10.3389/fncir.2010.00012
- Mihaljević, B., Benavides-Piccione, R., Bielza, C., DeFelipe, J., and Larrañaga, P. (2014a). Bayesian network classifiers for categorizing cortical GABAergic interneurons. *Neuroinformatics*. doi: 10.1007/s12021-014-9254-1. (in press).
- Mihaljević, B., Benavides-Piccione, R., Guerra, L., DeFelipe, J., Larrañaga, P., and Bielza, C. (2014b). Classifying GABAergic interneurons with semi-supervised projected model-based clustering. *Artif. Intell. Med.* (in press).
- Mountcastle, V. B. (1998). *Perceptual Neuroscience: The Cerebral Cortex*. Cambridge, MA: Harvard University Press.
- Nagarajan, R., Scutari, M., and Lébre, S. (2013). *Bayesian Networks in R: with Applications in Systems Biology (Use R!)*. New York, NY: Springer. doi: 10.1007/978-1-4614-6446-4
- Neapolitan, R. E. (2004). *Learning Bayesian Networks*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. San Francisco, CA: Morgan Kaufmann.
- Peña, J. M. (2011). Finding consensus Bayesian network structures. *J. Artif. Intell. Res.* 42, 661–687.
- Pennock, D. M., and Wellman, M. P. (1999). "Graphical representations of consensus belief," in *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (Stockholm: Morgan Kaufmann Publishers Inc.), 531–540.
- R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Raykar, V. C., and Yu, S. (2012). Eliminating spammers and ranking annotators for crowdsourced labeling tasks. *J. Mach. Learn. Res.* 13, 491–518.
- Raykar, V. C., Yu, S., Zhao, L. H., Valadez, G. H., Florin, C., Bogoni, L., et al. (2010). Learning from crowds. *J. Mach. Learn. Res.* 11, 1297–1322.
- Santana, R., McGarry, L. M., Bielza, C., Larrañaga, P., and Yuste, R. (2013). Classification of neocortical interneurons using affinity propagation. *Front. Neural Circuits* 7:185. doi: 10.3389/fncir.2013.00185
- Scherer, S., Kane, J., Gobl, C., and Schwenker, F. (2013). Investigating fuzzy-input fuzzy-output support vector machines for robust voice quality classification. *Comput. Speech Lang.* 27, 263–287. doi: 10.1016/j.csl.2012.06.001
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136
- Schwenker, F., and Trentin, E. (2014). Pattern classification and clustering: a review of partially supervised learning approaches. *Patt. Recogn. Lett.* 37, 4–14. doi: 10.1016/j.patrec.2013.10.017
- Scutari, M. (2010). Learning Bayesian networks with the bnlearn R package. *J. Stat. Soft.* 35, 1–22.
- Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. Y. (2008). "Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (Honolulu, HI: Association for Computational Linguistics), 254–263.
- Somogyi, P., Tamás, G., Lujan, R., and Buhl, E. H. (1998). Salient features of synaptic organization in the cerebral cortex. *Brain Res. Rev.* 26, 113–135. doi: 10.1016/S0165-0173(97)00061-1
- Sorokin, A., and Forsyth, D. (2008). "Utility data annotation with Amazon Mechanical Turk," in *First IEEE Workshop on Internet Vision at CVPR* (Anchorage, AK), 1–8.
- Sümbül, U., Song, S., McCulloch, K., Becker, M., Lin, B., Sanes, J. R., et al. (2014). A genetic and computational approach to structurally classify neuronal types. *Nat. Commun.* 5, 3512.
- Thiel, C., Scherer, S., and Schwenker, F. (2007). "Fuzzy-input fuzzy-output one-against-all support vector machines," in *Knowledge-Based Intelligent Information and Engineering Systems, Volume 4694 of Lecture Notes in Computer Science* (Vietri sul Mare: Springer), 156–165.
- Tsiola, A., Hamzei-Sichani, F., Peterlin, Z., and Yuste, R. (2003). Quantitative morphologic classification of layer 5 neurons from mouse primary visual cortex. *J. Comp. Neurol.* 461, 415–428. doi: 10.1002/cne.10628
- Turner, M. D., Chakrabarti, C., Jones, T. B., Xu, J. F., Fox, P. T., Luger, G. F., et al. (2013). Automated annotation of functional imaging experiments via multi-label classification. *Front. Neurosci.* 7:240. doi: 10.3389/fnins.2013.00240
- Van Der Gaag, L. C., and De Waal, P. R. (2006). "Multi-dimensional Bayesian network classifiers," in *Probabilistic Graphical Models* (Prague), 107–114.
- Wang, Y., Gupta, A., Toledo-Rodriguez, M., Wu, C. Z., and Markram, H. (2002). Anatomical, physiological, molecular and circuit properties of nest basket cells in the developing somatosensory cortex. *Cereb. Cortex* 12, 395–410. doi: 10.1093/cercor/12.4.395
- Welinder, P., Branson, S., Belongie, S., and Perona, P. (2010). "The multidimensional wisdom of crowds," in *Advances in Neural Information Processing Systems 23 (NIPS)* (Vancouver, BC), 2424–2432.
- Whitehill, J., Ruvolo, P., Wu, T., Bergsma, J., and Movellan, J. R. (2009). "Whose vote should count more: optimal integration of labels from labelers of unknown

expertise,” in *Advances in Neural Information Processing Systems (NIPS)*, Vol. 22 (Vancouver, BC), 2035–2043.

Yazdani, A., Ebrahimi, T., and Hoffmann, U. (2009). “Classification of EEG signals using Dempster Shafer theory and a k-nearest neighbor classifier,” in *4th International IEEE/EMBS Conference on Neural Engineering, NER’09* (Antalya: IEEE), 327–330.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 June 2014; accepted: 03 November 2014; published online: 25 November 2014.

Citation: Mihaljević B, Bielza C, Benavides-Piccione R, DeFelipe J and Larrañaga P (2014) Multi-dimensional classification of GABAergic interneurons with Bayesian network-modeled label uncertainty. *Front. Comput. Neurosci.* 8:150. doi: 10.3389/fncom.2014.00150

This article was submitted to the journal *Frontiers in Computational Neuroscience*. Copyright © 2014 Mihaljević, Bielza, Benavides-Piccione, DeFelipe and Larrañaga. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.