Linking Data Separation, Visual Separation, Classifier Performance Using Multidimensional Projections

Bárbara C. Benato[1](✉), Alexandre X. Falcão[1], and Alexandru C. Telea[2]

[1] Laboratory of Image Data Science, Institute of Computing, University of Campinas, Campinas, Brazil
{barbara.benato,afalcao}@ic.unicamp.br
[2] Department of Information and Computing Sciences, Faculty of Science, Utrecht University, Utrecht, The Netherlands
a.c.telea@uu.nl

Abstract. Understanding how data separation (DS), visual separation (VS), and classifier performance (CP) are related to each other is important for applications in both machine learning and information visualization. A recent study showed that, for a specific machine learning pipeline using a given multidimensional projection technique, high DS leads to high VS and next high CP. However, whether such correlations would stay the same (or not) when using other projection techniques was left open. We fill this gap by evaluating ten projection techniques in a pipeline that uses three contrastive learning methods (SimCLR, SupCon, and their combination) to produce latent spaces and next train and test classifiers for five image datasets of real-world application with human intestinal parasites. Our work identifies two classes of projection techniques – one leading to poor VS and next poor CS regardless of the available DS, and the other showing a good DS-VS-CP correlation. We argue that this last group of projections is a useful instrument in classifier engineering tasks.

Keywords: Data separation · Visual separation · Semi-supervised learning · Dimensionality reduction algorithms · Embedded pseudolabeling · Contrastive learning · Image classification

1 Introduction

Data separation (DS), visual separation (VS), and classifier performance (CP) are important concepts at the crossroads of machine learning (ML) and information visualization (infovis). DS relates to how well data samples of a high-dimensional dataset are separated in the respective space. Similarly, VS tells how well points of a 2D or 3D scatterplot, constructed from the high-dimensional data using dimensionality reduction, are separated. Finally, CP tells how well a classifier, constructed using the high-dimensional data and/or its mapping by a 2D or 3D scatterplot, succeeds in performing the task for which it was designed.

A. A. de Sousa et al. (Eds.): VISIGRAPP 2023, CCIS 2103, pp. 229–255, 2024.
https://doi.org/10.1007/978-3-031-66743-5_11

While DS, VS, and CP are well-defined concepts, the way they *relate* to each other is less well understood. Certain combinations of these characteristics have been studied: VS to assess DS (VS→DS) [27]; increasing DS to get easier-to-interpret VS (DS→VS) [23]; VS to find misclassified samples (VS→CP) [30]; VS to assess classification difficulty (VS→CP) [35,36]; and VS to build better classifiers (VS→CP) [5,6]. In all these connections, *projections*, or dimensionality-reduction methods, play a key role. These techniques reduce a high-dimensional dataset to a 2D or 3D scatterplot. If the scatterplot preserves well the data structure, such as separation of samples into distinct groups, it can be used as a 'proxy' to assess this structure in a cheaper way [12,30].

However, the end-to-end relationships between DS, VS, and CP – especially when using different projection techniques – are still not fully studied. Closest to this goal, we explored in earlier work [8] the links of DS, VS, and CP in the context of using pseudolabeling. For this, we created high DS from a given high-dimensional dataset using *contrastive learning* approaches. Those techniques have surpassed results of known semi-supervised and supervised losses functions [10] and become state-of-the-art [10,13,14,21]. We compared two contrastive learning models (SimCLR [10] and SupCon [21]) and proposed an approach that combines both. We evaluated DS by measuring the performance of a classifier trained with only 1% supervised samples. We measured VS on a 2D scatterplot created from the input data using t-SNE [26] by the pseudolabeling performance of a semi-supervised connectivity-based method (OPF-Semi [1]). Finally, we used our pseudolabeling to train a deep neural network and measured the network's CP. We performed all our experiments in the context of a challenging medical application (classifying human intestinal parasites in microscopy images).

Our earlier work outlined above has a key limitation. We used a *single* projection technique, t-SNE [26], which gave good VS results. This thus only shows that, if DS and VS are both high, then CP is also high. This does not explain the *full* link between DS, VS, and CP. For instance: Does a high DS *always* imply a high VS? Does a high VS *always* imply a high CP? What are these correlations when one uses a different projection technique than t-SNE?

To answer such questions, we extend our work in [8] to evaluate 10 projection techniques which produce a wide range of VS values. Our contributions are as follows:

C1: We explored contrastive learning to achieve high DS for several challenging datasets

C2: We identify projection techniques for which DS strongly correlates with VS and also techniques for which this does not happen;

C3: We show that good-VS projections are essential for training classifiers that reach a high CP.

Our work brings evidence that DS, VS, and CP are strongly correlated for a specific class of projection techniques. For these projections, one can use our proposed pipeline to design high-CP classifiers, specially for training sets with very few supervised (labeled) points. We also identify a class of projection techniques which lead to poor VS regardless of the available DS. We argue that these projections are less useful instruments in classifier design tasks and, more broadly, any infovis task where assessing DS via VS is important.

2 Related Work

2.1 Self-supervised Learning

Self-supervised contrastive methods in representation learning have been the choice for learning representations without using any labels [10,13,14,21]. Such methods aim to pull similar pairs of samples closer while pushing apart dissimilar pairs through using a *contrastive loss*. Data image transformations generate similar and dissimilar synthetic samples without using true label information – known as data views. For image data, SimCLR [10] used transformations such as cropping, Gaussian blur, color jittering, and grayscale bias. MoCo [14] explored a momentum contrast approach to learn a representation from a progressing encoder while increasing the number of dissimilar samples. BYOL [13] used only augmentations from similar examples. SimCLR has shown significant advances in (self-and-semi-) supervised alearning and achieved a new record for image classification with few labeled data. Supervised contrastive learning (SupCon) [21] generalized both SimCLR and N-pair losses and was proven to be closely related to triplet loss. SupCon surpasses cross-entropy, margin classifiers, and other self-supervised contrastive learning techniques.

2.2 Pseudolabeling

When lacking large supervised training sets to design accurate classifers, creating pseudo-labels by *propagating* labels from a few supervised samples to a large set of unsupervised ones is a well-kown possibility. Pseudolabeling, also called self-training, takes a training set with few supervised and many unsupervised samples and assigns pseudo-labels to the latter samples – a process known as data annotation – and re-trains the model with all (pseudo)labeled samples. [24] trained a neural network with 100 to 3000 supervised images and then assigned the class with maximum predicted probability to the remaining unsupervised ones. The network is then fine-tuned using both true and pseudo-labels to yield the final model. Yet, as the name suggests, pseudo-labels are not perfect, as they are *extrapolated* from actual labels, which can affect training performance [3,5]. Also, pseudolabeling methods still require training and validation sets with thousands of supervised samples per class to yield reasonable results [18,29,34].

2.3 Structure in (Embedded) Data

Although *Data structure* is a common term in ML literature, there is not an agreement in its formally definition. Here, we specify *Data structure* as data separability (DS). Simply put, for a dataset $D = \{\mathbf{x}_i \mid \mathbf{x}_i \in \mathbb{R}^n\}$, DS refers to the presence of *groups* of points which are similar and also separated from other point groups. DS is essential in ML, especially classification. Obviously, the stronger DS is, the easier is to build a classifier that separates points belonging to the various groups with high classifier performance (CP). CP can be measured by many metrics, *e.g.*, accuracy, F1 score, or AUROC [16]. Indeed, if different-class points are not separated via their features (coordinates in $\mathbb{R}^n$), then no (or poor) classification (CP) is possible.

Projections, or Dimensionality Reduction (DR) methods, take a dataset D and create a scatterplot, or embedding of D, $P(D) = \{\mathbf{y}_i = P(\mathbf{x}_i) | \mathbf{y}_i \in \mathbb{R}^q\}$, where typically $q \in \{2, 3\}$. One of the aims of projections is that the *visual structure* of $P(D)$ mimics the data structure of D. To test this, several so-called projection quality metrics have been proposed, such as trustworthiness [43], continuity [43], normalized stress [19], neighborhood hit [33], and Shepard correlation [19]. Extensive evaluations have proposed rankings of projection techniques based on these metrics [12]. However, such metrics are not precisely aiming at measuring the DS-VS structure preservation, but rather more general desirable properties of a projection (as discussed later on in Sect. 7.3.

2.4 VS, DS, and CP

The relationship among possible combinations of VS, DS, and CP have not been fully explored. Rauber *et al.* [36] used the VS of a t-SNE [25] projection to gauge the difficulty of a classification task (CP). They found that VS and CS are positively correlated when VS is medium to high but could not infer actionable insights for low-VS projections. Also, they did not address the task of *building* higher-CP classifiers using t-SNE, nor did they test other projection methods. In a related vein, Rodrigues *et al.* [37] used the VS in projections to construct so-called decision boundary maps to interpret classification performance (CP) but did not actually use these to improve classifiers. Kim *et al.* [22,23] showed that one can improve VS by increasing DS, the latter being done by mean shift [11]. However, their main goal differs when projections was not uses to build higher-CP classifiers but to achieve an easier comprehension by the user. Moreover, their approach actually changed the input data in ways not easy to control, which raises question as to the interpretability of the resulting projections. Benato *et al.* [5,9] used the VS of t-SNE projections to create pseudo-labels and train higher-CP classifiers from them. They showed that label propagation in the 2D projection space can lead to higher-CP classifiers than when propagating labels in the data space. Yet, they did not study how correlations between DS and VS can affect CP, nor did they test other projection methods than t-SNE.

2.5 Embedded Pseudolabeling (EPL)

The abovementioned topics of pseudolabeling and VS-CP correlation were connected recently by *Embedded Pseudolabeling* (EPL) [5], a method proposed to increase the number of labeled samples from only dozens of supervised samples, without needing validation sets with more supervised samples. To do this, EPL projects to 2D the latent feature space extracted from a deep neural network (DNN) using autoencoders [9] and pre-trained architectures [6]. Pseudo-labels are next propagated in the 2D projection from supervised to unsupervised samples using the OPFSemi [1] method. However, the success of EPL strongly depends on the VS in the projection space. Moreover, EPL was tested for a single projection technique (t-SNE), and it remains unclear how it would perform for other such techniques.

3 Pipeline Description

As mentioned earlier, our main intention is to evaluate DS when two *contrastive learning* models – SimCLR [10], SupCon [21], and a further combination of both – are used to learn an initial feature space. The encoder's output from contrastive learning models are used as input to EPL. Figure 1 illustrates the three main steps to investigate our claims (Sect. 1) using this process: DS has improved by contrastive learning (C1); EPL using a set of projetction techniques on this DS has led to an improved VS (C2); and pseudo-labels generated by EPL can be used to train a classifier with high CP (C3). In the same time, we aim to study how VS depends on DS and, next, how CP depends on VS. For this, we evaluate our end-to-end pipeline for 10 different projection techniques, record the obtained VS and CP values, and compute the DS-VS and VS-CP correlations. We next depict each step of our method.

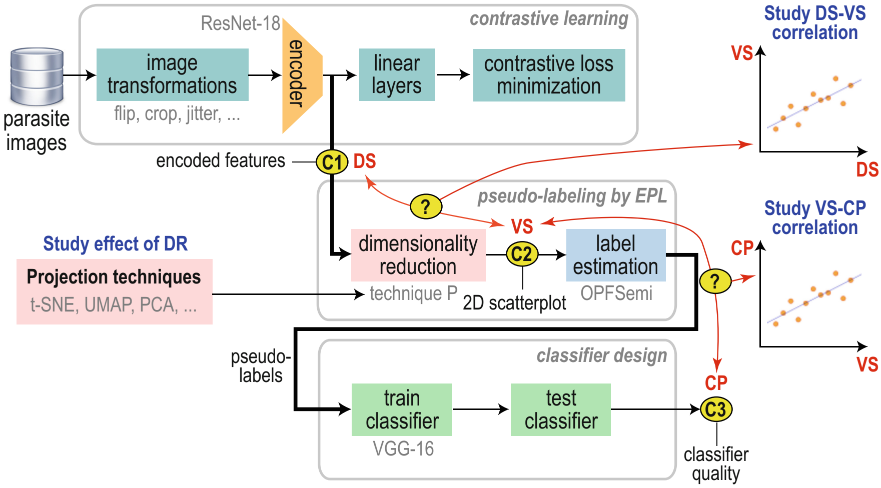


Fig. 1. Some image transformations are applied to the original data to generate synthetic examples. The model is trained over this enlarged image set using a contrastive learning loss. Encoder's output features are projected into 2D for pseudo labeling the unlabeled data. Next, original images and their assigned pseudo labels are used to train a classifier. Elements in **blue** indicate extensions of this paper *vs* [8], namely testing 10 different DR techniques and studying the obtained DS-VS and VS-CP correlations. (Color figure online)

3.1 Contrastive Learning

The latent space to be used by EPL (Fig. 1, top gray box) is learned in three different ways: (a) from the many unsupervised samples available by using SimCLR [10]; (b) using our 1% supervised samples with SupCon [21]; and (c) by combining the SimCLR and SupCon methods. All these approaches produce different amounts of DS which we study next.

3.2 Pseudolabeling by EPL

We use ResNet-18 [15] as bottleneck for both SimCLR and SupCon strategies. The encoder's output of ResNet-18 (hundreds of dimensions) is projected in a lower dimensional space (2D) using different projection techniques (Fig. 1, middle gray box). Propagating pseudo-labels in a 2D space was first observed to lead to high-CP classifiers particularly using t-SNE by EPL in [6,7]. Here, we explore the same EPL's function of exploring 2D points to propagate the (few) true labels to all unsupervised points for chosen projection techniques. This propagation is performed by OPFSemi [1] as long as it maps (un)supervised samples to nodes of a complete graph. Edges are weighted by the Euclidean distance between samples. Maximal weight of an edge in a path connecting two nodes is set as the cost of that path. The algorithm computes an optimum-path forest of minimum-cost paths rooted in the supervised samples on top of this graph. Supervised samples assign their labels to their most closely connected unsupervised nodes. OPFSemi was shown to perfom better for pseudo-label propagation than earlier semi-supervised methods [1,2,5]. Finally, we measure the VS of the 2D scatterplots created by the tested projection techniques by measuring the success of pseudolabeling.

3.3 Classifier Training with Pseudo-Labels

With the purpose of evaluating the quality of proposed pipeline, a deep neural network with ImageNet pre-trained weights – in this case, VGG-16 – is trained and tested on our parasite datasets (Fig. 1, bottom gray box). Earlier studies showed best results for our datasets using VGG-16 architecture [31].

4 Experimental Setup

4.1 Datasets

We use five image datasets (see Table 1) of color microscopy images of 200 × 200 pixels of Brazil's most common species of human intestinal parasites. Infection by these parasites in most tropical countries is a public health problem and can lead to death of infants and immunodeficient individuals [40]. Datasets are mainly the following three: (i) *Helminth larvae* (*H.larvae*, 2 classes, $3,514$ images); (ii) *Helminth eggs* (*H.eggs*, 9 classes, $5,112$ images, see examples in Fig. 2); and (iii) *Protozoan cysts* (*P.cysts*, 7 classes, $9,568$ images). These datasets are challenging since their unbalancing among distinct classes and also due to the impurity (adversarial) class, a class that is similar to other parasites classes. To evaluate different difficulty levels, we also explore (ii) and (iii) without the impurity class, which form our last two datasets. We refer to these last two datasets as *H.eggs** and *P.cysts** in our experiments.

4.2 Projection Methods

As outlined in Sect. 1, we want to evaluate the impact of different projection techniques on the measured DS-VS-CP correlations. For this, we chose the 10 most accessible and easy-to-implement projection techniques from the projection-quality benchmark

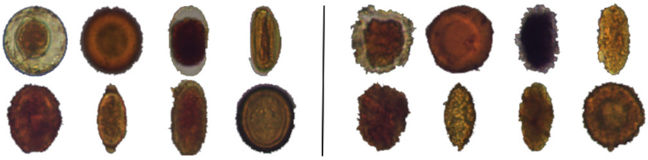


Fig. 2. Parasites dataset: H.eggs. On the left, eight classes of parasites' images in this dataset. On the right, corresponding impurity class for each class on the left – which jointly form class 9 (impurities). Reproduced from [8].

Table 1. Parasites datasets. The class names, number of classes, and number of samples per class are presented. Reproduced from [8].

dataset	classes	# samples
(i) *H.larvae* (2 classes)	*S.stercoralis*	446
	impurities	3068
	total	3,514
(ii) *H.eggs* (9 classes)	*H.nana*	348
	H.diminuta	80
	Ancilostomideo	148
	E.vermicularis	122
	A.lumbricoides	337
	T.trichiura	375
	S.mansoni	122
	Taenia	236
	impurities	3,444
	total	5,112
(iii) *P.cysts* (7 classes)	*E.coli*	719
	E.histolytica	78
	E.nana	724
	Giardia	641
	I.butschlii	1,501
	B.hominis	189
	impurities	5,716
	total	9,568

proposed in [12]. Table 2 shows our selection, with all techniques available in *scikit-learn*, except UMAP which has a separate Python implementation. Our selection covers linear *vs* nonlinear, and global *vs* local, projections; and also projections taking samples *vs* sample-distances as input. For all techniques, we fixed their parameters to the default values proposed by each author in *scikit-learn*.

Table 2. Projection techniques chosen for our evaluation. For each one, we list the linearity, type of input, and whether the technique is local or global.

projection	linearity	input	local or global
FA [20]	linear	samples	global
FICA [17]	linear	distances	global
ISO [41]	nonlinear	distances	local
KPCA [39]	nonlinear	samples	global
LLE [38]	nonlinear	samples	local
MDS [42]	nonlinear	samples	global
MLLE [44]	nonlinear	samples	local
PCA [20]	linear	samples	global
t-SNE [26]	nonlinear	samples	local
UMAP [28]	nonlinear	distances	local

4.3 Data Layout for Validation

As outlined in Sect. 1, our main goal it to build a classifier for the chosen datasets exploring only a small set of supervised samples. For this, we split each of the five considered datasets D (Sect. 4.1) into a supervised training-set S containing 1% supervised samples from D, an unsupervised training-set U with 69% of the samples in D, and a test set T with 30% of the samples in D (hence, $D = S \cup U \cup T$). We repeat the above division randomly and in a stratified manner to create three distinct splits of D in order to gain statistical relevance when evaluating results next. Table 3 shows the sizes $|S|$ and $|U|$ for each dataset.

Table 3. Number of samples in S and U for each dataset. Reproduced from [8].

sample set	H.eggs*	P. cysts*	H. larvae	H. eggs	P. cysts
S	17	38	35	51	95
U	1220	2658	2424	3527	6602

4.4 Quality Measure

To measure quality, we compute accuracy (number of correct classified or labeled samples over all the samples in a set) and Cohen's κ (since our datasets are unbalanced). κ gives the agreement level between two distinct predictions in a range $[-1, 1]$, where $\kappa \leq 0$ means no possibility, and $\kappa = 1$ means full possibility, of agreement.

4.5 Implementation Details

We next outline our end-to-end implementation.

Contrastive Learning: We implemented both SimCLR and SupCon using Pytorch in Python. We generate two augmented images (views) for each original image by random horizontal flip, resized crop (96×96), color jitter (brightness$= 0.5$, contrast$= 0.5$, saturation$= 0.5$, hue$= 0.1$) with probability of 0.8, gray-scale with probability of 0.2, Gaussian blur (9×9), and a normalization of 0.5.

Latent Space Generation: We replace ResNet-18's decision layer by a linear layer with $4,096$ neurons, a ReLU activation layer, and a linear layer with $1,024$ neurons respectively. We train the model by backpropagating errors of NT-Xent and SupCon losses for SimCLR and SupCon, respectively, with a fixed temperature of 0.07. We use the AdamW optimizer with a learning rate of 0.0005, weight decay of 0.0001, and a learning rate scheduler using cosine annealing, with a maximum temperature equal to the epochs and minimum learning rate of $0.0005/50$. We use 50 epochs and select the best model through a checkpoint obtained from the lowest validation loss during training. Finally, we use the 512 features of the ResNet-18's encoder to obtain our latent space. This setup is maintained for SimCLR+SupCon for each strategy.

Classifier Using Pseudo-Labels: We replace the original VGG-16 classifier with two linear layers with $4,096$ neurons followed by ReLU activations and a softmax decision layer. We train the model with the last four layers unfixed by backpropagating errors using categorical cross-entropy. We use stochastic gradient descent with a linear decay learning rate initialized at 0.1 and momentum of 0.9 over 15 epochs.

Parameter Setting: We used OPFSup and OPFSemi for pseudolabeling (Sect. 3.2), and they do not have any parameters. For Linear SVM (Sect. 5.1), we use the default parameters provided by scikit-learn. For the tested projection methods, see Sect. 4.2.

All our code and results are made openly available [4] for replication purposes.

5 Proposed Experiments

We next introduce a few notations to better describe our experiments. S, U, and T are the supervised (known labels), unsupervised (to be pseudo labeled), and test sets, respectively (see Sect. 4.3). Let I be the images in a given dataset having true labels L and pseudo-labels P. Let F be the latent features obtained by the three contrastive learning methods; and let F' be the features' projection to 2D via P. We use subscripts to denote on which subset I, L, P, and F are computed, *e.g.* F_S are the latent features for samples in S. Finally, let A be the initialization strategy for training a classifier C.

Figure 3 shows the several experiments we performed to explore the claims C1-C3 listed in Sect. 1. We next detail these experiments.

5.1 Experiments for Testing C1

Our first claim C1 is the following: contrastive learning methods produce high separability of classes (*i.e.*, DS) in the learned feature space. Also, we noticed that using contrastive learning increased the propagation accuracy in up to 20% *vs* using a simpler feature learning method, *i.e.*, generating the latent space via autoencoders [5]. Since the concept of data separability is not uniquely and formally defined (see Sect. 2), directly measuring DS is a difficult task. As such, we assess DS by a 'proxy' method: We train two distinct classifiers C, both using 1% supervised samples. For this, we use Linear SVM, a simple linear classifier used to check the linear separability of classes in the latent space; and OPFSup [32], an Euclidean distance-based classifier. Our assumption is if these classifiers yield high quality, then DS is high, and conversely. We measure classifiers' quality by accuracy and κ over correctly classified samples in T.

With the above, we conduct three experiments – one per method of latent space generation (see Sect. 3.1):

a) *SimCLR:* Train with A on $I_{S \cup U}$; extract features F_S and F_T; train C on F_S and L_S; test on F_T and L_T.
b) *SupCon:* Train with A on I_S and L_S; extract features F_S and F_T; train and test as above.
c) *SimCLR+SupCon:* Train SimCLR with A on $I_{S \cup U}$; fine-tune with SupCon on I_S and L_S; extract features F_S and F_T; train and test as above.

5.2 Experiments for Testing C2

For testing C2, many strategies could be used to evaluate the VS of projections since visual separation of clusters in a 2D scatterplot is a broad concept. Several metrics have been proposed for this task in DR literature – see surveys [12,30]. However, existing metrics are usually used to gauge the projection quality when explored by a *human*. Rather, in our context, we use projections *automatically* to drive pseudo labeling and improve classification (Sect. 3.2) – and, in this process, we aim to find which projection techniques are best for this task. In this way, evaluating our projections' VS by how well they can do this label propagation is a good assessment for this purpose. We compare the computed pseudo labels with the true, supervised, labels by computing accuracy and κ for the correctly computed pseudo-labels over U. This comparison is performed using distinct projection techniques P (Sect. 4.2) to understand how P is directly related to the different VS resulted from each projection. We proposed three experiments aim to achieve precisely this:

a) *SimCLR:* Train with A on $I_{S \cup U}$; extract features $F_{S \cup U}$; compute 2D features F' with P from $F_{S \cup U}$; propagate labels L_S with OPFSemi from F'_S to F'_U;
b) *SupCon:* Train with A on I_S and L_S; extract features $F_{S \cup U}$; compute 2D features F' with P from $F_{S \cup U}$; propagate labels as above;
c) *SimCLR+SupCon:* Train SimCLR with A on $I_{S \cup U}$; fine-tune with SupCon on I_S and L_S; extract features $I_{S \cup U}$; compute 2D features F' with P from $F_{S \cup U}$; propagate labels as above.

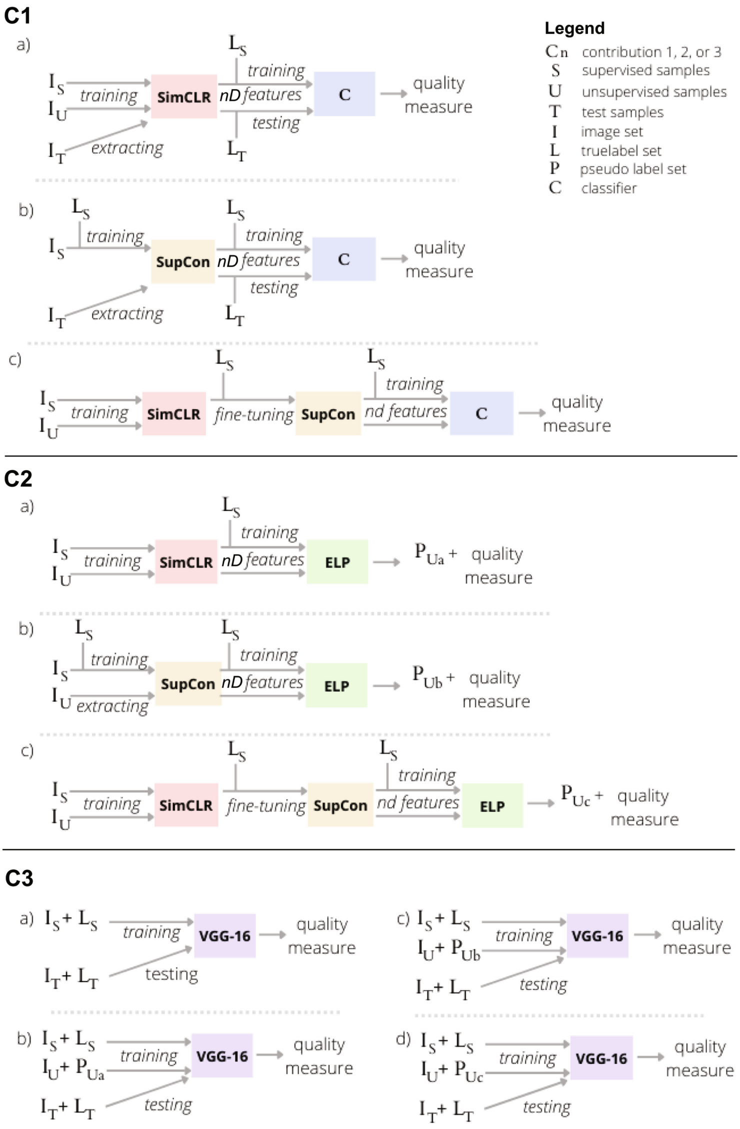


Fig. 3. Summary of the proposed experiments for testing our claims C1, C2, and C3. We added C1.c) to complement the experimental setup first proposed in [8].

5.3 Experiments for Testing C3

Finally, the computed pseudo labels are used to train and test a DNN classifier, in this case VGG-16, to test how CS is correlated (or not) with VS and DS. For this, we do the following experiments:

a) *baseline:* train with I_S and L_S; test on I_T and L_T;

b) *SimCLR:* train with $I_{S \cup U}$ and $L_{S \cup P_U}$, with pseudo-labels P_U from (Sect. 5.2,a); test as above;
c) *SupCon:* train with $I_{S \cup U}$ and $L_{S \cup P_U}$, with pseudo-labels P_U from (Sect. 5.2,b); test as above;
d) *SimCLR+SupCon:* train with $I_{S \cup U}$ and $L_{S \cup P_U}$, with pseudo-labels P_U from (Sect. 5.2,c); test as above.

6 Experiments and Results

We next show the results of the experiments in Sect. 5 along our claims C1-C3.

6.1 C1: Contrastive Learning Yields High DS

Table 4 shows the classification results for the experiments in Sect. 5.1 in terms of accuracy and κ (mean and standard deviation) for the trained Linear SVM and OPFSup classifiers. To ease interpretation, we next summarize these results by averaging classification values using a heatmap in Fig. 4.

Table 4. C1: DS assessment of SimCLR's, SupCon's, and SimCLR+SupCon's latent spaces using Linear SVM and OPFSup on T. The three methods are compared trained from scratch and with pre-trained weights during 50 epochs. Best values per dataset and model initialization are in bold.

datasets	metrics	from scratch					
		SimCLR		SupCon		SimCLR+SupCon	
		LinearSVM	OPFSemi	LinearSVM	OPFSemi	LinearSVM	OPFSemi
H.eggs*	acc	0.842436 ± 0.047	0.861268 ± 0.019	0.811048 ± 0.048	0.752668 ± 0.058	0.890144 ± 0.015	**0.902072** ± **0.006**
	κ	0.810976 ± 0.057	0.835316 ± 0.023	0.770732 ± 0.058	0.705686 ± 0.067	0.868792 ± 0.018	**0.883541** ± **0.007**
P.cysts*	acc	**0.722607** ± **0.034**	0.678489 ± 0.025	0.668973 ± 0.094	0.660611 ± 0.016	0.719723 ± 0.010	0.673876 ± 0.032
	κ	0.616310 ± 0.047	0.569074 ± 0.031	0.559129 ± 0.121	0.553795 ± 0.019	**0.616431** ± **0.011**	0.562813 ± 0.039
H.larvae	acc	0.935545 ± 0.003	0.905213 ± 0.003	0.929858 ± 0.009	0.901422 ± 0.037	**0.934913** ± **0.015**	0.932070 ± 0.017
	κ	**0.710109** ± **0.040**	0.564216 ± 0.050	0.673113 ± 0.020	0.554242 ± 0.165	0.708803 ± 0.064	0.683580 ± 0.094
H.eggs	acc	0.772056 ± 0.005	0.710778 ± 0.012	0.657975 ± 0.004	0.561930 ± 0.037	**0.783572** ± **0.022**	0.730335 ± 0.025
	κ	0.565300 ± 0.041	0.524010 ± 0.017	0.122930 ± 0.116	0.259556 ± 0.045	**0.595363** ± **0.062**	0.543848 ± 0.038
P.cysts	acc	0.733078 ± 0.028	0.627772 ± 0.009	0.628701 ± 0.017	0.527342 ± 0.015	**0.766284** ± **0.018**	0.677464 ± 0.026
	κ	0.561195 ± 0.025	0.409251 ± 0.010	0.254190 ± 0.056	0.260470 ± 0.017	**0.600513** ± **0.037**	0.482015 ± 0.033
datasets	metrics	pre-trained					
		SimCLR		SupCon		SimCLR+SupCon	
		LinearSVM	OPFSemi	LinearSVM	OPFSemi	LinearSVM	OPFSemi
H.eggs*	acc	0.809793 ± 0.031	0.834903 ± 0.032	0.854990 ± 0.017	0.842436 ± 0.018	0.839297 ± 0.013	**0.880728** ± **0.020**
	κ	0.773903 ± 0.037	0.803842 ± 0.038	0.825372 ± 0.021	0.811695 ± 0.023	0.809970 ± 0.015	**0.858649** ± **0.024**
P.cysts*	acc	0.685410 ± 0.039	0.580450 ± 0.012	**0.742215** ± **0.012**	0.697232 ± 0.005	0.690312 ± 0.030	0.614764 ± 0.011
	κ	0.579214 ± 0.048	0.444268 ± 0.019	**0.652441** ± **0.012**	0.595298 ± 0.006	0.571982 ± 0.039	0.483336 ± 0.013
H.larvae	acc	0.949447 ± 0.007	0.947551 ± 0.016	0.950079 ± 0.008	0.949447 ± 0.010	**0.952607** ± **0.007**	0.951343 ± 0.005
	κ	**0.779016** ± **0.039**	0.767287 ± 0.080	0.755562 ± 0.061	0.748063 ± 0.072	0.777646 ± 0.049	0.775075 ± 0.045
H.eggs	acc	0.772490 ± 0.022	0.755976 ± 0.040	0.780965 ± 0.053	0.703390 ± 0.065	0.765971 ± 0.038	**0.782051** ± **0.017**
	κ	0.606751 ± 0.031	0.586526 ± 0.055	0.566952 ± 0.146	0.469467 ± 0.108	0.591410 ± 0.056	**0.617267** ± **0.022**
P.cysts	acc	0.616278 ± 0.083	0.615117 ± 0.020	**0.721932** ± **0.013**	0.635551 ± 0.007	0.709857 ± 0.043	0.681528 ± 0.024
	κ	0.378195 ± 0.125	0.370532 ± 0.043	0.505566 ± 0.043	0.419729 ± 0.011	**0.526674** ± **0.052**	0.486552 ± 0.030

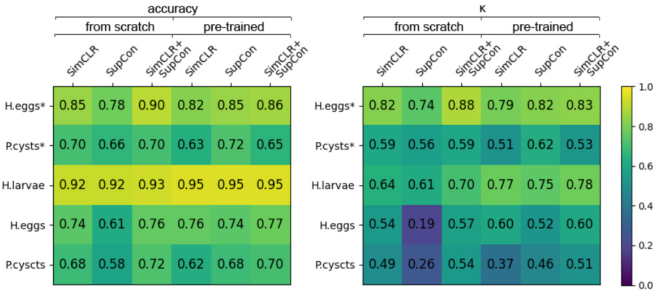


Fig. 4. C1: Values from Table 4 averaged per contrastive learning technique.

We first discuss the contrastive learning methods trained from scratch *vs* using ImageNet pre-trained weights. The best accuracy and κ for all datasets exceed 0.70 and 0.59, respectively. Linear SVM obtained the best results for most datasets, showing that these latent spaces have a reasonable *linear separation* between classes even when classified with only 1% supervised samples. In contrast, OPFSup seems to suffer from the dimensionality curse as it uses Euclidean distances in the latent space. This further motivates the latent space's dimensionality reduction when using an OPF classifier. In cases where OPFSemi values are higher than LinearSVM, both are relatively close, i.e., for *H.eggs** using SimCLR+SupCon. Separately, we see that the ImageNet pre-trained weights helped the three compared methods for most methods and datasets, while the best values for *H.eggs** and *P.cysts* are obtained when trained from scratch. SimCLR had an increase of around 0.10 in κ for *H.larvae* with pre-trained weights. SupCon also had an extra 0.10 accuracy and κ for all datasets with pre-trained weights. SimCLR+SupCon achieved its best results compared to SimCLR and SupCon for most datasets with pre-trained weights. We also see this in Fig. 4 – for the κ plot, brighter cells are mainly in the SimCLR+SupCon columns. We also see that higher accuracy values do not always reflect higher κ values, e.g., for the *H.larvae* dataset. This can be justified by the unbalancing of classes presented in such datasets, as discussed in Sect. 4.1.

Although contrastive learning yields quite high DS values, we see both in Table 4 and Fig. 4 that there exists quite some DS variation across datasets. This will help us next explore how different projection methods map these values to visual separation (VS).

6.2 C2: Correlation Between Different Projections and VS

Table 5 shows the results for the experiments in Sect. 5.2, *i.e.*, the mean propagation accuracy and κ in pseudolabeling for the correctly assigned labels in U for EPL run using the selected P on latent spaces created by SimCLR, SupCon, and SimCLR+SupCon. Figure 5(left) shows the same data using a heat map for easier interpretation.

We got the best results using the ImageNet pre-trained weights – see the lightgreen-yellow cells in the three rightmost columns marked *pre-trained* in Fig. 5(left). This shows that such pre-trained weights favor the pseudolabeling on the contrastive latent space. Additionally, SimCLR+SupCon obtained the best results among compared the compared projections – see the brighter colors in the rightmost column in the same figure.

More interestingly, we see a clear pattern of similar horizontal colors in Fig. 5(left). For κ, these 'color bands' match very well the projection techniques. For instance, the dark-blue block of cells in the figure tells that FICA, ISO, KPCA, LLE, and MLLE score very poor values, *i.e.*, create a poor VS. In contrast, t-SNE, UMAP, FA, and MDS have brighter cells, so, they create better VS. We will explore this further in Sect. 7.

6.3 C3: Classifiers Trained by Pseudo-Labels Obtained from High-VS Projections Have a High CP

Table 6 shows the results of classification for VGG-16 trained from the pseudolabeling performed on latent spaces from SimCLR, SupCon, and SimCLR+SupCon. Figure 5(right) shows the same data as a heatmap plot for interpretation ease. We see that the *"C2: Propagation results"* (left) and *"C3: Classification results"* (right) subfigures in Fig. 5 are very similar – hence, CP is indeed highly correlated with VS.

We got the best CP results by the methods using the ImageNet pre-trained weights – see the three rightmost columns in Fig. 5(right) which have brighter cells. Also, SimCLR+SupCon obtained the best results for most datasets, as shown by the lighter color in the last column. Projection-wise, we see again that t-SNE, UMAP, and MDS lead to the highest overall CP values. Separately, κ shows a bigger gap between projection techniques with high values (lighter/yellow cells) and low values (darker/purple cells). These results show that VGG-16 can learn from the pseudo-labels since it provided good classification accuracies and κ – higher than 0.80 and 0.70, respectively – among the studied datasets and projections.

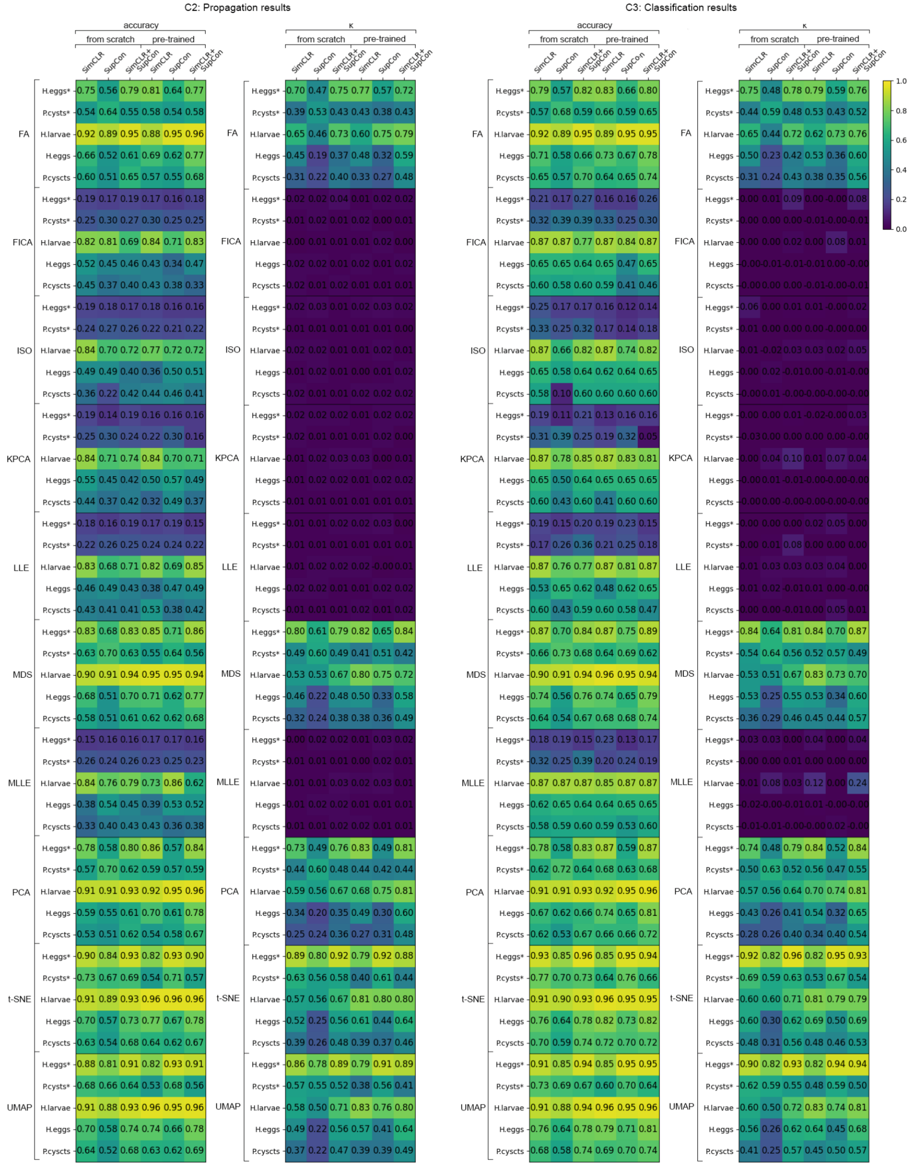
Table 5. C2: Propagation results for pseudo-labeling U on the projected SimCLR's and SupCon's latent spaces, from scratch and using ImageNet pre-trained weights. Best values per dataset are in bold.

techniques	datasets	metric	from scratch			pre-trained		
			SimCLR	SupCon	SimCLR+SupCon	SimCLR	SupCon	SimCLR+SupCon
FA	H.eggs*	acc	0.747777 ± 0.068	0.560496 ± 0.050	0.791431 ± 0.040	**0.807060** ± **0.026**	0.641337 ± 0.071	0.766370 ± 0.065
		κ	0.697796 ± 0.082	0.473382 ± 0.062	0.750259 ± 0.047	**0.767607** ± **0.032**	0.566708 ± 0.086	0.721839 ± 0.077
	P.cysts*	acc	0.539688 ± 0.073	**0.644288** ± **0.054**	0.552300 ± 0.022	0.576286 ± 0.039	0.537092 ± 0.038	0.576781 ± 0.038
		κ	0.389467 ± 0.085	**0.530016** ± **0.059**	0.427268 ± 0.046	0.428285 ± 0.047	0.383229 ± 0.065	0.428245 ± 0.059
	H.larvae	acc	0.920835 ± 0.021	0.889115 ± 0.015	0.947540 ± 0.012	0.882201 ± 0.056	0.950386 ± 0.005	**0.955809** ± **0.008**
		κ	0.649369 ± 0.093	0.455761 ± 0.078	0.726970 ± 0.092	0.599440 ± 0.164	0.751991 ± 0.044	**0.785414** ± **0.053**
	H.eggs	acc	0.659959 ± 0.030	0.518726 ± 0.016	0.607322 ± 0.040	0.693032 ± 0.014	0.618129 ± 0.078	**0.768586** ± **0.049**
		κ	0.445596 ± 0.026	0.187524 ± 0.014	0.368271 ± 0.066	0.480564 ± 0.027	0.319189 ± 0.101	**0.587166** ± **0.078**
	P.cysts	acc	0.596337 ± 0.031	0.513513 ± 0.018	0.649445 ± 0.038	0.573540 ± 0.004	0.550197 ± 0.029	**0.675427** ± **0.018**
		κ	0.305130 ± 0.020	0.219125 ± 0.034	0.399547 ± 0.084	0.331404 ± 0.048	0.271909 ± 0.046	**0.482847** ± **0.007**
FICA	H.eggs*	acc	0.191054 ± 0.003	0.174616 ± 0.034	**0.191592** ± **0.017**	0.167071 ± 0.005	0.164915 ± 0.028	0.180005 ± 0.001
		κ	0.015913 ± 0.004	0.015189 ± 0.009	0.036693 ± 0.009	0.011312 ± 0.021	0.015913 ± 0.008	**0.020143** ± **0.009**
	P.cysts*	acc	0.245920 ± 0.032	0.295128 ± 0.008	0.269288 ± 0.017	**0.301681** ± **0.043**	0.248393 ± 0.043	0.247404 ± 0.045
		κ	0.013333 ± 0.009	0.017266 ± 0.010	0.012307 ± 0.005	**0.018144** ± **0.009**	0.004888 ± 0.005	0.009506 ± 0.009
	H.larvae	acc	0.818219 ± 0.030	0.814152 ± 0.027	0.688762 ± 0.132	**0.839637** ± **0.026**	**0.711265** ± **0.153**	0.829606 ± 0.017
		κ	0.001788 ± 0.003	0.009610 ± 0.012	0.013793 ± 0.026	**0.014248** ± **0.006**	0.017165 ± 0.013	0.003038 ± 0.022
	H.eggs	acc	**0.518539** ± **0.048**	0.452674 ± 0.064	0.456307 ± 0.046	0.433016 ± 0.026	0.343767 ± 0.100	0.472331 ± 0.077
		κ	0.016130 ± 0.002	**0.015664** ± **0.006**	0.013574 ± 0.005	0.018069 ± 0.001	0.010873 ± 0.002	0.016699 ± 0.001
	P.cysts	acc	**0.450948** ± **0.031**	0.367179 ± 0.079	0.395749 ± 0.031	0.429745 ± 0.072	0.380469 ± 0.129	0.330695 ± 0.070
		κ	0.018332 ± 0.004	0.013987 ± 0.007	**0.020642** ± **0.008**	0.014538 ± 0.008	0.014562 ± 0.003	0.007237 ± 0.003
ISO	H.eggs*	acc	**0.194826** ± **0.011**	0.182700 ± 0.014	0.172460 ± 0.017	0.176772 ± 0.028	0.162220 ± 0.028	0.164915 ± 0.032
		κ	0.019858 ± 0.007	0.028010 ± 0.004	0.009387 ± 0.003	0.021445 ± 0.006	**0.034176** ± **0.008**	0.023816 ± 0.006
	P.cysts*	acc	0.235534 ± 0.020	**0.268917** ± **0.056**	0.260880 ± 0.024	0.224901 ± 0.008	0.211795 ± 0.035	0.219955 ± 0.010
		κ	0.011823 ± 0.004	**0.013657** ± **0.005**	0.013012 ± 0.001	0.008791 ± 0.003	0.010802 ± 0.006	0.002970 ± 0.003
	H.larvae	acc	**0.844923** ± **0.015**	0.698116 ± 0.224	0.724278 ± 0.156	0.774976 ± 0.041	0.717365 ± 0.188	0.722380 ± 0.168
		κ	0.022980 ± 0.016	**0.023176** ± **0.015**	0.009617 ± 0.008	0.010355 ± 0.018	0.017956 ± 0.010	0.009347 ± 0.011
	H.eggs	acc	0.486398 ± 0.067	0.489379 ± 0.125	0.399665 ± 0.023	0.364263 ± 0.079	0.503168 ± 0.112	**0.505496** ± **0.012**
		κ	0.014483 ± 0.013	**0.020610** ± **0.005**	0.010909 ± 0.007	0.004325 ± 0.007	0.013514 ± 0.004	0.019615 ± 0.003
	P.cysts	acc	0.359365 ± 0.008	0.223732 ± 0.014	0.419790 ± 0.027	0.435717 ± 0.016	**0.463541** ± **0.022**	0.406202 ± 0.034
		κ	0.015452 ± 0.008	0.009035 ± 0.001	0.014798 ± 0.002	0.016392 ± 0.006	**0.019200** ± **0.012**	0.017231 ± 0.004
KPCA	H.eggs*	acc	**0.190245** ± **0.009**	0.143088 ± 0.018	0.188898 ± 0.007	0.158717 ± 0.041	0.155753 ± 0.023	0.163837 ± 0.023
		κ	**0.022804** ± **0.011**	0.017998 ± 0.006	0.020763 ± 0.004	0.009970 ± 0.008	0.016143 ± 0.015	0.016003 ± 0.003
	P.cysts*	acc	0.252596 ± 0.046	**0.304772** ± **0.056**	0.243571 ± 0.047	0.222552 ± 0.008	0.302423 ± 0.049	0.164565 ± 0.012
		κ	0.020874 ± 0.015	0.010778 ± 0.007	0.008015 ± 0.008	0.007448 ± 0.007	**0.021246** ± **0.002**	0.002761 ± 0.003
	H.larvae	acc	0.838824 ± 0.006	0.709231 ± 0.179	**0.741223** ± **0.126**	0.837332 ± 0.029	0.697438 ± 0.162	0.709773 ± 0.181
		κ	0.010105 ± 0.017	0.023394 ± 0.012	**0.034326** ± **0.015**	0.025248 ± 0.005	0.003343 ± 0.009	0.007443 ± 0.004
	H.eggs	acc	0.550307 ± 0.052	0.452022 ± 0.185	0.421837 ± 0.030	0.499348 ± 0.054	**0.574064** ± **0.022**	0.490963 ± 0.064
		κ	0.014382 ± 0.005	0.019979 ± 0.009	**0.020720** ± **0.008**	0.007635 ± 0.006	0.019011 ± 0.006	0.019884 ± 0.007
	P.cysts	acc	0.441292 ± 0.024	0.367080 ± 0.125	0.417849 ± 0.047	0.323727 ± 0.088	**0.490817** ± **0.017**	0.371310 ± 0.053
		κ	0.008013 ± 0.009	0.013086 ± 0.006	0.012913 ± 0.001	**0.014376** ± **0.008**	0.011941 ± 0.007	0.008518 ± 0.004
LLE	H.eggs*	acc	0.177580 ± 0.009	0.163029 ± 0.034	**0.194557** ± **0.003**	0.165185 ± 0.011	0.189976 ± 0.014	0.153058 ± 0.026
		κ	0.012537 ± 0.006	0.010076 ± 0.003	**0.024128** ± **0.006**	0.016414 ± 0.006	0.033328 ± 0.010	0.000560 ± 0.012
	P.cysts*	acc	0.217483 ± 0.016	**0.260633** ± **0.039**	0.252967 ± 0.020	0.238501 ± 0.070	0.244931 ± 0.047	0.220450 ± 0.002
		κ	0.005612 ± 0.002	0.009387 ± 0.002	0.011296 ± 0.008	**0.011855** ± **0.013**	0.007083 ± 0.005	0.006551 ± 0.003
	H.larvae	acc	0.833401 ± 0.006	0.677918 ± 0.162	0.712214 ± 0.188	0.817812 ± 0.038	0.688762 ± 0.160	**0.847499** ± **0.029**
		κ	**0.013540** ± **0.026**	0.018801 ± 0.010	0.018322 ± 0.004	0.016302 ± 0.005	-0.004704 ± 0.009	0.006490 ± 0.016
	H.eggs	acc	0.457798 ± 0.129	0.485374 ± 0.075	0.430501 ± 0.116	0.379356 ± 0.110	0.468604 ± 0.138	**0.492920** ± **0.072**
		κ	0.017533 ± 0.002	**0.022688** ± **0.005**	0.012804 ± 0.007	0.010157 ± 0.005	0.018747 ± 0.003	0.015604 ± 0.008
	P.cysts	acc	0.431039 ± 0.069	0.408242 ± 0.118	0.408541 ± 0.081	**0.526106** ± **0.011**	0.375890 ± 0.077	0.418446 ± 0.102
		κ	0.013445 ± 0.006	0.009230 ± 0.004	0.014226 ± 0.008	**0.018359** ± **0.007**	0.012149 ± 0.006	0.015147 ± 0.004

continued

Table 5. continued

techniques	datasets	metric	from scratch			pre-trained		
			SimCLR	SupCon	SimCLR+SupCon	SimCLR	SupCon	SimCLR+SupCon
MDS	H.eggs*	acc	0.831851 ± 0.010	0.679332 ± 0.055	0.825923 ± 0.028	0.852331 ± 0.047	0.710051 ± 0.024	**0.863110 ± 0.016**
		κ	0.797733 ± 0.012	0.614286 ± 0.068	0.789018 ± 0.035	0.823455 ± 0.056	0.647216 ± 0.028	**0.835812 ± 0.019**
	P.cysts*	acc	0.627473 ± 0.040	**0.699060 ± 0.014**	0.629204 ± 0.028	0.550198 ± 0.050	0.639343 ± 0.017	0.563922 ± 0.031
		κ	0.494613 ± 0.061	**0.595492 ± 0.015**	0.490610 ± 0.043	0.408429 ± 0.058	0.514499 ± 0.022	0.416427 ± 0.040
	H.larvae	acc	0.896570 ± 0.008	0.907279 ± 0.021	0.938728 ± 0.012	**0.952691 ± 0.010**	0.949573 ± 0.006	0.944286 ± 0.014
		κ	0.530915 ± 0.025	0.532948 ± 0.100	0.670950 ± 0.103	**0.799463 ± 0.039**	0.746419 ± 0.050	0.716183 ± 0.115
	H.eggs	acc	0.677008 ± 0.002	0.510062 ± 0.035	0.699273 ± 0.032	0.711384 ± 0.042	0.620179 ± 0.041	**0.774734 ± 0.035**
		κ	0.455775 ± 0.042	0.223363 ± 0.025	0.480271 ± 0.045	0.501590 ± 0.055	0.325114 ± 0.050	**0.584801 ± 0.064**
	P.cysts	acc	0.577920 ± 0.018	0.508685 ± 0.010	0.606142 ± 0.063	0.615699 ± 0.013	0.616694 ± 0.016	**0.677517 ± 0.043**
		κ	0.318017 ± 0.034	0.241293 ± 0.016	0.383062 ± 0.068	0.382044 ± 0.036	0.359747 ± 0.024	**0.489937 ± 0.053**
MLLE	H.eggs*	acc	0.154136 ± 0.039	0.164915 ± 0.022	0.155753 ± 0.036	**0.173269 ± 0.005**	0.168418 ± 0.026	0.158717 ± 0.028
		κ	0.004386 ± 0.008	0.024440 ± 0.004	0.019636 ± 0.003	0.012537 ± 0.004	**0.030116 ± 0.005**	0.018084 ± 0.001
	P.cysts*	acc	**0.261746 ± 0.027**	0.244313 ± 0.039	0.258902 ± 0.006	0.225766 ± 0.023	0.250495 ± 0.057	0.228116 ± 0.005
		κ	0.012912 ± 0.008	0.013391 ± 0.002	0.007871 ± 0.006	0.004547 ± 0.003	**0.015660 ± 0.006**	0.010145 ± 0.005
	H.larvae	acc	0.836248 ± 0.009	0.764810 ± 0.062	0.790159 ± 0.048	0.728074 ± 0.064	**0.862003 ± 0.011**	0.621933 ± 0.135
		κ	0.012424 ± 0.011	0.014140 ± 0.029	0.025816 ± 0.018	0.016783 ± 0.015	**0.027929 ± 0.008**	0.005776 ± 0.008
	H.eggs	acc	0.379635 ± 0.038	**0.539035 ± 0.034**	0.449972 ± 0.049	0.389510 ± 0.077	0.531582 ± 0.065	0.520682 ± 0.031
		κ	0.013558 ± 0.010	**0.023266 ± 0.003**	0.016370 ± 0.005	0.013985 ± 0.012	0.008897 ± 0.006	0.014190 ± 0.003
	P.cysts	acc	0.331492 ± 0.042	0.400080 ± 0.050	**0.429994 ± 0.023**	0.427356 ± 0.085	0.362152 ± 0.077	0.377731 ± 0.022
		κ	0.012621 ± 0.003	0.014195 ± 0.002	**0.019642 ± 0.003**	0.017896 ± 0.004	0.007093 ± 0.002	0.009973 ± 0.004
PCA	H.eggs*	acc	0.776341 ± 0.022	0.575047 ± 0.072	0.801132 ± 0.043	**0.855026 ± 0.052**	0.565616 ± 0.061	0.838857 ± 0.030
		κ	0.730955 ± 0.027	0.485557 ± 0.093	0.760041 ± 0.053	**0.826294 ± 0.063**	0.485068 ± 0.065	0.808110 ± 0.035
	P.cysts*	acc	0.572700 ± 0.042	**0.696588 ± 0.023**	0.615356 ± 0.055	0.585188 ± 0.019	0.573318 ± 0.033	0.590999 ± 0.040
		κ	0.438663 ± 0.050	**0.597055 ± 0.026**	0.483103 ± 0.066	0.442557 ± 0.022	0.420640 ± 0.051	0.440307 ± 0.057
	H.larvae	acc	0.910804 ± 0.005	0.910804 ± 0.026	0.931951 ± 0.017	0.915006 ± 0.033	0.950522 ± 0.007	**0.958520 ± 0.001**
		κ	0.587131 ± 0.050	0.563387 ± 0.151	0.668527 ± 0.105	0.676664 ± 0.116	0.751620 ± 0.053	**0.811768 ± 0.001**
	H.eggs	acc	0.587572 ± 0.038	0.545929 ± 0.022	0.612167 ± 0.033	0.703186 ± 0.041	0.612540 ± 0.088	**0.777809 ± 0.042**
		κ	0.340487 ± 0.048	0.201855 ± 0.015	0.351600 ± 0.053	0.493198 ± 0.055	0.302314 ± 0.122	**0.596395 ± 0.072**
	P.cysts	acc	0.530536 ± 0.034	0.508536 ± 0.005	0.615599 ± 0.039	0.544124 ± 0.031	0.584092 ± 0.007	**0.669653 ± 0.048**
		κ	0.245823 ± 0.084	0.241375 ± 0.011	0.361185 ± 0.080	0.273647 ± 0.076	0.314512 ± 0.030	**0.475270 ± 0.058**
t-SNE	H.eggs*	acc	0.904069 ± 0.022	0.836163 ± 0.026	0.931016 ± 0.025	0.824306 ± 0.027	**0.933980 ± 0.015**	0.901374 ± 0.015
		κ	0.885518 ± 0.026	0.804574 ± 0.031	0.918209 ± 0.030	0.791693 ± 0.031	**0.921045 ± 0.018**	0.882814 ± 0.017
	P.cysts*	acc	**0.726014 ± 0.060**	0.674085 ± 0.025	0.691642 ± 0.027	0.537834 ± 0.030	0.710064 ± 0.007	0.572082 ± 0.028
		κ	**0.631292 ± 0.078**	0.564841 ± 0.040	0.583401 ± 0.043	0.397794 ± 0.033	0.608087 ± 0.016	0.441656 ± 0.026
	H.larvae	acc	0.906059 ± 0.013	0.892233 ± 0.039	0.925037 ± 0.022	**0.957299 ± 0.008**	0.955537 ± 0.002	0.955537 ± 0.005
		κ	0.572966 ± 0.086	0.556742 ± 0.151	0.670628 ± 0.099	**0.809426 ± 0.042**	0.795117 ± 0.014	0.801494 ± 0.016
	H.eggs	acc	0.703279 ± 0.032	0.570989 ± 0.040	0.731787 ± 0.013	0.772592 ± 0.009	0.669648 ± 0.085	**0.783492 ± 0.030**
		κ	0.524161 ± 0.043	0.248194 ± 0.041	0.558122 ± 0.030	0.608138 ± 0.021	0.436442 ± 0.124	**0.636986 ± 0.046**
	P.cysts	acc	0.630481 ± 0.012	0.538948 ± 0.013	**0.678413 ± 0.011**	0.644368 ± 0.035	0.616644 ± 0.019	0.667214 ± 0.028
		κ	0.393737 ± 0.002	0.263230 ± 0.010	**0.478606 ± 0.032**	0.389898 ± 0.062	0.366433 ± 0.047	0.460750 ± 0.030
UMAP	H.eggs*	acc	0.880625 ± 0.020	0.811372 ± 0.056	0.907033 ± 0.026	0.823767 ± 0.070	**0.926435 ± 0.018**	0.908919 ± 0.011
		κ	0.857432 ± 0.024	0.777552 ± 0.066	0.889462 ± 0.032	0.790407 ± 0.084	**0.911833 ± 0.022**	0.891483 ± 0.013
	P.cysts*	acc	**0.679896 ± 0.045**	0.661597 ± 0.023	0.640826 ± 0.034	0.527201 ± 0.042	0.678783 ± 0.012	0.564293 ± 0.028
		κ	**0.566764 ± 0.059**	0.554236 ± 0.022	0.518054 ± 0.051	0.382443 ± 0.048	0.562840 ± 0.029	0.414946 ± 0.045
	H.larvae	acc	0.907957 ± 0.011	0.877999 ± 0.043	0.931815 ± 0.017	**0.961095 ± 0.004**	0.949980 ± 0.006	0.958113 ± 0.001
		κ	0.584348 ± 0.059	0.499324 ± 0.130	0.710935 ± 0.049	**0.830896 ± 0.019**	0.763058 ± 0.053	0.803964 ± 0.011
	H.eggs	acc	0.695640 ± 0.021	0.578349 ± 0.006	0.743432 ± 0.051	0.742035 ± 0.044	0.660145 ± 0.061	**0.784796 ± 0.037**
		κ	0.488702 ± 0.012	0.217511 ± 0.009	0.563820 ± 0.081	0.567092 ± 0.056	0.410992 ± 0.121	**0.637002 ± 0.051**
	P.cysts	acc	0.635707 ± 0.011	0.518342 ± 0.024	0.677318 ± 0.041	0.634214 ± 0.024	0.623264 ± 0.008	**0.687223 ± 0.031**
		κ	0.368501 ± 0.023	0.215635 ± 0.010	0.470794 ± 0.064	0.388012 ± 0.058	0.387465 ± 0.006	**0.491069 ± 0.035**



C2: Propagation results

		accuracy						κ					
		from scratch			pre-trained			from scratch			pre-trained		
		SimCLR	SupCon	SimCLR+ SupCon	SimCLR	SupCon	SimCLR+ SupCon	SimCLR	SupCon	SimCLR+ SupCon	SimCLR	SupCon	SimCLR+ SupCon
FA	H.eggs*	0.75	0.56	0.79	0.81	0.64	0.77	0.70	0.47	0.75	0.77	0.57	0.72
	P.cysts*	0.54	0.64	0.55	0.58	0.54	0.58	0.39	0.53	0.43	0.43	0.38	0.43
	H.larvae	0.92	0.89	0.95	0.88	0.95	0.96	0.65	0.46	0.73	0.60	0.75	0.79
	H.eggs	0.66	0.52	0.61	0.69	0.62	0.77	0.45	0.19	0.37	0.48	0.32	0.59
	P.cyscts	0.60	0.51	0.65	0.57	0.55	0.68	0.31	0.22	0.40	0.33	0.27	0.48
FICA	H.eggs*	0.19	0.17	0.19	0.17	0.16	0.18	0.02	0.02	0.04	0.01	0.02	0.02
	P.cysts*	0.25	0.30	0.27	0.30	0.25	0.25	0.01	0.02	0.01	0.02	0.00	0.01
	H.larvae	0.82	0.81	0.69	0.84	0.71	0.83	0.00	0.01	0.01	0.01	0.02	0.00
	H.eggs	0.52	0.45	0.46	0.43	0.34	0.47	0.02	0.02	0.01	0.02	0.01	0.02
	P.cyscts	0.45	0.37	0.40	0.43	0.38	0.33	0.02	0.01	0.02	0.01	0.01	0.01
ISO	H.eggs*	0.19	0.18	0.17	0.18	0.16	0.16	0.02	0.03	0.01	0.02	0.03	0.02
	P.cysts*	0.24	0.27	0.26	0.22	0.21	0.22	0.01	0.01	0.01	0.01	0.01	0.00
	H.larvae	0.84	0.70	0.72	0.77	0.72	0.72	0.02	0.02	0.01	0.01	0.02	0.01
	H.eggs	0.49	0.49	0.40	0.36	0.50	0.51	0.01	0.02	0.01	0.00	0.01	0.02
	P.cyscts	0.36	0.22	0.42	0.44	0.46	0.41	0.02	0.01	0.01	0.02	0.02	0.02
KPCA	H.eggs*	0.19	0.14	0.19	0.16	0.16	0.16	0.02	0.02	0.02	0.01	0.02	0.02
	P.cysts*	0.25	0.30	0.24	0.22	0.30	0.16	0.02	0.01	0.01	0.01	0.02	0.00
	H.larvae	0.84	0.71	0.74	0.84	0.70	0.71	0.01	0.02	0.03	0.03	0.00	0.01
	H.eggs	0.55	0.45	0.42	0.50	0.57	0.49	0.01	0.02	0.02	0.01	0.02	0.02
	P.cyscts	0.44	0.37	0.42	0.32	0.49	0.37	0.01	0.01	0.01	0.01	0.01	0.01
LLE	H.eggs*	0.18	0.16	0.19	0.17	0.19	0.15	0.01	0.01	0.02	0.02	0.03	0.00
	P.cysts*	0.22	0.26	0.25	0.24	0.24	0.22	0.01	0.01	0.01	0.01	0.01	0.01
	H.larvae	0.83	0.68	0.71	0.82	0.69	0.85	0.01	0.02	0.02	0.02	-0.00	0.01
	H.eggs	0.46	0.49	0.43	0.38	0.47	0.49	0.02	0.02	0.01	0.01	0.02	0.02
	P.cyscts	0.43	0.41	0.41	0.53	0.38	0.42	0.01	0.01	0.01	0.02	0.01	0.02
MDS	H.eggs*	0.83	0.68	0.83	0.85	0.71	0.86	0.80	0.61	0.79	0.82	0.65	0.84
	P.cysts*	0.63	0.70	0.63	0.55	0.64	0.56	0.49	0.60	0.49	0.41	0.51	0.42
	H.larvae	0.90	0.91	0.94	0.95	0.95	0.94	0.53	0.53	0.67	0.80	0.75	0.72
	H.eggs	0.68	0.51	0.70	0.71	0.62	0.77	0.46	0.22	0.48	0.50	0.33	0.58
	P.cyscts	0.58	0.51	0.61	0.62	0.62	0.68	0.32	0.24	0.38	0.38	0.36	0.49
MLLE	H.eggs*	0.15	0.16	0.16	0.17	0.17	0.16	0.00	0.02	0.02	0.01	0.03	0.02
	P.cysts*	0.26	0.24	0.26	0.23	0.25	0.23	0.01	0.01	0.01	0.00	0.02	0.01
	H.larvae	0.84	0.76	0.79	0.73	0.86	0.62	0.01	0.01	0.03	0.02	0.03	0.01
	H.eggs	0.38	0.54	0.45	0.39	0.53	0.52	0.01	0.02	0.02	0.01	0.01	0.01
	P.cyscts	0.33	0.40	0.43	0.43	0.36	0.38	0.01	0.01	0.02	0.02	0.01	0.01
PCA	H.eggs*	0.78	0.58	0.80	0.86	0.57	0.84	0.73	0.49	0.76	0.83	0.49	0.81
	P.cysts*	0.57	0.70	0.62	0.59	0.57	0.59	0.44	0.60	0.48	0.44	0.42	0.44
	H.larvae	0.91	0.91	0.93	0.92	0.95	0.96	0.59	0.56	0.67	0.68	0.75	0.81
	H.eggs	0.59	0.55	0.61	0.70	0.61	0.78	0.34	0.20	0.35	0.49	0.30	0.60
	P.cyscts	0.53	0.51	0.62	0.54	0.58	0.67	0.25	0.24	0.36	0.27	0.31	0.48
t-SNE	H.eggs*	0.90	0.84	0.93	0.82	0.93	0.90	0.89	0.80	0.92	0.79	0.92	0.88
	P.cysts*	0.73	0.67	0.69	0.54	0.71	0.57	0.63	0.56	0.58	0.40	0.61	0.44
	H.larvae	0.91	0.89	0.93	0.96	0.96	0.96	0.57	0.56	0.67	0.81	0.80	0.80
	H.eggs	0.70	0.57	0.73	0.77	0.67	0.78	0.52	0.25	0.56	0.61	0.44	0.64
	P.cyscts	0.63	0.54	0.68	0.64	0.62	0.67	0.39	0.26	0.48	0.39	0.37	0.46
UMAP	H.eggs*	0.88	0.81	0.91	0.82	0.93	0.91	0.86	0.78	0.89	0.79	0.91	0.89
	P.cysts*	0.68	0.66	0.64	0.53	0.68	0.56	0.57	0.55	0.52	0.38	0.56	0.41
	H.larvae	0.91	0.88	0.93	0.96	0.95	0.96	0.58	0.50	0.71	0.83	0.76	0.80
	H.eggs	0.70	0.58	0.74	0.74	0.66	0.78	0.49	0.22	0.56	0.57	0.41	0.64
	P.cyscts	0.64	0.52	0.68	0.63	0.62	0.69	0.37	0.22	0.47	0.39	0.39	0.49

C3: Classification results

		accuracy						κ					
		from scratch			pre-trained			from scratch			pre-trained		
		SimCLR	SupCon	SimCLR+ SupCon	SimCLR	SupCon	SimCLR+ SupCon	SimCLR	SupCon	SimCLR+ SupCon	SimCLR	SupCon	SimCLR+ SupCon
FA	H.eggs*	0.79	0.57	0.82	0.83	0.66	0.80	0.75	0.48	0.78	0.79	0.59	0.76
	P.cysts*	0.57	0.68	0.59	0.66	0.59	0.65	0.44	0.59	0.48	0.53	0.43	0.52
	H.larvae	0.92	0.89	0.95	0.89	0.95	0.95	0.65	0.44	0.72	0.62	0.73	0.76
	H.eggs	0.71	0.58	0.66	0.73	0.67	0.78	0.50	0.23	0.42	0.53	0.36	0.60
	P.cyscts	0.65	0.57	0.70	0.64	0.65	0.74	0.31	0.24	0.43	0.38	0.35	0.56
FICA	H.eggs*	0.21	0.17	0.27	0.16	0.16	0.26	0.00	0.01	0.09	0.00	-0.00	0.08
	P.cysts*	0.32	0.39	0.39	0.33	0.25	0.30	0.00	0.00	0.00	-0.01	-0.00	-0.01
	H.larvae	0.87	0.87	0.77	0.87	0.84	0.87	0.00	0.00	0.02	0.00	0.08	0.01
	H.eggs	0.65	0.65	0.64	0.65	0.47	0.65	-0.00	-0.01	-0.01	-0.01	0.00	-0.00
	P.cyscts	0.60	0.58	0.60	0.59	0.41	0.46	-0.00	0.00	0.00	-0.01	-0.00	-0.01
ISO	H.eggs*	0.25	0.17	0.17	0.16	0.12	0.14	0.06	0.00	0.01	0.01	-0.00	0.02
	P.cysts*	0.33	0.25	0.32	0.17	0.14	0.18	0.01	0.00	0.00	0.00	-0.00	0.00
	H.larvae	0.87	0.66	0.82	0.87	0.74	0.82	0.01	-0.02	0.03	0.03	0.02	0.05
	H.eggs	0.65	0.58	0.64	0.62	0.64	0.65	0.00	0.02	-0.01	0.00	-0.01	-0.00
	P.cyscts	0.58	0.10	0.60	0.60	0.60	0.60	-0.00	0.01	-0.00	-0.00	-0.00	-0.00
KPCA	H.eggs*	0.19	0.11	0.21	0.13	0.16	0.16	0.00	0.00	0.01	-0.02	-0.00	0.03
	P.cysts*	0.31	0.39	0.25	0.19	0.32	0.05	-0.03	0.00	0.00	0.00	0.00	-0.00
	H.larvae	0.87	0.78	0.85	0.87	0.83	0.81	0.00	0.04	0.10	0.01	0.07	0.04
	H.eggs	0.65	0.50	0.64	0.65	0.65	0.65	-0.00	0.01	-0.01	-0.00	-0.00	-0.00
	P.cyscts	0.60	0.43	0.60	0.41	0.60	0.60	-0.00	0.00	-0.00	-0.00	-0.00	-0.00
LLE	H.eggs*	0.19	0.15	0.20	0.19	0.23	0.15	0.00	0.00	0.00	0.02	0.05	0.00
	P.cysts*	0.17	0.26	0.36	0.21	0.25	0.18	0.00	0.01	0.08	0.00	0.00	0.00
	H.larvae	0.87	0.76	0.77	0.87	0.81	0.87	0.01	0.03	0.03	0.03	0.04	0.00
	H.eggs	0.53	0.65	0.62	0.48	0.62	0.65	0.01	0.02	-0.01	0.01	0.00	-0.00
	P.cyscts	0.60	0.43	0.59	0.60	0.58	0.47	0.00	0.00	-0.01	0.00	0.05	0.01
MDS	H.eggs*	0.87	0.70	0.84	0.87	0.75	0.89	0.84	0.64	0.81	0.84	0.70	0.87
	P.cysts*	0.66	0.73	0.68	0.64	0.69	0.62	0.54	0.64	0.56	0.52	0.57	0.49
	H.larvae	0.90	0.91	0.94	0.96	0.95	0.94	0.53	0.51	0.67	0.83	0.73	0.70
	H.eggs	0.74	0.56	0.76	0.74	0.65	0.79	0.53	0.25	0.55	0.53	0.34	0.60
	P.cyscts	0.64	0.54	0.67	0.68	0.68	0.74	0.36	0.29	0.46	0.45	0.44	0.57
MLLE	H.eggs*	0.18	0.19	0.15	0.23	0.13	0.17	0.03	0.03	0.00	0.04	0.00	0.04
	P.cysts*	0.32	0.25	0.39	0.20	0.24	0.19	0.00	0.00	0.01	0.00	0.00	0.00
	H.larvae	0.87	0.87	0.87	0.85	0.87	0.87	0.01	0.08	0.03	0.12	0.00	0.24
	H.eggs	0.62	0.65	0.64	0.64	0.65	0.65	-0.02	-0.00	-0.01	0.01	-0.00	-0.00
	P.cyscts	0.58	0.59	0.60	0.59	0.53	0.60	-0.01	-0.01	-0.00	-0.00	0.02	-0.00
PCA	H.eggs*	0.78	0.58	0.83	0.87	0.59	0.87	0.74	0.48	0.79	0.84	0.52	0.84
	P.cysts*	0.62	0.72	0.64	0.68	0.63	0.68	0.50	0.63	0.52	0.56	0.47	0.55
	H.larvae	0.91	0.91	0.93	0.92	0.95	0.96	0.57	0.56	0.64	0.70	0.74	0.81
	H.eggs	0.67	0.62	0.66	0.74	0.65	0.81	0.43	0.26	0.41	0.54	0.32	0.65
	P.cyscts	0.62	0.53	0.67	0.66	0.66	0.72	0.28	0.26	0.40	0.34	0.40	0.54
t-SNE	H.eggs*	0.93	0.85	0.96	0.85	0.95	0.94	0.92	0.82	0.96	0.82	0.95	0.93
	P.cysts*	0.77	0.70	0.73	0.64	0.76	0.66	0.69	0.59	0.63	0.53	0.67	0.54
	H.larvae	0.91	0.90	0.93	0.96	0.95	0.95	0.60	0.60	0.71	0.81	0.79	0.79
	H.eggs	0.76	0.64	0.78	0.82	0.73	0.82	0.60	0.30	0.62	0.69	0.50	0.69
	P.cyscts	0.70	0.59	0.74	0.72	0.70	0.72	0.48	0.31	0.56	0.48	0.46	0.53
UMAP	H.eggs*	0.91	0.85	0.94	0.85	0.95	0.95	0.90	0.82	0.93	0.82	0.94	0.94
	P.cysts*	0.73	0.69	0.67	0.60	0.70	0.64	0.62	0.59	0.55	0.48	0.59	0.50
	H.larvae	0.91	0.88	0.94	0.96	0.95	0.96	0.60	0.50	0.72	0.83	0.74	0.81
	H.eggs	0.76	0.64	0.78	0.79	0.71	0.81	0.56	0.26	0.62	0.64	0.45	0.68
	P.cyscts	0.68	0.58	0.74	0.69	0.70	0.74	0.41	0.25	0.57	0.45	0.50	0.57

Fig. 5. *"C2: Propagation results"* and *"C3: Classification results"*. For the proposed experiments, results of accuracy and κ are shown for pseudolabeling using 10 projection techniques P on SimCLR's, SupCon's, and SimCLR+SupCon's latent spaces trained from scratch or with pre-trained weights for five datasets.

7 Discussion

We next discuss the main findings emerging from our results.

Table 6. C3: VGG-16's classification results on T when using pseudo labels from SimCLR's, SupCon and SimCLR+SupCon latent spaces, from scratch and with ImageNet pre-trained weights. Best values per dataset are in bold.

techniques	datasets	metric	baseline	from scratch			pre-trained		
				SimCLR	SupCon	SimCLR+SupCon	SimCLR	SupCon	SimCLR+SupCon
FA	H.eggs*	acc	0.812932 ± 0.059	0.791588 ± 0.071	0.569994 ± 0.077	0.816070 ± 0.037	**0.826114 ± 0.043**	0.660389 ± 0.087	0.797238 ± 0.074
		κ	0.775954 ± 0.073	0.750532 ± 0.086	0.481551 ± 0.096	0.780221 ± 0.044	**0.790394 ± 0.052**	0.589562 ± 0.106	0.758894 ± 0.088
	P.cysts*	acc	**0.757209 ± 0.015**	0.574395 ± 0.132	0.684544 ± 0.056	0.590542 ± 0.015	0.663783 ± 0.037	0.586217 ± 0.049	0.648212 ± 0.033
		κ	**0.651933 ± 0.023**	0.435072 ± 0.157	0.585071 ± 0.064	0.482646 ± 0.019	0.533400 ± 0.052	0.426891 ± 0.093	0.516758 ± 0.055
	H.larvae	acc	0.930806 ± 0.026	0.917535 ± 0.033	0.889099 ± 0.015	0.948183 ± 0.018	0.891627 ± 0.057	0.947551 ± 0.010	**0.951343 ± 0.014**
		κ	0.613432 ± 0.230	0.645066 ± 0.129	0.441950 ± 0.135	0.723843 ± 0.126	0.624029 ± 0.171	0.731107 ± 0.071	**0.758448 ± 0.094**
	H.eggs	acc	**0.862234 ± 0.010**	0.708822 ± 0.015	0.582790 ± 0.047	0.655367 ± 0.035	0.730552 ± 0.010	0.667753 ± 0.058	0.782269 ± 0.057
		κ	**0.740861 ± 0.028**	0.500625 ± 0.016	0.230899 ± 0.071	0.415120 ± 0.056	0.525028 ± 0.024	0.358934 ± 0.112	0.602154 ± 0.095
	P.cysts	acc	**0.850691 ± 0.018**	0.650528 ± 0.010	0.569604 ± 0.016	0.698595 ± 0.039	0.638221 ± 0.013	0.646697 ± 0.024	0.736445 ± 0.012
		κ	**0.751667 ± 0.028**	0.312113 ± 0.026	0.241764 ± 0.048	0.430761 ± 0.136	0.380899 ± 0.060	0.345129 ± 0.061	0.556038 ± 0.038
FICA	H.eggs*	acc	**0.812932 ± 0.059**	0.209040 ± 0.010	0.165725 ± 0.075	0.271186 ± 0.088	0.161331 ± 0.008	0.161958 ± 0.058	0.259887 ± 0.056
		κ	**0.775954 ± 0.073**	0.002840 ± 0.004	0.009936 ± 0.022	0.092821 ± 0.104	0.002707 ± 0.037	-0.001412 ± 0.005	0.077978 ± 0.066
	P.cysts*	acc	**0.757209 ± 0.015**	0.315744 ± 0.106	0.390138 ± 0.000	0.388408 ± 0.002	0.333045 ± 0.081	0.247981 ± 0.101	0.299884 ± 0.085
		κ	**0.651933 ± 0.023**	0.000569 ± 0.001	0.000000 ± 0.000	0.002938 ± 0.004	-0.008047 ± 0.011	-0.000579 ± 0.001	-0.010702 ± 0.015
	H.larvae	acc	**0.930806 ± 0.026**	0.871722 ± 0.002	0.872986 ± 0.000	0.769668 ± 0.141	0.872670 ± 0.000	0.841706 ± 0.039	0.872670 ± 0.001
		κ	**0.613432 ± 0.230**	0.001145 ± 0.002	0.000000 ± 0.000	0.024169 ± 0.009	0.003042 ± 0.006	0.080839 ± 0.073	0.006713 ± 0.005
	H.eggs	acc	**0.862234 ± 0.010**	0.649935 ± 0.005	0.647762 ± 0.004	0.644937 ± 0.009	0.646893 ± 0.005	0.467623 ± 0.257	0.651456 ± 0.003
		κ	**0.740861 ± 0.028**	-0.004383 ± 0.005	-0.005765 ± 0.003	-0.007738 ± 0.012	-0.005335 ± 0.004	0.000909 ± 0.006	-0.001991 ± 0.002
	P.cysts	acc	**0.850691 ± 0.018**	0.595147 ± 0.003	0.584814 ± 0.015	0.597353 ± 0.000	0.592825 ± 0.006	0.408452 ± 0.267	0.462441 ± 0.191
		κ	**0.751667 ± 0.028**	-0.001485 ± 0.002	0.003837 ± 0.010	0.000000 ± 0.000	-0.005944 ± 0.008	-0.002462 ± 0.003	-0.008557 ± 0.012
ISO	H.eggs*	acc	**0.812932 ± 0.059**	0.248588 ± 0.062	0.173886 ± 0.028	0.173258 ± 0.033	0.164470 ± 0.058	0.118644 ± 0.051	0.137477 ± 0.050
		κ	**0.775954 ± 0.073**	0.057848 ± 0.080	0.003443 ± 0.003	0.010141 ± 0.012	0.009016 ± 0.013	-0.000439 ± 0.000	0.018088 ± 0.012
	P.cysts*	acc	**0.757209 ± 0.015**	0.326701 ± 0.098	0.247981 ± 0.101	0.315455 ± 0.106	0.173299 ± 0.010	0.141003 ± 0.065	0.180219 ± 0.010
		κ	**0.651933 ± 0.023**	0.008227 ± 0.007	0.000000 ± 0.000	0.000000 ± 0.000	0.000000 ± 0.000	-0.000103 ± 0.000	0.000000 ± 0.000
	H.larvae	acc	**0.930806 ± 0.026**	0.872986 ± 0.000	0.662243 ± 0.298	0.822117 ± 0.072	0.872670 ± 0.001	0.740916 ± 0.189	0.822749 ± 0.072
		κ	**0.613432 ± 0.230**	0.014444 ± 0.013	-0.015757 ± 0.030	0.031433 ± 0.044	0.031303 ± 0.015	0.016744 ± 0.020	0.048051 ± 0.059
	H.eggs	acc	**0.862234 ± 0.010**	0.649283 ± 0.006	0.577357 ± 0.107	0.642329 ± 0.009	0.621252 ± 0.044	0.643850 ± 0.013	0.650152 ± 0.004
		κ	**0.740861 ± 0.028**	0.003225 ± 0.005	0.020891 ± 0.030	-0.006907 ± 0.003	0.001426 ± 0.003	-0.006799 ± 0.010	-0.003234 ± 0.004
	P.cysts	acc	**0.850691 ± 0.018**	0.580634 ± 0.023	0.099036 ± 0.006	0.596424 ± 0.001	0.596424 ± 0.001	0.597237 ± 0.000	0.596308 ± 0.001
		κ	**0.751667 ± 0.028**	-0.003610 ± 0.005	0.009051 ± 0.002	-0.001265 ± 0.001	-0.001084 ± 0.002	-0.000161 ± 0.000	-0.001397 ± 0.001
KPCA	H.eggs*	acc	**0.812932 ± 0.059**	0.193974 ± 0.003	0.112994 ± 0.031	0.209040 ± 0.010	0.131199 ± 0.054	0.157564 ± 0.063	0.163214 ± 0.059
		κ	**0.775954 ± 0.073**	0.000112 ± 0.000	0.000674 ± 0.001	0.012136 ± 0.016	-0.018456 ± 0.013	-0.000097 ± 0.000	0.027955 ± 0.052
	P.cysts*	acc	**0.757209 ± 0.015**	0.305651 ± 0.119	0.390138 ± 0.000	0.249423 ± 0.100	0.187139 ± 0.000	0.322376 ± 0.096	0.051038 ± 0.002
		κ	**0.651933 ± 0.023**	-0.025975 ± 0.037	0.000000 ± 0.000	0.003966 ± 0.005	0.000000 ± 0.000	0.000000 ± 0.000	-0.002450 ± 0.003
	H.larvae	acc	**0.930806 ± 0.026**	0.872670 ± 0.001	0.780411 ± 0.132	0.848657 ± 0.036	0.872354 ± 0.002	0.825908 ± 0.070	0.807899 ± 0.091
		κ	**0.613432 ± 0.230**	0.003063 ± 0.007	0.037027 ± 0.027	0.104310 ± 0.066	0.013242 ± 0.010	0.066086 ± 0.022	0.040867 ± 0.052
	H.eggs	acc	**0.862234 ± 0.010**	0.650804 ± 0.004	0.500217 ± 0.217	0.644285 ± 0.007	0.651021 ± 0.002	0.651890 ± 0.002	0.652977 ± 0.001
		κ	**0.740861 ± 0.028**	-0.003327 ± 0.004	0.006654 ± 0.010	-0.008322 ± 0.006	-0.001763 ± 0.002	-0.001528 ± 0.001	-0.001004 ± 0.001
	P.cysts	acc	**0.850691 ± 0.018**	0.596772 ± 0.001	0.427029 ± 0.240	0.596773 ± 0.001	0.411471 ± 0.260	0.597237 ± 0.000	0.597005 ± 0.000
		κ	**0.751667 ± 0.028**	-0.000805 ± 0.001	0.001823 ± 0.003	-0.000298 ± 0.000	-0.000119 ± 0.002	-0.000161 ± 0.000	-0.000483 ± 0.001
LLE	H.eggs*	acc	**0.812932 ± 0.059**	0.192090 ± 0.003	0.152542 ± 0.075	0.199623 ± 0.010	0.185185 ± 0.028	0.227872 ± 0.045	0.151287 ± 0.075
		κ	**0.775954 ± 0.073**	0.000017 ± 0.000	0.002360 ± 0.002	0.000179 ± 0.000	0.018815 ± 0.017	0.054332 ± 0.071	0.000028 ± 0.000
	P.cysts*	acc	**0.757209 ± 0.015**	0.173299 ± 0.010	0.262976 ± 0.090	0.355824 ± 0.126	0.209054 ± 0.140	0.247981 ± 0.101	0.180219 ± 0.010
		κ	**0.651933 ± 0.023**	0.000000 ± 0.000	0.006750 ± 0.010	0.080931 ± 0.115	0.000000 ± 0.000	0.000000 ± 0.000	0.000000 ± 0.000
	H.larvae	acc	**0.930806 ± 0.026**	0.872354 ± 0.002	0.763349 ± 0.154	0.774723 ± 0.141	0.870774 ± 0.006	0.806319 ± 0.093	0.872986 ± 0.000
		κ	**0.613432 ± 0.230**	0.009639 ± 0.007	0.034258 ± 0.030	0.027194 ± 0.019	0.027385 ± 0.021	0.035078 ± 0.027	0.000000 ± 0.000
	H.eggs	acc	**0.862234 ± 0.010**	0.533681 ± 0.165	0.645589 ± 0.009	0.621903 ± 0.041	0.478705 ± 0.225	0.617557 ± 0.051	0.650804 ± 0.002
		κ	**0.740861 ± 0.028**	0.005994 ± 0.013	0.019156 ± 0.029	-0.006981 ± 0.005	0.008658 ± 0.022	0.001140 ± 0.002	-0.002347 ± 0.002
	P.cysts	acc	**0.850691 ± 0.018**	0.597353 ± 0.000	0.427145 ± 0.240	0.589690 ± 0.011	0.597353 ± 0.000	0.577383 ± 0.016	0.473006 ± 0.176
		κ	**0.751667 ± 0.028**	0.000000 ± 0.000	0.000985 ± 0.002	-0.008861 ± 0.012	0.000303 ± 0.001	0.054789 ± 0.067	0.014402 ± 0.019
MDS	H.eggs*	acc	0.812932 ± 0.059	0.870057 ± 0.007	0.698054 ± 0.053	0.839924 ± 0.030	0.868801 ± 0.056	0.750785 ± 0.044	**0.889517 ± 0.009**
		κ	0.775954 ± 0.073	0.844017 ± 0.009	0.635306 ± 0.066	0.806060 ± 0.038	0.843305 ± 0.067	0.696556 ± 0.054	**0.867837 ± 0.011**
	P.cysts*	acc	**0.757209 ± 0.015**	0.660035 ± 0.042	0.729815 ± 0.018	0.679642 ± 0.008	0.675026 ± 0.019	0.688004 ± 0.049	0.623414 ± 0.047
		κ	**0.651933 ± 0.023**	0.535152 ± 0.066	0.636650 ± 0.017	0.558141 ± 0.010	0.452008 ± 0.046	0.574029 ± 0.067	0.485190 ± 0.040
	H.larvae	acc	0.930806 ± 0.026	0.896366 ± 0.010	0.907425 ± 0.021	0.941864 ± 0.022	**0.961137 ± 0.003**	0.948183 ± 0.014	0.944392 ± 0.020
		κ	0.613432 ± 0.230	0.527845 ± 0.047	0.512390 ± 0.132	0.670216 ± 0.171	**0.828970 ± 0.016**	0.732767 ± 0.094	0.702961 ± 0.166
	H.eggs	acc	**0.862234 ± 0.010**	0.735767 ± 0.006	0.556280 ± 0.010	0.755324 ± 0.023	0.736636 ± 0.040	0.649283 ± 0.049	0.787701 ± 0.039
		κ	**0.740861 ± 0.028**	0.527193 ± 0.052	0.250879 ± 0.023	0.552969 ± 0.033	0.529298 ± 0.057	0.342224 ± 0.071	0.598315 ± 0.075
	P.cysts	acc	**0.850691 ± 0.018**	0.640660 ± 0.022	0.541159 ± 0.012	0.672820 ± 0.066	0.639273 ± 0.041	0.684663 ± 0.022	0.742947 ± 0.021
		κ	**0.751667 ± 0.028**	0.360158 ± 0.031	0.293691 ± 0.017	0.460289 ± 0.091	0.516147 ± 0.048	0.437128 ± 0.046	0.572537 ± 0.027

continued

Table 6. continued

techniques	datasets	metric	baseline	from scratch			pre-trained		
				SimCLR	SupCon	SimCLR+SupCon	SimCLR	SupCon	SimCLR+SupCon
MLLE	H.eggs*	acc	**0.812932 ± 0.059**	0.176397 ± 0.069	0.186440 ± 0.007	0.149404 ± 0.053	0.231638 ± 0.028	0.133710 ± 0.051	0.172630 ± 0.061
		κ	**0.775954 ± 0.073**	0.029830 ± 0.024	0.030248 ± 0.055	0.003127 ± 0.004	0.038996 ± 0.038	0.000489 ± 0.001	0.039277 ± 0.051
	P.cysts*	acc	**0.757209 ± 0.015**	0.322664 ± 0.095	0.254902 ± 0.096	0.394175 ± 0.005	0.198673 ± 0.015	0.240773 ± 0.106	0.187139 ± 0.000
		κ	**0.651933 ± 0.023**	0.000096 ± 0.000	0.000000 ± 0.000	0.009169 ± 0.009	0.001291 ± 0.006	0.000000 ± 0.000	0.000000 ± 0.000
	H.larvae	acc	**0.930806 ± 0.026**	0.873618 ± 0.001	0.872670 ± 0.003	0.870774 ± 0.002	0.854029 ± 0.032	0.872986 ± 0.000	0.870774 ± 0.007
		κ	**0.613432 ± 0.230**	0.008591 ± 0.012	0.080602 ± 0.071	0.027161 ± 0.028	0.121338 ± 0.009	0.000000 ± 0.000	0.236101 ± 0.148
	H.eggs	acc	**0.862234 ± 0.010**	0.619513 ± 0.024	0.651456 ± 0.002	0.638418 ± 0.007	0.636679 ± 0.012	0.653629 ± 0.000	0.651456 ± 0.001
		κ	**0.740861 ± 0.028**	-0.019102 ± 0.011	-0.002446 ± 0.003	-0.010811 ± 0.001	0.006862 ± 0.024	-0.000263 ± 0.000	-0.001735 ± 0.000
	P.cysts	acc	**0.850691 ± 0.018**	0.583769 ± 0.015	0.585743 ± 0.016	0.596889 ± 0.000	0.592128 ± 0.007	0.528387 ± 0.095	0.596076 ± 0.001
		κ	**0.751667 ± 0.028**	-0.013466 ± 0.014	-0.008483 ± 0.012	-0.000645 ± 0.001	-0.003423 ± 0.005	0.019087 ± 0.030	-0.001528 ± 0.001
PCA	H.eggs*	acc	0.812932 ± 0.059	0.782800 ± 0.022	0.575644 ± 0.086	0.827997 ± 0.033	0.865035 ± 0.065	0.593220 ± 0.072	**0.868801 ± 0.027**
		κ	0.775954 ± 0.073	0.739026 ± 0.028	0.483446 ± 0.110	0.792577 ± 0.041	0.837979 ± 0.079	0.518984 ± 0.080	**0.844098 ± 0.031**
	P.cysts*	acc	**0.757209 ± 0.015**	0.619666 ± 0.083	0.719147 ± 0.025	0.641580 ± 0.059	0.677047 ± 0.019	0.626586 ± 0.027	0.679354 ± 0.048
		κ	**0.651933 ± 0.023**	0.496203 ± 0.104	0.632033 ± 0.030	0.520132 ± 0.069	0.556229 ± 0.027	0.474462 ± 0.041	0.551953 ± 0.072
	H.larvae	acc	0.930806 ± 0.026	0.906793 ± 0.005	0.912480 ± 0.029	0.931438 ± 0.019	0.923539 ± 0.026	0.949447 ± 0.014	**0.957662 ± 0.002**
		κ	0.613432 ± 0.230	0.567678 ± 0.037	0.557973 ± 0.171	0.644459 ± 0.162	0.696070 ± 0.105	0.737386 ± 0.095	**0.807875 ± 0.003**
	H.eggs	acc	**0.862234 ± 0.010**	0.668622 ± 0.046	0.618644 ± 0.054	0.664711 ± 0.029	0.744242 ± 0.027	0.647327 ± 0.066	0.812690 ± 0.032
		κ	**0.740861 ± 0.028**	0.433312 ± 0.065	0.260372 ± 0.051	0.411077 ± 0.062	0.543626 ± 0.048	0.321544 ± 0.130	0.647622 ± 0.057
	P.cysts	acc	**0.850691 ± 0.018**	0.615000 ± 0.017	0.531638 ± 0.005	0.666899 ± 0.031	0.658307 ± 0.032	0.664229 ± 0.001	0.720771 ± 0.027
		κ	**0.751667 ± 0.028**	0.284629 ± 0.104	0.263112 ± 0.026	0.395488 ± 0.115	0.344865 ± 0.153	0.403226 ± 0.034	0.535129 ± 0.032
t-SNE	H.eggs*	acc	0.812932 ± 0.059	0.932831 ± 0.032	0.846202 ± 0.025	**0.964846 ± 0.022**	0.850596 ± 0.051	0.954802 ± 0.023	0.938481 ± 0.017
		κ	0.775954 ± 0.073	0.919898 ± 0.038	0.816217 ± 0.030	**0.958337 ± 0.026**	0.822694 ± 0.059	0.946002 ± 0.027	0.927126 ± 0.020
	P.cysts*	acc	0.757209 ± 0.015	**0.771626 ± 0.050**	0.697232 ± 0.020	0.732122 ± 0.023	0.644464 ± 0.047	0.756920 ± 0.012	0.658881 ± 0.028
		κ	0.651933 ± 0.023	**0.688570 ± 0.066**	0.594524 ± 0.035	0.633941 ± 0.042	0.534253 ± 0.057	0.665824 ± 0.024	0.544564 ± 0.034
	H.larvae	acc	0.930806 ± 0.026	0.910901 ± 0.013	0.902685 ± 0.029	0.933965 ± 0.017	**0.956714 ± 0.003**	0.953871 ± 0.003	0.952291 ± 0.004
		κ	0.613432 ± 0.230	0.599203 ± 0.111	0.595899 ± 0.123	0.705463 ± 0.079	**0.810965 ± 0.018**	0.785615 ± 0.022	0.787406 ± 0.014
	H.eggs	acc	**0.862234 ± 0.010**	0.761191 ± 0.027	0.636028 ± 0.025	0.777705 ± 0.013	0.824641 ± 0.003	0.725772 ± 0.070	0.819426 ± 0.024
		κ	**0.740861 ± 0.028**	0.599112 ± 0.036	0.299141 ± 0.077	0.621008 ± 0.032	0.685570 ± 0.014	0.500910 ± 0.118	0.686552 ± 0.042
	P.cysts	acc	**0.850691 ± 0.018**	0.698711 ± 0.003	0.592476 ± 0.025	0.735748 ± 0.019	0.719378 ± 0.007	0.701614 ± 0.011	0.719959 ± 0.032
		κ	**0.751667 ± 0.028**	0.484463 ± 0.017	0.313296 ± 0.022	0.559176 ± 0.038	0.483280 ± 0.034	0.460516 ± 0.054	0.527912 ± 0.030
UMAP	H.eggs*	acc	0.812932 ± 0.059	0.913999 ± 0.039	0.846202 ± 0.055	0.937225 ± 0.033	0.849341 ± 0.088	**0.950408 ± 0.024**	0.945386 ± 0.004
		κ	0.775954 ± 0.073	0.897254 ± 0.047	0.817901 ± 0.065	0.925412 ± 0.039	0.820393 ± 0.105	**0.940671 ± 0.028**	0.935039 ± 0.005
	P.cysts*	acc	**0.757209 ± 0.015**	0.728662 ± 0.044	0.686851 ± 0.023	0.666667 ± 0.040	0.601499 ± 0.098	0.702422 ± 0.022	0.636678 ± 0.004
		κ	**0.651933 ± 0.023**	0.621347 ± 0.063	0.590353 ± 0.021	0.547625 ± 0.062	0.480729 ± 0.113	0.590077 ± 0.043	0.500455 ± 0.014
	H.larvae	acc	0.930806 ± 0.026	0.911848 ± 0.010	0.883096 ± 0.034	0.935861 ± 0.015	**0.959242 ± 0.006**	0.947867 ± 0.013	0.958610 ± 0.005
		κ	0.613432 ± 0.230	0.601489 ± 0.075	0.503848 ± 0.098	0.718764 ± 0.056	**0.826163 ± 0.025**	0.739471 ± 0.099	0.805239 ± 0.026
	H.eggs	acc	**0.862234 ± 0.010**	0.755107 ± 0.009	0.644285 ± 0.019	0.781399 ± 0.045	0.790960 ± 0.025	0.707084 ± 0.063	0.814863 ± 0.036
		κ	**0.740861 ± 0.028**	0.563245 ± 0.010	0.261623 ± 0.039	0.616187 ± 0.079	0.635362 ± 0.039	0.449519 ± 0.148	0.680416 ± 0.051
	P.cysts	acc	**0.850691 ± 0.018**	0.684547 ± 0.012	0.575061 ± 0.010	0.743295 ± 0.036	0.691165 ± 0.025	0.703819 ± 0.017	0.741089 ± 0.014
		κ	**0.751667 ± 0.028**	0.408555 ± 0.082	0.251174 ± 0.026	0.565152 ± 0.073	0.445321 ± 0.084	0.500526 ± 0.021	0.566037 ± 0.016

7.1 Data Separation *vs* Visual Separation Depends on the Projection Technique

Figures 4 and 5(left) give us two main insights.

First, we see a similar pattern in κ (lighter/yellow or darker/purple colors) for distinct *datasets*. DS and VS are correlated for some projections P, *e.g.*, MDS, t-SNE, and UMAP; and somewhat less for FA and PCA. For other projections, *e.g.*, FICA, ISO, KPCA, LLE, and MLLE, this correlation is absent. For these last projections, while there is some variation in κ for distinct datasets in DS, VS is close to 0 for *all* datasets. This tells that some projections can map DS to VS quite well, whereas others cannot and tend to create a low VS no matter how high DS is. In short, VS *strongly depends* on the projection technique P.

Secondly, we see a distinct pattern for those projections that capture well DS in their VS. Datasets with medium-high DS (accuracy in $[0.8, 1]$, κ in $[0.7, 1.00]$) tends to yield also medium-high VS accuracy and κ (compare Figs. 4 and 5). *H.eggs** and *H.larvae* are examples of these datasets. We see a similar pattern for low-medium accuracy and κ:

For DS accuracy in $[0, 0.5]$ and κ in $[0, 0.4]$, we get medium-high accuracy and κ for VS (see Fig. 5 for *P.cysts**, *P.cysts*, or *H.eggs*). This tells that some projection techniques can not only correlate VS with DS, but also keep differences for distinct datasets, while others do not.

7.2 Assessing the Quality of Visual Separation

As explained earlier, we use κ as a measure of visual separation of (same-labeled) points in a projection. While this is arguably intuitive – one can propagate labels easier when surrounding points are unlabeled or have the same label than when surrounding points would have many different labels – we would like to directly test how κ and VS *as perceived by humans* agree.

To assess the correlation between κ and perceived VS, we ranked our results as follows. We computed the average accuracy and κ for each P (over all contrastive learning methods, initialization strategies, and datasets). Next, we sorted the projections P over accuracy and κ (see Table 7). Finally, we show in Fig. 6 the actual projection results for the best, medium, and worst projection following the above ranking.

The best-ranked projection in Fig. 6 is t-SNE. In line with this ranking, we indeed see a quite good separation of points having a label from those having different labels (or gray, *i.e.*, unlabeled). For *H.eggs**, all three latent space projections (SimCLR, SupCon, and SimCLR+SupCon) show a clear VS, and we see that this leads to almost no color mixing in the propagated pseudo-labels. For *P.cysts**, there is a clearly separated group (red) in all three projections which also has a single color (label). The remaining projections, which have no clear VS in terms of distinct groups, show a mix of different colors. For *H.larvae*, the larvae class (red) is better separated from the big group of impurities (gray), and this correlates with the larvae samples being all located in a tail-like periphery of the projection – thus, better visually separated from the rest. For *H.eggs*, we see how the visually separated groups show almost no color mixing, whereas the parts of the projection where no VS is present show color mixing. For *P.cysts*, the projections have even less VS, and we see how labels get even more mixed – for instance, the impurity class (gray) is spread all over the projection.

The medium-ranked projection technique in Fig. 6 is FA. Its scatterplots show a less clear correlation between VS and lack of label mixing in distinct groups. For *H.eggs**, we notice some VS for SimCLR and SupCon. Some groups are better clustered (red and yellow) than others (brown and gray), but with few whitespace among those groups. No clear VS can be seen for all other datasets and contrastive learning approaches – the points are condensed in a single group with similar colors close to each other in this group.

Finally, LLE scored as the worst method in our ranking. Figure 6 shows, indeed, no clear separation of points into groups having the same color – not only are colors intermixed all over the projection, bit it is often hard to even visually 'split' the projection into distinct point groups.

All above results show, first of all, that the κ ranking of projections is indeed in line with our perception of visual separation. This empirically validates our decision to measure the latter by computing the former. Also, our results show that a good VS leads to a low mixing of the propagated labels, and conversely. In turn, a low mixing leads to a

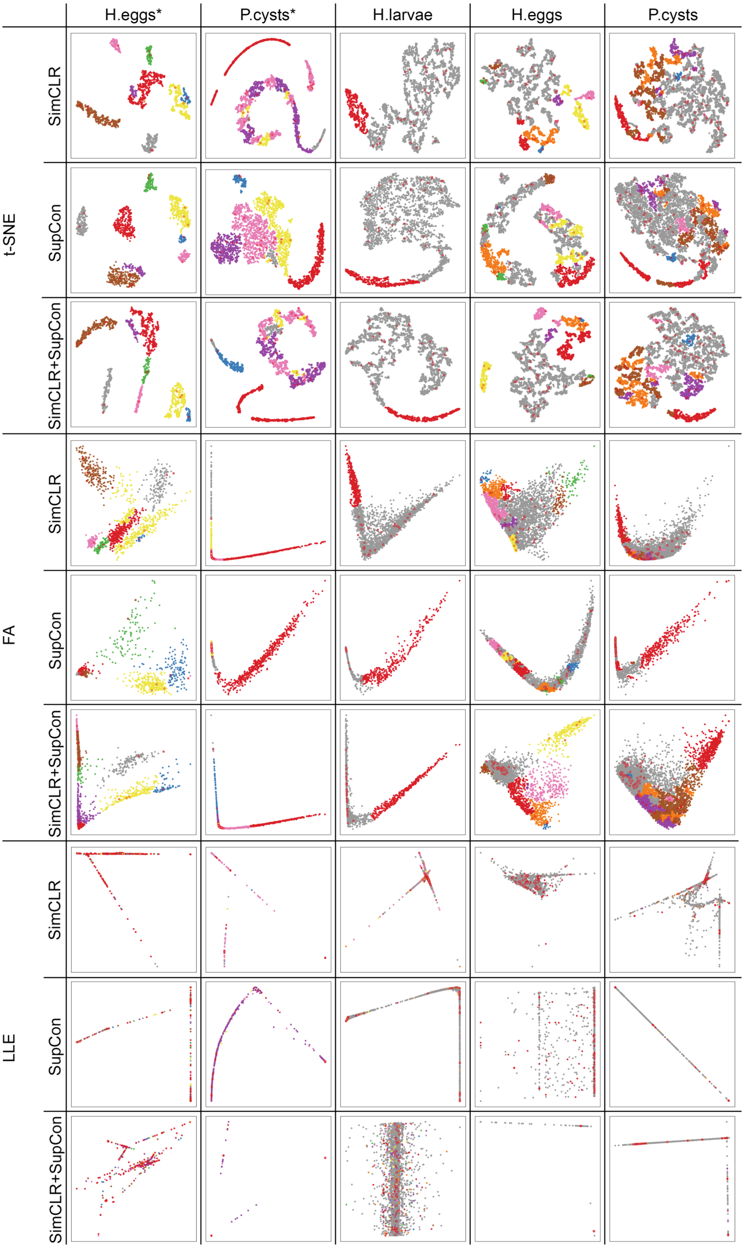


Fig. 6. 2D projections of the best (t-SNE), medium (FA), and worst (LLE) P for the three contrastive latent spaces (SimCLR, SupCon, SimCLR+SupCon) and the six studied datasets (columns). Colored points show the computed pseudo-labels for distinct classes. Outlined points in red represent the supervised points. (Color figure online)

high classification performance (CP), and conversely, *i.e.*, our claim C3. Table 7 shows this by ranking the average results of each projection by each metric for the baseline and VGG-16 trained with the generated pseudo-labels. We see the best κ value for t-SNE with a clear VS and little label mixing in the projections. Conversely, we see the medium-low κ values $[0, 0.6]$ for FA and LLE with poor VS and color-mixing in their projections.

7.3 Data Separation *vs* Visual Separation *vs* Classifier Performance

Section 7.1 discussed how a good projection P can map high DS to high VS. Section 7.2 showed that VS can be measured by κ. Section 6.3 showed that CP is correlated with VS. Let us now put together all these observations.

Figure 7 shows plots of the correlation between (i) DS and VS and (ii) VS and CP, using both accuracy or κ. To simplify the plot, we averaged over initialization strategies (pre-trained, scratch) per contrastive learning method. As such, every point in a plot is a (dataset, contrastive learning method, projection technique) combination. Points are colored to indicate projection techniques. For all same-projection points (15 of them), we also plot a trend line showing their correlation.

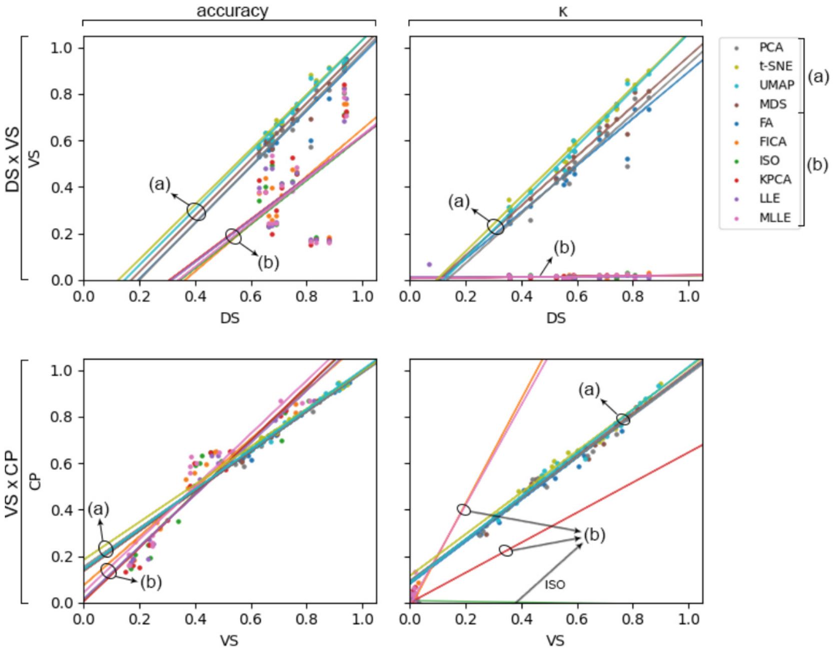


Fig. 7. Correlation plots *DS-VS* and *VS-CP* for accuracy and κ. In each plot, a point corresponds to the average of initialization strategies (scratch or pre-trained) for each dataset and contrastive learning method projected by a distinct P (in different colors). Each line represents the trend of same-P (same-color) points. Projections P are grouped into those which (a) map well DS to VS and (b) create poor VS regardless of DS.

Figure 7 gives several insights. The DS-VS correlation plots confirm our findings in Sect. 7.1. For accuracy, all trend lines have an increasing slope which tells a positive DS-VS correlation, i.e., good DS leads to good VS. We notice two main groups of trend lines: one for (a) UMAP, t-SNE, MDS, PCA, and FA, and another for (b) all other projection techniques. Indeed, Fig. 5(left) showed a pattern of darker/blue points for (b) (see e.g. the *H.eggs**'s row). For κ, we see an increasing slope for trend lines only for (a), and horizontal lines for (b). This says that only projections in group (a) have a positive DS-VS correlation. Figure 5(left) confirms this by showing a stronger pattern of darker/blue cells, i.e., low κ, for *all datasets* in (b). We conclude that projection techniques in group (a) yield a strong DS-VS correlation, whereas projections in group (b) do not do that but rather generate poor-VS results for any DS value.

The VS-CP correlation plots strengthens our earlier findings from Fig. 5 in Sect. 6.3. For accuracy, all trend lines show an increasing slope and thus a positive VS-CP, with a slight difference between projections in group (a) and group (b). For κ, we also see the positive VS-CP correlation except for ISO which has an almost flat line. Additionally, we see that the trend lines in the middle of the plot correspond for projections in group (a). For these projections, both VS and CP spread widely over the $[0, 1]$ range. In contrast, lines for projections in group (b) have values lower than 0.05 for VS and CP. This is the same pattern found in our earlier analysis for DS *vs.* VS, i.e., the presence of two groups of projection techniques with distinct correlation values. The existence of these groups highlights the importance of the chosen projection technique in supporting the DS, VS, and CP links. In detail, for UMAP, t-SNE, MDS, PCA, and FA (group (a)), we confirm a strong DS-VS and VS-CP, thus DS-VS-CP, correlation. In contrast, for FICA, ISO, KPCA, LLE, and MLLE (group (b)), we find a VS-CP *relation* in the sense that both these values are low. However, we did not find any DS-VS *correlation* since these projections map any DS values in $[0, 1]$ to very poor VS values (close to zero).

Table 7. Average values of propagation results (C2) for accuracy and κ and studied P. P is ordered in increasing order for the respective metrics.

metrics	techniques ordered by metric									
acc	ISO	M-LLE	LLE	FICA	K-PCA	FA	PCA	MDS	UMAP	t-SNE
	0.400488	0.406233	0.411292	0.413671	0.414182	0.690367	0.697184	0.721874	0.752606	0.761406
κ	LLE	FICA	M-LLE	K-PCA	ISO	FA	PCA	MDS	UMAP	t-SNE
	0.013256	0.014352	0.014943	0.014971	0.015558	0.497711	0.505837	0.538983	0.583886	0.600066

The above insights connect to our qualitative and ranked analysis of the scatterplots in Fig. 6 and Table 7. The presence of two groups in the correlation plots is also connected to the quality of the perceived projection scatterplots. Scatterplots with a clear VS among groups of different classes/colors separated by whitespace come from t-SNE, a projection in the group (a). Conversely, scatterplots with no VS among groups of different classes/colors and without any whitespace separation come from LLE, a projection in group (b). Our key finding – that projections split in two groups – (a) and (b), with good, respectively poor, DS-VS correlation – complements and extends the largest former quantitative comparison of projection techniques [12]. Our set (a), i.e., projections which best preserve DS-VS and are best for building high-CP classifiers, matches quite closely the projections found best in [12]. However, our quality criteria – VS measured by pseudolabeling performance and CP measured by classifier performance are *completely* different than the criteria used in [12] to measure projection quality. The latter criteria include metrics such as trustworthiness, continuity, normalized stress, neighborhood hit, and Shepard correlation. As outlined in Sect. 2, such metrics only measure how a projection preserves *local* data structure. This is typically done by using small-sized neighborhoods of under 10 points in both the data and projection space. Visual separation (VS), however, occurs at much larger scales in a scatterplot. Moreover, we measure VS in a completely different way, namely, as the performace of a ML algorithm (pseudolabeling) that handles all the scatterplot points globally rather than in terms of small-scale, independent, neighborhoods. As such, the fact that we 'rank' projections quite similarly to [12], but using completely different metrics, has several *potentially* far-reaching implications (all to be tested by future work):

- Projections assessed by local quality metrics [12] can be used to predict classifier performance;
- Higher-level properties like visual separation (VS) can be predicted by lower-level metrics [12];
- the VS of a projection, measured by κ or accuracy during pseudolabeling, can be an additional quality metric for generic projection assessment. Apart from their ability to gauge a projection more globally, such metrics are also much faster to compute than neighborhood-based metrics such as trustworthiness, continuity, normalized stress, or Shepard correlation.

7.4 Contrastive Learning from Few Supervised Samples

Our experiments with ImageNet pre-trained weights and higher results show that SimCLR – even trained with *thousands* of unsupervised samples (69%) – and having more information on the data distribution of the original space – did not surpass SupCon which used only *dozens* of supervised samples (1%). Our explanation for this is that the latent space generated when SupCon was used to fine-tune SimCLR (SupCon+SimCLR) had a *better DS* than the one created by SimCLR – this is by Table 4. This shows the benefit of using SupCon with supervised data restriction as compared to SimCLR, a finding that up to our knowledge is novel. Separately, the fact that a higher DS lead to a higher CP further supports our claim C3.

8 Conclusion

We presented a detailed study of the link between data separation (DS) in a high-dimensional partially-labeled dataset, the visual separation (VS) in a 2D projection of that dataset, and the performance of a classifier (CP) constructed by pseudolabeling the abovementioned projection.

In our work, we used two contrastive learning approaches (SimCLR and SupCon) as well as their combination. We projected the latent spaces produced by these methods to

2D using ten projection techniques (FA, FICA, ISO, KPCA, LLE, MDS, MLLE, PCA, t-SNE, and UMAP). We propagated labels in these projections and finally used these pseudo-labels to train a deep-learning classifier for a challenging problem involving the classification of human intestinal parasite images.

Our results show that SimCLR+SupCon performed better than using only SimCLR or SupCon to create a data space with strong DS. In turn, this allowed us to construct an end-to-end classifier with higher accuracy and κ values than earlier reported for the respective datasets in the literature.

Separately, we showed that the 10 studied projection techniques can be split into two groups. Projections in the first group (FICA, ISO, KPCA, LLE, and MLLE) yield very poor VS results for any DS values of their input data, and consequently also very poor CP results. These projections are hence not useful for our classifier engineering pipeline and, arguably, they will also have challenges for other infovis applications where VS is important. Projections in the second group (FA, MDS, PCA, t-SNE, and UMAP) show a good DS-VS correlation and, next, a good VS-CP correlation. These projections are thus ideal for our classifier engineering task and, arguably, for other infovis applications where VS is important. Our work shows, to our knowledge, for the first time how *specific* projection techniques preserve a strong DS-VS-CP correlation (or not). Our findings can assist additional applications in infovis or machine learning where projections are used.

Several future work directions are possible. First, the connection between pseudolabeling quality and visual separation could be further exploited to e.g. design new metrics for visual separation using labeling algorithms or, conversely, to guide labeling by existing visual quality metrics. Secondly, user experiments could be designed and executed to assess more formally the relationship between pseudolabeling quality and perceived visual separation. Finally, we aim to involve users in the loop to assist the automatic pseudolabeling process by *e.g.* adjusting some of the automatically propagated labels based on the human assessment of VS. We believe that this will lead to even more accurate pseudo-labels and, ultimately, more accurate classifiers for the problem at hand.

Acknowledgments. The authors acknowledge FAPESP grants #2014/12236-1, #2019/10705-8, #2022/12668-5, CAPES grants with Finance Code 001, and CNPq grants #303808/2018-7.

References

1. Amorim, W., Falcão, A., Papa, J., Carvalho, M.: Improving semi-supervised learning through optimum connectivity. Pattern Recognit. **60**, 72–85 (2016)
2. Amorim, W., et al.: Semi-supervised learning with connectivity-driven convolutional neural networks: Pattern Recognit. Lett. **128**, 16–22 (2019)
3. Arazo, E., Ortego, D., Albert, P., O'Connor, N.E., McGuinness, K.: Pseudo-labeling and confirmation bias in deep semi-supervised learning. In: IJCNN, pp. 1–8. IEEE (2020)
4. Benato, B.C., Falcão, A.X., Telea, A.C.: Linking data separation, visual separation, and classifier performance using multidimensional projections (2023). https://github.com/barbarabenato/linking_ds_vs_cp_proj

5. Benato, B.C., Telea, A.C., Falcão, A.X.: Semi-supervised learning with interactive label propagation guided by feature space projections. In: Proceedings of SIBGRAPI, pp. 392–399 (2018)
6. Benato, B.C., Gomes, J.F., Telea, A.C., Falcão, A.X.: Semi-supervised deep learning based on label propagation in a 2d embedded space. In: Proceedings of CIARP, pp. 371–381 (2021)
7. Benato, B.C., Telea, A.C., Falcao, A.X.: Iterative pseudo-labeling with deep feature annotation and confidence-based sampling. In: Proceedings of SIBGRAPI, pp. 192–198. IEEE (2021)
8. Benato, B.C., Falcão, A.X., Telea, A.C.: Linking data separation, visual separation, and classifier performance using pseudo-labeling by contrastive learning. In: Proceedings of VISAPP, pp. 315–324 (2023)
9. Benato, B.C., Gomes, J.F., Telea, A.C., Falcão, A.X.: Semi-automatic data annotation guided by feature space projection. Pattern Recognit. **109**, 107612 (2021)
10. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607. PMLR (2020)
11. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE TPAMI **24**(5), 603–619 (2002)
12. Espadoto, M., Martins, R., Kerren, A., Hirata, N., Telea, A.: Toward a quantitative survey of dimension reduction techniques. IEEE TVC **27**(3), 2153–2173 (2019)
13. Grill, J.B., et al.: Bootstrap your own latent: a new approach to self-supervised learning. arXiv preprint arXiv:2006.07733 (2020)
14. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of IEEE CVPR, pp. 9729–9738 (2020)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of IEEE CVPR, pp. 770–778 (2016)
16. Hossin, M., Sulaiman, M.N.: A review on evaluation metrics for data classification evaluations. IJDKP **5**(2), 1 (2015)
17. Hyvarinen, A.: Fast ICA for noisy data using gaussian moments. In: Proceedings of IEEE ISCAS, vol. 5, pp. 57–61 (1999)
18. Jing, L., Tian, Y.: Self-supervised visual feature learning with deep neural networks: a survey. In: IEEE TPAMI, p. 1 (2020). https://doi.org/10.1109/TPAMI.2020.2992393
19. Joia, P., Coimbra, D., Cuminato, J.A., Paulovich, F.V., Nonato, L.G.: Local affine multidimensional projection. In: Proceedings of IEEE TVCG, vol. 17, pp. 2563–2571 (2011)
20. Jolliffe, I.T.: Principal component analysis and factor analysis, pp. 115–128 (1986)
21. Khosla, P., et al.: Supervised contrastive learning. Proc. NeurIPS **33**, 18661–18673 (2020)
22. Kim, Y., Espadoto, M., Trager, S., Roerdink, J., Telea, A.: SDR-NNP: sharpened dimensionality reduction with neural networks. In: Proceedings of IVAPP (2022)
23. Kim, Y., Telea, A.C., Trager, S.C., Roerdink, J.B.: Visual cluster separation using high-dimensional sharpened dimensionality reduction. Inf. Vis. **21**(3), 197–219 (2022)
24. Lee, D.H.: Pseudo-label : the simple and efficient semi-supervised learning method for deep neural networks. In: Proceedings of ICML-WREPL (2013)
25. van der Maaten, L.: Accelerating t-SNE using tree-based algorithms. JMLR **15**(1), 3221–3245 (2014)
26. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. JMLR **9**, 2579–2605 (2008)
27. Maaten, L.V.D., Postma, E., den Herik, J.V.: Dimensionality reduction: a comparative review. J. Mach. Learn. Res. **10**, 66–71 (2009)
28. McInnes, L., Healy, J., Melville, J.: Umap: uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 (2018)
29. Miyato, T., Maeda, S.i., Koyama, M., Ishii, S.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE TPAMI **41**(8), 1979–1993 (2018)

30. Nonato, L., Aupetit, M.: Multidimensional projection for visual analytics: linking techniques with distortions, tasks, and layout enrichment. In: IEEE TVCG (2018)
31. Osaku, D., Cuba, C.F., Suzuki, C.T., Gomes, J.F., Falcão, A.X.: Automated diagnosis of intestinal parasites: a new hybrid approach and its benefits. Comput. Biol. Med. **123**, 103917 (2020)
32. Papa, J.P., Falcão, A.X.: A learning algorithm for the optimum-path forest classifier. In: Torsello, A., Escolano, F., Brun, L. (eds.) GbRPR 2009. LNCS, vol. 5534, pp. 195–204. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02124-4_20
33. Paulovich, F.V., Nonato, L.G., Minghim, R., Levkowitz, H.: Least square projection: a fast high-precision multidimensional projection technique and its application to document mapping. IEEE TVCG **14**(3), 564–575 (2008)
34. Pham, H., Dai, Z., Xie, Q., Le, Q.V.: Meta pseudo labels. In: Proceedings of IEEE CVPR, pp. 11557–11568, June 2021
35. Rauber, P.E., Fadel, S.G., Falcão, A.X., Telea, A.: Visualizing the hidden activity of artificial neural networks. IEEE TVCG **23**(1), 101–110 (2017)
36. Rauber, P., Falcão, A., Telea, A.: Projections as visual aids for classification system design. Inf. Vis. **17**(4), 282–305 (2017)
37. Rodrigues, F.C.M., Espadoto, M., Jr., R.H., Telea, A.: Constructing and visualizing high-quality classifier decision boundary maps. Information **10**(9), 280–297 (2019)
38. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**(5500), 2323–2326 (2000)
39. Schölkopf, B., Smola, A., Müller, K.R.: Kernel principal component analysis. In: Proceedings of ICANN, pp. 583–588 (1997)
40. Suzuki, C., Gomes, J., Falcão, A., Shimizu, S., Papa, J.: Automated diagnosis of human intestinal parasites using optical microscopy images. In: Proceedings of Symposium Biomedical Imaging, pp. 460–463, April 2013
41. Tenenbaum, J.B., Silva, V.D., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**(5500), 2319–2323 (2000)
42. Torgerson, W.S.: Theory and Methods of Scaling. Wiley, Hoboken (1958)
43. Venna, J., Kaski, S.: Visualizing gene interaction graphs with local multidimensional scaling. In: Proceedings of ESANN, vol. 6, pp. 557–562 (2006)
44. Zhang, Z., Wang, J.: Mlle: modified locally linear embedding using multiple weights. In: Proceedings of NIPS, vol. 19, pp. 1593–1600 (2006)