01 Jan 2022

Incremental Cluster Validity Index-Guided Online Learning for Performance and Robustness to Presentation Order

Leonardo Enzo Brito da Silva

Nagasharath Rayapati

Donald C. Wunsch
Missouri University of Science and Technology, dwunsch@mst.edu

Recommended Citation

L. E. da Silva et al., "Incremental Cluster Validity Index-Guided Online Learning for Performance and Robustness to Presentation Order," *IEEE Transactions on Neural Networks and Learning Systems*, Institute of Electrical and Electronics Engineers, Jan 2022.
The definitive version is available at https://doi.org/10.1109/TNNLS.2022.3212345

Incremental Cluster Validity Index-Guided Online Learning for Performance and Robustness to Presentation Order

Leonardo Enzo Brito da Silva, *Member, IEEE*, Nagasharath Rayapati, and Donald C. Wunsch, II, *Fellow, IEEE*

***Abstract*— In streaming data applications, the incoming samples are processed and discarded, and therefore, intelligent decision-making is crucial for the performance of lifelong learning systems. In addition, the order in which the samples arrive may heavily affect the performance of incremental learners. The recently introduced incremental cluster validity indices (iCVIs) provide valuable aid in addressing such class of problems. Their primary use case has been cluster quality monitoring; nonetheless, they have been recently integrated in a streaming clustering method. In this context, the work presented, here, introduces the first adaptive resonance theory (ART)-based model that uses iCVIs for unsupervised and semi-supervised online learning. Moreover, it shows how to use iCVIs to regulate ART vigilance via an iCVI-based match tracking mechanism. The model achieves improved accuracy and robustness to ordering effects by integrating an online iCVI module as module B of a topological ART predictive mapping (TopoARTMAP)—thereby being named iCVI-TopoARTMAP—and using iCVI-driven post-processing heuristics at the end of each learning step. The online iCVI module provides assignments of input samples to clusters at each iteration in accordance to any of the several iCVIs. The iCVI-TopoARTMAP maintains useful properties shared by the ART predictive mapping (ARTMAP) models, such as stability, immunity to catastrophic forgetting, and the many-to-one mapping capability via the map field module. The performance and robustness to the presentation order of iCVI-TopoARTMAP were evaluated via experiments with synthetic and real-world datasets.**

***Index Terms*— Adaptive resonance theory predictive mapping (ARTMAP), clustering, data streams, incremental cluster validity index (iCVI), online learning, semi-supervised learning.**

I. Introduction

Cluster validation [1] is an essential part of cluster analysis and consists of measuring the quality of data partitions yielded by the clustering algorithms. In the past, this task was performed using batch cluster validity indices (bCVIs) [2], [3], [4], [5] to evaluate the clustering algorithms after completion. In 2018, Moshtaghi et al. [6] introduced the idea of incremental cluster validity indices (iCVIs) to visually monitor and evaluate cluster footprints yielded by the online clustering algorithms in data stream applications, thus being referred to as incremental stream monitoring functions (iSMFs) [7], [8]. A recursive formulation was derived for a quantity present in sum-of-squares (SS)-based bCVIs, namely, the fuzzy compactness, consequently allowing for the incremental computation of a related class of bCVIs. Since then, efforts have been used to introduce incremental versions of popular SS- and non-SS-based bCVIs [9], [10], [11].

bCVIs have also been traditionally used as fitness functions in the optimization algorithms to perform clustering [12] or to aid in the clustering process—for instance, Brito da Silva and Wunsch [13] presented a fuzzy adaptive resonance theory (ART) [14] augmented with bCVIs as an additional vigilance test, while Smith and Wunsch [15] used them for vigilance parameter setting within the neural network. Chenaghlou [16] introduced the usage of iCVIs in the online clustering task via systems that combined the incremental clustering algorithms and an iCVI-based controller to determine the creation and merging of cluster prototypes. In addition, Ibrahim et al. [17] used the compactness to aid in making decisions regarding the emergence of clusters when performing stream clustering. Although iCVIs were not widely used prior to 2018, we did find earlier contributions: Lughofer [18] presented an ART-like online incremental clustering algorithm that used a non-SS-based iCVI [19] to guide a splitting and merging heuristic.

ART predictive mapping (ARTMAP) neural networks [20], [21], [22], [23] are built upon elementary ART models such as [14], [24], and their classical applications are for supervised learning [25]. However, some variants exist that have been repurposed for unsupervised learning applications, such as biclustering (BARTMAP [26], hierarchical BARTMAP [27], and TopoBARTMAP [28]), hierarchical divisive clustering (SMART [29]), and offline iCVI-based clustering (iCVI-ARTMAP [30]). Furthermore, the unified ART model and its extended version [31] can perform mixed-modality learning. In particular, hierarchical BARTMAP introduced a hierarchy of biclusters whose levels are evaluated using a bCVI; TopoBARTMAP introduced the usage of fuzzy TopoART [32] as fuzzy ARTMAP's building blocks; iCVI-ARTMAP introduced the usage of an iCVI-based module to replace fuzzy ARTMAP's labeling module; unified and extended unified ARTs introduced integrated learning dynamics which depends on the type of inputs provided to the model, wherein

Manuscript received 13 December 2021; revised 7 July, 2022; accepted 3 October 2022. *(Corresponding author: Leonardo Enzo Brito da Silva.)*

Leonardo Enzo Brito da Silva and Nagasharath Rayapati are with Guise AI, Inc., Rolla, MO 65401 USA (e-mail: leonardo@guise.ai; naga@guise.ai).

Donald C. Wunsch, II, is with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65401 USA (e-mail: wunsch@ieee.org).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TNNLS.2022.3212345.

Digital Object Identifier 10.1109/TNNLS.2022.3212345

supervised, unsupervised, or reinforcement learning modes have different priorities.

Our previous work [30] demonstrated how to integrate fuzzy ARTMAP [21], [22] and iCVIs for offline incremental unsupervised learning and argued that for the iterative algorithms, there is usually no reason to use a bCVI for which an iCVI has been developed. The iCVI-TopoARTMAP model presented in this work goes further to allow for online learning by incorporating: 1) online normalization, online complement coding, and online weight vector rescaling based on [33], [34]; 2) iCVI-based merge, iCVI-based swap, split, compression, and prune-and-reassign heuristics during online learning; 3) summary statistics associated with each category; and 4) an iCVI-based match tracking procedure, to aid in dynamically restructuring clusters using feedback from iCVIs. The iCVI-TopoARTMAP model considers not only the local similarities of samples to categories but also the global effect of assigning the presented sample to a given cluster (i.e., the outcome with respect to the entire cluster structure), which is considered by the iCVI-based label generation module.

The main contribution of this work is the iCVI-TopoARTMAP model. The latter is, to the best of our knowledge, the first combination of a model from the ARTMAP family with iCVIs for the purposes of online unsupervised and semi-supervised learning. No prior ART method has been able to combine iCVIs and multiprototype representation-capable supervised methods such as ARTMAP to perform online learning. Moreover, to the best of our knowledge, with the exception of [16], [18], no other streaming clustering algorithm makes use of iCVIs. However, these approaches do not make use of multiprototype-based representation along with the iCVI-driven aforementioned operations. The remainder of this article is organized as follows. Section II describes the iCVI-TopoARTMAP model; Section III discusses the experiments and our findings; finally, Section IV concludes this article.

II. iCVI-TopoARTMAP Design

Clustering is an unsupervised learning task; thus, no labels are provided to classify data into groups. The iCVI-TopoARTMAP model depicted in Fig. 1 circumvents this problem in online learning using iCVIs for the purpose of label generation. Specifically, it allows a user to select an iCVI and perform online incremental multiprototype-based unsupervised and semi-supervised learning. The model presented, here, uses an iCVI for online decision-making regarding: (a) sample allocation to clusters, (b) prototype allocation to clusters, (c) merging clusters, and d) iCVI-based match tracking, while incrementally building associative mappings between prototypes (categories) to clusters. Items (b)–(d) above can be individually enabled, while item (a) is automatically disregarded by the system if a supervised label is presented alongside the current sample.

The iCVI-TopoARTMAP is a winner-take-all ARTMAP-based neural network; therefore, it consists of three main parts: 1) module A; 2) module B; and 3) map field. The remainder of this section provides the design details of these modules. For clarity, the main hyperparameters of iCVI-TopoARTMAP are listed in Table I.

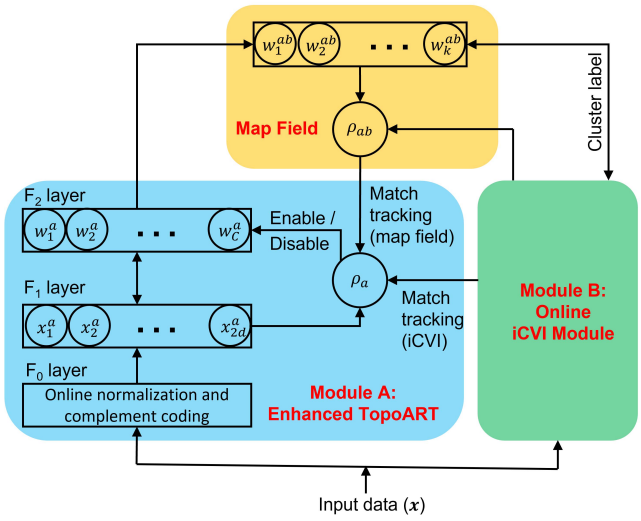


Fig. 1. iCVI-TopoARTMAP neural network. The enhanced TopoART is a variant of fuzzy TopoART [32] equipped with online min–max normalization and complement coding [21] for input samples $\boldsymbol{x}$, rescaling of weight vectors $\boldsymbol{w}^a$ [33], [34], summary statistics and inactivity counters per category, and a CONN matrix [35]. The online iCVI module is the labeling component that provides assignments of input samples to clusters at each input presentation in accordance to any of several iCVIs. If the iCVI-based match tracking is enabled, following successive worsening of the selected iCVI, the vigilance parameter ρ_a is changed until the iCVI value restarts to improve. The map field is a variant of fuzzy ARTMAP's map field [21], [22] that can process (in unsupervised learning mode) a label matrix containing multiple cluster assignments in one-hot encoded form and provide multiprototype representation of clusters via its mapping weight vectors $\boldsymbol{w}^{ab}$. At the end of each learning step, if enabled, the postprocessing strategies can be performed.

TABLE I
iCVI-TopoARTMAP Hyperparameters

Hyper-parameter	Description
ρ_a	Module A's vigilance parameter.
β_1	Module A's learning rate for the 1st resonant category.
β_2	Module A's learning rate for the 2nd resonant category.
α	Module A's choice parameter.
M_{type}	Module A's match function type (Eq. (7) or (8)).
ε	Map field's standard match tracking parameter.
ρ_{ab}	Map field's vigilance parameter.
β_{ab}	Map field's learning rate.
L_{type}	Map field learning type (fixed or variable).
iCVI	Selected iCVI and corresponding parameters.
ε_{icvi}	iCVI-based match tracking parameter.
$\rho_{MT_{icvi}}$	iCVI-based match tracking vigilance parameter.
ρ_c	Vigilance parameter for the compression strategy.
τ	Number of checks threshold for iCVI-based match tracking.
ϕ	Number of samples (within a cluster) threshold for pruning.
ξ	Category inactivity threshold for compression and pruning.

The post-processing strategies (Section II-D) can be individually enabled using their corresponding flags: EN_swap, EN_merge, EN_split, $EN\_compress$ and $EN\_prune\_reassign$. The iCVI-based match tracking and uncommitted category can also be individually enabled using their flags: EN_MT_{icvi} and EN_T^u, respectively.

A. Module A: Enhanced Topological Fuzzy ART

Module A of iCVI-TopoARTMAP is a variant of topological fuzzy ART (topoFA) [32], namely, the enhanced topoFA, which features the following.

1) Online Normalization of Inputs, Online Complement Coding, and Online Rescaling of Weight Vectors: In general, knowledge of the data range is absent in online learning use cases. This is a challenge for the fuzzy ART-based neural networks [25] because they need normalized (i.e., in the range $[0, 1]$) and complement coded inputs [21]. Moreover, when input samples push the lower and upper bounds of the data range interval, its weight vectors must be rescaled. Recent works have used different approaches to address this problem [33], [34], [36]. Here, the recursive rescaling formulation presented in [33] is used to adapt the weight vectors of module A and normalize the input samples. However, unlike [33], the minimum and maximum statistics stored are immediately overwritten if the incoming sample pushes the data interval like in [34].

If the maximum and/or minimum values of the data features are different from the current stored statistics (i.e., $\boldsymbol{x}_{\max} \in \mathbb{R}^d$ and/or $\boldsymbol{x}_{\min} \in \mathbb{R}^d$, respectively), module A's weight vectors (i.e., $\boldsymbol{w}_j^a = [\boldsymbol{u}_j, \bar{\boldsymbol{v}}_j] \in \mathbb{R}^{2d}$) undergo online rescaling. First, the components $\boldsymbol{u}_j = [\boldsymbol{w}_{(j,1)}^a, \ldots, \boldsymbol{w}_{(j,d)}^a]$ and $\boldsymbol{v}_j = \vec{\mathbf{1}} - \bar{\boldsymbol{v}}_j = \vec{\mathbf{1}} - [w_{(j,d+1)}^a, \ldots, w_{(j,2d)}^a]$ of each category j of module A undergo an inverse transformation given by $\mathcal{T}^{-1}(\cdot)$. Transformation $\mathcal{T}(\boldsymbol{z})$ represents a componentwise min–max normalization of a vector $\boldsymbol{z}$ using $\boldsymbol{x}_{\max}$ and $\boldsymbol{x}_{\min}$

$$\mathcal{T}(z_i) = \frac{z_i - x_{\min,i}}{x_{\max,i} - x_{\min,i}}. \tag{1}$$

Then, $\boldsymbol{x}_{\max}$ and $\boldsymbol{x}_{\min}$ are updated to reflect the new feature-wise maxima and minima, respectively. Finally, components $\boldsymbol{u}_j$ and $\boldsymbol{v}_j$ are rescaled using $\boldsymbol{u}_j^{\text{new}} \leftarrow \mathcal{T}(\boldsymbol{u}_j) \vee \vec{\mathbf{0}}$ and $\boldsymbol{v}_j^{\text{new}} \leftarrow \mathcal{T}(\boldsymbol{v}_j) \wedge \vec{\mathbf{1}}$, where the operators $\vee$ and $\wedge$ represent componentwise maximum and minimum between two vectors, respectively. Finally, the weight vector of each category j is updated by concatenating the rescaled components: $\boldsymbol{w}_j^a \leftarrow [\boldsymbol{u}_j^{\text{new}}, \vec{\mathbf{1}} - \boldsymbol{v}_j^{\text{new}}]$.

2) Summary Statistics, Inactivity Counters, and Connectivity Matrix: The data stream algorithms usually store summary statistics (or "footprints") in the form of multiple "micro-clusters." These are incrementally updated once an assignment has been made for the current input sample and can be used to alter the perceived cluster structure underlying the streaming data (e.g., via heuristics for splitting, merging) [37].

In module A, each category j has a weight vector $\boldsymbol{w}_j^a$, an inactivity counter ($\bar{a}_j$) and the summary statistics of frequency ($n_j \in \mathbb{N}$), mean ($\boldsymbol{\mu}_j \in \mathbb{R}^d$), and hard compactness ($\text{CP}_j \in \mathbb{R}$)

$$\boldsymbol{\mu}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \boldsymbol{x}_i, \quad \boldsymbol{x}_i \text{ mapped to category } j \tag{2}$$

$$\text{CP}_j = \sum_{i=1}^{n_j} \|\boldsymbol{x}_i - \boldsymbol{\mu}_j\|_2^2, \quad \boldsymbol{x}_i \text{ mapped to category } j. \tag{3}$$

Furthermore, module A also stores a $P \times P$ connectivity matrix ($\text{CONN}_{P\times P}$) [35] containing local density information, where P represents the total number of categories. All these quantities (n, $\boldsymbol{\mu}$, CP, and CONN) are continuously and incrementally updated during learning. Note that TopoFA [32] only stores the accumulated frequency (i.e., sample count) for each category and a binary adjacency matrix in addition to the standard weight vectors, while the probabilistic fuzzy ARTMAP [38] only stores additional centroids.

3) Learning (Algorithm 1): When a d-dimensional sample $\boldsymbol{x}$ is presented, first the online rescaling procedure previously discussed in item 1 takes place. Next, a copy of $\boldsymbol{x}$ undergoes min–max normalization using the updated $\boldsymbol{x}_{\max}$ and $\boldsymbol{x}_{\min}$ statistics and subsequent complement coding [21]

$$\boldsymbol{x}^a \leftarrow \left[\mathcal{T}(\boldsymbol{x}), \vec{\mathbf{1}} - \mathcal{T}(\boldsymbol{x})\right]. \tag{4}$$

Module A retains the main dynamics of fuzzy topoART [32], which are akin to fuzzy ART's [14]. After presenting the input $\boldsymbol{x}^a$, the activation function T_j^a of each category is computed as

$$T_j^a = \frac{\left\|\boldsymbol{x}^a \wedge \boldsymbol{w}_j^a\right\|_1}{\alpha + \left\|\boldsymbol{w}_j^a\right\|_1}, \quad \alpha > 0 \tag{5}$$

where T_j^a and $\boldsymbol{w}_j^a$ are the activation and weight vector of category j, respectively, and $\|\cdot\|_1$ is the ℓ_1-norm. To be considered resonant, an existing category J_1 must simultaneously satisfy the following criteria:

$$T_{J_1}^a > T^u \tag{6}$$

$$M_{J_1}^a = \frac{\left\|\boldsymbol{x}^a \wedge \boldsymbol{w}_{J_1}^a\right\|_1}{\|\boldsymbol{x}^a\|_1} \geq \rho_a \tag{7}$$

where $T^u = d/(\alpha + d)$ (if EN_T^u is enabled—see Table I), $M_{J_1}^a$ is the match value of category J_1 with weight vector $\boldsymbol{w}_{J_1}^a$, and $0 \leq \rho_a \leq 1$ is the vigilance parameter of module A. According to the problem at hand, a cosine-distance-based match function $M_{J_1}^a$ can be used instead of (7)

$$M_{J_1}^a = 1 - \frac{\langle \boldsymbol{x}^b, \boldsymbol{\mu}_{J_1} \rangle}{\|\boldsymbol{x}^b\|_2 \left\|\boldsymbol{\mu}_{J_1}\right\|_2} \leq \rho_a, \quad 0 \leq \rho_a \leq 2 \tag{8}$$

where $\boldsymbol{x}^b = \boldsymbol{x}$ is the original input to the online iCVI module, $\boldsymbol{\mu}_{J_1}$ is the local mean statistic associated with category J_1, $\|\cdot\|_2$ is the ℓ_2-norm, and $\langle \cdot, \cdot \rangle$ is the inner product. The cosine similarity[1] has been extensively used to compare embeddings generated by deep feature extractors (e.g., person reidentification [40], face recognition [41], [42]).

When a resonant category J_1 is found for sample $\boldsymbol{x}$, then its associated $\boldsymbol{w}_{J_1}^a$, n_{J_1}, $\boldsymbol{\mu}_{J_1}$, and CP_{J_1} are updated incrementally using the following formulas [6], [21], [30]:

$$\boldsymbol{w}_{J_1}^a \leftarrow (1 - \beta_1)\boldsymbol{w}_{J_1}^a + \beta_1 \left(\boldsymbol{x}^a \wedge \boldsymbol{w}_{J_1}^a\right), \quad 0 < \beta_1 \leq 1 \tag{9}$$

$$\boldsymbol{\mu}_{J_1} \leftarrow \frac{n_{J_1}}{n_{J_1} + 1}\boldsymbol{\mu}_{J_1} + \frac{1}{n_{J_1} + 1}\boldsymbol{x}^b \tag{10}$$

$$\text{CP}_{J_1} \leftarrow \text{CP}_{J_1} + \frac{n_{J_1}}{n_{J_1} + 1}\left\|\boldsymbol{x}^b - \boldsymbol{\mu}_{J_1}\right\|_2^2 \tag{11}$$

$$n_{J_1} \leftarrow n_{J_1} + 1. \tag{12}$$

If category J_1 does not satisfy (6) and (7) [or (6) and (8)], then it is inhibited and the search continues with the next highest ranked category. This process is repeated until either

[1]The cosine similarity is a dilation invariant measure defined as a normalized inner product. It carries angle (but not magnitude) information between two real-valued vectors [1], [39].

1) a category meets all the resonance criteria or 2) the set of categories with activation greater than T^u is exhausted, in which case a new category is created and initialized as

$$\boldsymbol{w}_{J_1}^a \leftarrow \boldsymbol{x}^a, \quad n_{J_1} \leftarrow 1, \quad \boldsymbol{\mu}_{J_1} \leftarrow \boldsymbol{x}^b, \quad \mathrm{CP}_{J_1} \leftarrow 0. \tag{13}$$

The inactivity counters of all the categories are then incremented by one, i.e., $\bar{a}_j \leftarrow \bar{a}_j + 1, \forall j$. However, the resonant category J_1 has its inactivity counter reset to zero (or initialized in case a new category is created), i.e., $\bar{a}_{J_1} \leftarrow 0$.

If a first resonant category J_1 is found within the set of existing categories, then the search for the second resonant category J_2 starts from the subsequent highest ranked (if any) as per the activation function (5). However, if a new category is created, then the search for a second resonant category is performed on the full set of previously existing categories in descending order of activation (5). Note that the second resonant category only needs to satisfy the constraints expressed by inequalities (6) and (7) or (6) and (8) while using the baseline vigilance parameter (i.e., the original user-defined module A vigilance parameter). Within the iCVI-TopoARTMAP architecture, the first resonant category also needs to satisfy the map field vigilance test (Section II-C). If a second resonant category J_2 is found, then only its weight vector is updated

$$\boldsymbol{w}_{J_2}^a \leftarrow (1-\beta_2)\boldsymbol{w}_{J_2}^a + \beta_2\left(\boldsymbol{x}^a \wedge \boldsymbol{w}_{J_2}^a\right), \quad 0 \le \beta_2 \le \beta_1 \le 1. \tag{14}$$

Finally, when both the first and second resonant categories are found, then their corresponding entries in the CONN matrix [i.e., (J_1, J_2) and (J_2, J_1)] are incremented by 1.

B. Module B: Online iCVI Module

Module B of iCVI-TopoARTMAP extends the offline iCVI-framework component of iCVI-ARTMAP [30] to online learning applications. This work uses the following iCVIs: incremental Xie-Beni (iXB) [6], incremental Davies–Bouldin (iDB) [6], incremental Calinski–Harabasz (iCH) [11], incremental Pakhira–Bandyopadhyay–Maulik (iPBM) [11], incremental WB index (iWB) [11], and incremental Conn_index (iConn_index) [11]. For brevity, we refer the reader to [6], [11] for a comprehensive treatment of iCVI formulations.

The online iCVI module stores and updates the iCVI value and iCVI-specific quantities, which collective define its *state*. All SS-based iCVI-specific quantities include clusterwise and entire data statistics (i.e., frequencies, means, and compactnesses). In particular, the following are examples (not an exhaustive list) of specific quantities: between-group SSs and within-group SSs for iCH and iWB; within-group SSs and prototype dissimilarity matrix for iXB and iPBM; prototype dissimilarity matrix and cluster similarity matrix for iDB; mapping of prototypes to clusters, frequency statistics, within-cluster interconnectivity, and cluster-wise interconnectivity matrix for the iConn_index.[2] Hereafter, the set of specific quantities associated with a given iCVI will be referred to as $\mathcal{Q}$. The online iCVI module is responsible for the following tasks.

[2]In the computation of iConn_index, the frequency statistics are considered independently of the CONN matrix, and thus, while the sum of the frequency statistics is equal to the total number of samples seen, the sum of the upper (or lower) triangular part of CONN may not be because of the vigilance criteria imposed to the first and second resonant categories within module A.

Algorithm 1 Module A Learning

Input : $\boldsymbol{X}$, ρ_a, β_1, β_2, α, ϵ, ϕ, ξ, M_{type}, EN_T^u.
Output : Trained Enhanced TopoART.

```
/* Notation */
W^a = [w^a_{i,j}]_{P×2d}: Module A's weight matrix.
ν^a_j: Boolean output of module A's vigilance test for category j.
ν^{ab}_j: Boolean output of map field's vigilance test for category j.
J_i: ith best matching category (i ∈ {1,2}).
s_j: statistics of category j {n_j, μ_j, CP_j}.
I: list of indices of categories sorted in descending order of activation.
for x ∈ X do
    if W^a = {∅} then
        Initialize: w^a_1, s_1, ā_1, CONN, x_max, x_min.
    else
        Rescale w^a's and update x_max and x_min.
        x^a ← [T(x), 1⃗ − T(x)].
        ā_j ← ā_j + 1, ∀j.
        Set vigilance parameter to ρ_a.
        Compute T^a_j, M^a_j, ∀j, and get I.
        mismatch ← True.
        for j ← 1 to P do
            J_1 ← I[j].
            if T^a_{J_1} > T^u then
                if ν^a_{J_1} then
                    if ν^{ab}_{J_1} then
                        mismatch ← False.
                        Update: w^a_{J_1}, s_{J_1}, ā_{J_1}.
                        Set vigilance parameter to ρ_a.
                        for i ← j + 1 to P do
                            if T^a_{I[i]} > T^u and ν^a_{I[i]} then
                                J_2 ← I[i].
                                Update: w^a_{J_2}, CONN.
                                break.
                        break.
                    else
                        Engage standard match tracking.
        if mismatch = True then
            Set vigilance parameter to ρ_a, J_2 ← {∅}.
            for j ← 1 to P do
                if T^a_{I[j]} > T^u and ν^a_{I[j]} then
                    J_2 ← I[j].
                    break.
            J_1 ← P + 1. Initialize: w^a_{J_1}, s_{J_1}, ā_{J_1}.
            Expand CONN matrix.
            if J_2 ≠ {∅} then
                Update: w_{J_2}, CONN.
```

1) iCVI-Based Label Generation: The online iCVI module generates a cluster label matrix $\boldsymbol{Y}$ that represents the best assignments for the current sample with respect to the current cluster structures according to the selected iCVI. Let T_i^b be the temporary iCVI value (or $-T_i^b$ if the iCVI is min-optimal) corresponding to assigning the current sample $\boldsymbol{x}$ to cluster i of the current data partition thereby yielding the vector $\boldsymbol{T}^b = [T_1^b, \ldots, T_i^b, \ldots, T_k^b]$, where k is the current number of clusters. Moreover, let $\mathcal{C} = \{c_i | c_i = \arg\max_l (T_l^b)\}$, $|\mathcal{C}| \geq 1$, be the set of current clusters c_i that optimize the iCVI for $\boldsymbol{x}$. Then, a matrix of crisp labels $\boldsymbol{Y} = [y_{i,j}]_{m \times k}$ is generated as

$$y_{i,j} = \begin{cases} 1, & \text{if } j = c_i \\ 0, & \text{otherwise,} \end{cases} \quad i \in \{1, \ldots, m\}, \;\; j \in \{1, \ldots, k\} \tag{15}$$

where $1 \leq m \leq k$ is the number of clusters that equally optimize the selected iCVI, and $c_i \in \mathcal{C}$ is the ith cluster that optimizes the iCVI. Each row $\boldsymbol{y}_i$ of $\boldsymbol{Y}$ is thus a one-hot encoded label vector indicating a cluster assignment that optimizes the selected iCVI. If there is a unique optimal assignment for the presented sample $\boldsymbol{x}$ (i.e., $|\mathcal{C}| = 1$—which is the case for SS-based iCVIs), then $\boldsymbol{Y}$ reduces to a single row vector as defined in [30]; otherwise, matrix $\boldsymbol{Y}$ addresses the case of multiple optimal assignments (i.e., $|\mathcal{C}| > 1$—which may occur when using the iConn_index). The generation of $\boldsymbol{Y}$ depends on the nature of the iCVI: Algorithms 2 and 3 summarize the main steps for the SS-based iCVIs and iConn_index, respectively.

2) iCVI-Based Match Tracking: If the iCVI-based match tracking is enabled, then it is engaged as follows: when the iCVI tracker variable (υ) reaches a user-defined predetermined threshold value (τ), then the vigilance parameter of module A (ρ_a) is altered as $\rho_a \leftarrow \max[\min(\rho_a + \varepsilon_{\text{icvi}}, \rho_{\text{MT}_{\text{icvi}}}), 0]$ if the match function is (7) or as $\rho_a \leftarrow \min[\max(\rho_a - \varepsilon_{\text{icvi}}, \rho_{\text{MT}_{\text{icvi}}}), 2]$ if the match function is (8), where $\varepsilon_{\text{icvi}}$ is the iCVI-based match tracking parameter and $\rho_{\text{MT}_{\text{icvi}}}$ is the iCVI-based match tracking vigilance parameter. For an instantaneous change, $\varepsilon_{\text{icvi}}$ can be automatically set, for instance, as $(\rho_{\text{MT}_{\text{icvi}}} - \rho_a)$ when using the standard match function (7). When υ falls below τ, then ρ_a is reset to its original user-defined value.

3) iCVI Update: After the dynamics of module A and the map field have taken place during training, the online iCVI module incrementally updates its state. This is accomplished using the label associated module A's resonant category J_1 i.e., $\arg\max_i (\boldsymbol{w}_{J_1,i}^{ab})$ and either the presented sample $\boldsymbol{x}$ for SS-based iCVIs or the first and second (if any) prototype winners and the CONN matrix for the iConn_index.

Algorithm 2 Label Generation Using SS-Based iCVIs

Input : Current sample ($\boldsymbol{x}$).
Output : Label matrix $\boldsymbol{Y}$.

```
/* Notation */
s_i = {n_i, μ_i, CP_i}: statistics of cluster i.
Q: set of iCVI-specific quantities.
s'_i, Q': temporary s_i and Q, respectively.
T^b ← 0⃗.
for i ← 1 to k (i.e., each cluster) do
    x is temporarily assigned to cluster i
    Compute s'_i using s_i and x ((10)–(12) with i in place
    of J_1 to represent cluster i).
    Compute Q' using {Q \ s_i} ∪ {s'_i}.
    Compute the iCVI value (T_i^b) using Q'.
Get the set C = {c_i | c_i = arg max_l (T_l^b)}.
Generate the label matrix Y (15).
```

Algorithm 3 Label Generation Using iConn_index

Input : Current sample ($\boldsymbol{x}$), module A (MA), and map field (MF).
Output : Label matrix $\boldsymbol{Y}$.

```
/* Notation */
MA', MF': copies of module A and map
field, respectively.
J = {J_1, J_2}: 1st & 2nd resonant categories.
L = {l_1, l_2}: labels of 1st & 2nd resonant
categories (l_i = arg max_j (w^ab_{J_i,j}), i ∈ {1, 2}).
CONN': CONN matrix within MA'.
Q: set of iCVI-specific quantities.
Q': temporary Q.
T^b ← 0⃗.
for i ← 1 to k (i.e., each cluster) do
    MA' ← MA and MF' ← MF.
    Create a one-hot encoded label y for cluster i.
    Present x to MA' (Alg. 1 using y for the vigilance
    test of MF'). Get set J.
    Present J_1 and y to MF' (Alg. 4). Get set L.
    Compute Q' using Q, CONN', J, L.
    Compute the iCVI value (T_i^b) using Q'.
Get the set C = {c_i | c_i = arg max_l (T_l^b)}.
Generate the label matrix Y (15).
```

C. Map Field

The map field associative network is a variant of fuzzy ARTMAPs [21], [22] that has been enhanced as follows.

1) Vigilance Test and Match Tracking: The iCVI-TopoARTMAP allows for a label matrix that can possibly contain multiple cluster assignments in the one-hot encoded form, and this must be considered by the dynamics of the map field. Therefore, if necessary, the map field may perform multiple vigilance tests during its dynamics, and then output the maximum across all possible cluster assignments corresponding to the label matrix generated by the online iCVI module.

Specifically, the effective map field match value with respect to category J_1 ($M_{J_1}^{ab}$) is given by the maximum among the match values $M_{J_1,i}^{ab}$

$$M_{J_1}^{ab} = \max_i \left[M_{J_1,i}^{ab} \right] \tag{16}$$

where

$$M_{J_1,i}^{ab} = \frac{\left\| \boldsymbol{y}_i \wedge \boldsymbol{w}_{J_1}^{ab} \right\|_1}{\|\boldsymbol{y}_i\|_1} \tag{17}$$

and $\boldsymbol{w}_{J_1}^{ab}$ is the J_1^{th} row vector of the map field mapping matrix $\boldsymbol{W}^{ab} \in \mathbb{R}^{P\times k}$ associated with category J_1, and $\boldsymbol{y}_i$ corresponds to the ith row of the label matrix $\boldsymbol{Y}$. Note that if a supervised label $\boldsymbol{y}^{(\text{SL})}$ is provided, then the following map field vigilance test would be performed instead of (16), (17):

$$M_{J_1}^{ab} = \frac{\left\| \boldsymbol{y}^{(\text{SL})} \wedge \boldsymbol{w}_{J_1}^{ab} \right\|_1}{\left\| \boldsymbol{y}^{(\text{SL})} \right\|_1}. \tag{18}$$

To learn, the current module A resonant category J_1 must satisfy (6) and (7) [or (6) and (8)] as well as the map field vigilance test

$$M_{J_1}^{ab} \geq \rho_{ab} \tag{19}$$

where $0 \leq \rho_{ab} \leq 1$ is the vigilance parameter of the map field. If module A's category J_1 does not also satisfy (19), then a mismatch occurs, and match tracking engages [20], [21], [43] thus modifying the vigilance parameter by a small value ε. In case of a standard match function (7), the vigilance is changed to

$$\rho_a \leftarrow \max\left[\min\left(M_{J_1}^{a} + \varepsilon, 1\right), 0\right] \tag{20}$$

and in case of the cosine-distance-based match function (8), then ρ_a is instead changed using

$$\rho_a \leftarrow \min\left[\max\left(M_{J_1}^{a} - \varepsilon, 0\right), 2\right] \tag{21}$$

where ε is the map field match tracking parameter. Then, category J_1 is inhibited, the search continues with the next highest ranked category, and the process repeats until a resonant category is found or the set of categories is exhausted and a new one is created (see Section II-A).

Note that the map field vigilance test is disregarded until a second cluster is detected. This will delay the computation of the selected iCVI value until a second category is created as per the user-selected vigilance parameter of module A; however, iCVI-specific quantities are still continuously and incrementally updated.

2) Learning (Algorithm 4): The map field adaptation is given by

$$\boldsymbol{w}_{J_1}^{ab} \leftarrow (1-\beta_{ab})\boldsymbol{w}_{J_1}^{ab} + \beta_{ab}\left(\boldsymbol{x}^{F_{ab}} \wedge \boldsymbol{w}_{J_1}^{ab}\right) \tag{22}$$

where $0 < \beta_{ab} \leq 1$ is the map field learning rate and $\boldsymbol{x}^{F_{ab}}$ is equal to $\boldsymbol{y}^{(\text{SL})}$, the lth row of $\boldsymbol{Y}$ ($l = \arg\max_i(M_{J_1,i}^{ab})$), or $\vec{\boldsymbol{1}}$, corresponding to the cases in which a supervised label is provided, the online iCVI module generated a label matrix, or the map field vigilance test was disabled, respectively. Note that $\boldsymbol{x}^{F_{ab}} = \boldsymbol{Y}$ for the SS-based iCVIs used here because they yield a single row label matrix. When a new category is created in module A, $\boldsymbol{W}^{ab}$ is updated according to hyperparameter L_{type} (Table I).

1) *Variable Number of Clusters:* $\boldsymbol{W}^{ab}$ is expanded (1 row and 1 column) and $\boldsymbol{x}^{F_{ab}} = [0\cdots 0\ 1] \in \mathbb{R}^{k+1}$.

Algorithm 4 Map Field Learning

Input : resonant category (J_1), L_{type}, one-hot encoded label ($\boldsymbol{y}$).
Output : Trained Map Field.

```
if Module A created a new category then
    if L_type = variable then
        x^{F_ab} ← [0 ··· 0 1]_{1×k+1}.
        Concatenate W^ab and column vector 0⃗.
    else
        x^{F_ab} ← y
    Concatenate W^ab and row vector 1⃗.
else
    if the map field vigilance test is disabled then
        x^{F_ab} ← 1⃗
    else
        x^{F_ab} ← y
Update: w_{J_1}^ab (22).
```

2) *Fixed Number of Clusters:* $\boldsymbol{W}^{ab}$ is expanded (1 row). The new category must be assigned to a single cluster for learning to take place (line 6 of Algorithm 4), and thus, only SS-based iCVIs are supported in this mode (because, in this case, $\boldsymbol{Y}$ is a row vector).

D. Postprocessing Strategies

When processing data streams, incorrect decisions may be made due to, for instance, ordering effects, no knowledge of the data range, and system hyperparameterization. Therefore, online learning systems usually make use of heuristics to improve performance such as some combination of split and merge [18], [37], [44], [45]; however, to the best of our knowledge, with the exception of [18] and [16], these do not make use of iCVIs. Here, iCVI-TopoARTMAP can use any combination of the following five strategies.

1) Swap: The system attempts to swap categories among clusters if such an operation incurs a better iCVI value. The system emulates a constrained swap operation for all the categories from their original to another cluster. The set of clusters to which a category can be swapped is limited to the clusters it is connected to as per the CONN matrix: a category c from cluster i can be swapped to cluster j if and only if it is connected to at least one category belonging to cluster j, i.e., if $\text{CONN}(c, g) > 0$ for some category g of cluster j. If the best iCVI value among these swap operations incurs an improvement of the original clustering structure as measured by iCVI, then this operation effectively takes place. The swap strategy is repeated until no swap operation leads to an iCVI improvement. In addition, a swap operation is only carried out if there are currently more than two categories. Note that a cluster represented by a single category will disappear if it is selected to undergo a swap operation; however, a swap whose outcome would be a single unique cluster is not allowed. Concretely, a swap operation entails

removing a category c (i.e., a split) from a cluster i and immediately assigning it to another cluster j (i.e., a merge).

The swap strategy relates to the heuristic in the iCVI-ARTMAP model [30]. However, the key difference is that instead of swapping samples, categories are swapped between clusters. The swap strategy is similar to the approach presented in [46] in the context of offline clustering, wherein prototype labels are swapped to optimize an information-theoretic bCVI.

2) Merge: The system emulates consecutive cluster merging until only two clusters are left, and the best cluster structure, according to the selected iCVI, is retained. Two clusters i and j are merged if they yield the best iCVI value compared with all other cluster merges at that stage. If the best iCVI value across all the merge stages is an improvement over the current clustering state, then this operation effectively takes place, and the associated cluster structure is retained. The merge strategy is only carried out if there are currently more than two clusters and the iCVI tracker variable is zero ($\upsilon = 0$).

3) Split: The most recently active category i (as per $\bar{a}_i$) is enforced to constitute a cluster on its own. Note, however, that if the category selected for splitting is a singleton, then the search proceeds with the subsequent most recent category j ($\bar{a}_j \leq \bar{a}_k, \forall k \neq i$). This process continues until 1) a new cluster consisting of a single category is created or 2) no category could become a new cluster and thus the split operation could not take place. The split strategy is only carried out if the iCVI tracker variable is greater than a user-defined threshold ($\upsilon > \tau$). This activity-based split heuristic is used in the experiments reported in Section III; iCVI-based split alternatives are discussed in the supplementary material.

4) Compress: In many applications, memory requirements must be observed, and thus iCVI-TopoARTMAP is equipped with a compression strategy to reduce its memory footprint (considering the number of categories as a proxy for model size). The system attempts to compress the categories of module A using a modified fuzzy ARTMAP. Let $\mathcal{H}$ be the set of iCVI-TopoARTMAP categories whose inactivity is greater or equal to a user-defined threshold ξ and let $\bar{\mathcal{H}}$ be its complement, i.e., $\mathcal{H} = \{\text{category } i | \bar{a}_i \geq \xi\}$ and $\bar{\mathcal{H}} = \{\text{category } i | \bar{a}_i < \xi\}$. The compression strategy is only carried out if $\mathcal{H} \neq \{\emptyset\}$. First, iCVI-TopoARTMAP module A's and map field's weights associated with categories from $\bar{\mathcal{H}}$ are copied to initialize the modified fuzzy ARTMAP module A's weight vector matrix $\boldsymbol{H}^a$ and map field's mapping matrix $\boldsymbol{H}^{ab}$. These will not be allowed to learn. Next, the input pairs $(\boldsymbol{w}_i^a, \boldsymbol{w}_i^{ab})$, $i \in \mathcal{H}$, are fed to the modified fuzzy ARTMAP module with its module A vigilance parameter ρ_c and the remaining parameters identical to the ones of iCVI-TopoARTMAP (i.e., ε, α, β_1, and ρ_{ab}).

The dynamics of the modified fuzzy ARTMAP are like a standard fuzzy ARTMAP. When $(\boldsymbol{w}_i^a, \boldsymbol{w}_i^{ab})$ is presented, the activations are computed as

$$T_j = \frac{\left\| \boldsymbol{w}_i^a \wedge \boldsymbol{h}_j^a \right\|_1}{\alpha + \left\| \boldsymbol{h}_j^a \right\|_1} \tag{23}$$

where $\boldsymbol{h}_j^a$ is the jth row of $\boldsymbol{H}^a$, representing the weight vector of category j within module A of the modified fuzzy ARTMAP. A resonant category must simultaneously 1) not belong to $\bar{\mathcal{H}}$ and 2) satisfy the following two equations

$$\frac{\left\| \boldsymbol{w}_i^a \wedge \boldsymbol{h}_j^a \right\|_1}{d} \geq \rho_c, \quad 0 \leq \rho_c \leq 1, \ i \in \mathcal{H} \tag{24}$$

$$\frac{\left\| \boldsymbol{w}_i^{ab} \wedge \boldsymbol{h}_j^{ab} \right\|_1}{d} \geq \rho_{ab}, \quad 0 \leq \rho_{ab} \leq 1, \ i \in \mathcal{H}. \tag{25}$$

The resonant category J updates its weight vector as per (9) (with $\boldsymbol{h}_J^a$ and $\boldsymbol{w}_i^a$ in place of $\boldsymbol{w}_{J_1}$ and $\boldsymbol{x}^a$, respectively). Note that the entry of the map field matrix associated with a category remains fixed after it is first assigned (e.g., by initialization or creation). If a category satisfies (24) but not (25), then standard match tracking engages with parameter ε. If a new category is created, then $\boldsymbol{h}_{\text{new}}^a \leftarrow \boldsymbol{w}_i^a$ and $\boldsymbol{h}_{\text{new}}^{ab} \leftarrow \boldsymbol{w}_i^{ab}$.

Note that categories created by the modified fuzzy ARTMAP are appended to set $\mathcal{H}$. The training is performed until there is no change in weight vectors $\boldsymbol{h}_i^a$ from one epoch to another. This procedure can be thought of as a form of "self-supervised" learning for iCVI-TopoARTMAP, since the label for each category $i \in \mathcal{H}$ with weight vector $\boldsymbol{w}_i^a$ is

$$\text{label}_i = \arg\max_i \left(\boldsymbol{w}_i^{ab} \right), \quad i \in \mathcal{H}. \tag{26}$$

After compression, if the number of categories of the modified fuzzy ARTMAP (P_{new}) is smaller than the number of categories of iCVI-TopoARTMAP (P_{old}), then the category and map field weights of the former are used to overwrite the latter: $\boldsymbol{W}^a \leftarrow \boldsymbol{H}^a$ and $\boldsymbol{W}^{ab} \leftarrow \boldsymbol{H}^{ab}$, where $\boldsymbol{W}^a$ and $\boldsymbol{W}^{ab}$ represent the category weight and map field matrices of iCVI-TopoARTMAP, respectively. Furthermore, iCVI-TopoARTMAP's variables must be updated to reflect such change. For instance, let $\mathcal{B}_i$ be the set of iCVI-TopoARTMAP module A's categories mapped to the trained modified fuzzy ARTMAP category i, and thus, the following is performed.

1) *Local Statistics:* The local statistics of the categories belonging to $\mathcal{B}_i$ are used to incrementally compute the local statistic associated with category i of the trained modified fuzzy ARTMAP via merges [(33)–(35), where both j and c would represent categories].
2) *Inactivity:* The inactivity of category i of the trained modified fuzzy ARTMAP is defined as the smallest among the categories belonging to $\mathcal{B}_i$.

Items (a) and (b) above are performed for all the categories of the trained modified fuzzy ARTMAP and used to overwrite the respective quantities of iCVI-TopoARTMAP's module A. Finally, iCVI-TopoARTMAP's $\text{CONN}_{P_{\text{old}} \times P_{\text{old}}}$ matrix is updated to reflect such compression. Specifically, a new $\text{CONN}_{P_{\text{new}} \times P_{\text{new}}}$ is constructed from $\text{CONN}_{P_{\text{old}} \times P_{\text{old}}}$

$$\text{CONN}_{P_{\text{new}} \times P_{\text{new}}}(i, j) \leftarrow \sum_{k \in \mathcal{B}_i}^{|\mathcal{B}_i|} \sum_{l \in \mathcal{B}_j}^{|\mathcal{B}_j|} \text{CONN}_{P_{\text{old}} \times P_{\text{old}}}(k, l) \tag{27}$$

$\text{CONN}_{P_{\text{new}} \times P_{\text{new}}}$ is then used to overwrite $\text{CONN}_{P_{\text{old}} \times P_{\text{old}}}$ of iCVI-TopoARTMAP. Additional changes include, but are not limited to, removing categories and associated variables for which $\mathcal{B}_i = \{\emptyset\}$.

Note that the online iCVI module's state remains the same for the SS-based iCVIs, whereas these must be recomputed for the iConn_index because of the CONN matrix change.

The compress operation is inspired by the compression step used in [45], the key difference is that here a fuzzy ARTMAP is used for "self-supervised" learning, i.e., the inputs to its module A are the current weight vectors of the iCVI-TopoARTMAP's module A, whereas the inputs to module B are the weight vectors of iCVI-TopoARTMAP's map field. In this manner, the mapping of categories to clusters obtained so far is assumed to encode the true labels of iCVI-TopoARTMAP's module A categories. This operation has the goal of mitigating category proliferation, thus making the iCVI-TopoARTMAP model more compact. The outcome of this operation is a modification to the weight vectors ($\boldsymbol{W}^a$), local statistics, and inactivity counter of module A's categories and the CONN matrix, as well as the map field mapping matrix $\boldsymbol{W}^{ab}$ entries—a change in the online iCVI module's state may also occur if the selected iCVI is iConn_index.

5) Prune-and-Reassign: A category is pruned if 1) its inactive counter is above the threshold parameter ξ and 2) the cluster to which it is associated encoded less than ϕ samples. Let $\mathcal{P}$ be the set of pruned categories; these categories are reassigned to the cluster associated with the closest category that does not belong to $\mathcal{P}$, i.e., the assignment of a category $i \in \mathcal{P}$ is given by

$$\text{label}_i = \arg\max_j \left(T_j^a \right), \quad j \notin \mathcal{P} \tag{28}$$

where

$$T_j^a = \frac{\left\| \boldsymbol{w}_i^a \wedge \boldsymbol{w}_j^a \right\|_1}{\alpha + \left\| \boldsymbol{w}_j^a \right\|_1}, \quad i \in \mathcal{P}, \ j \notin \mathcal{P}. \tag{29}$$

This approach is reminiscent of the fuzzy topoART pruning strategy [32]; however, the latter considers the number of samples encoded by a category (not the cluster to which it belongs) and deletes such category. Moreover, it does not consider a category inactivity counter. The prune-and-reassign strategy is only carried out if $0 < |\mathcal{P}| < P$. Furthermore, this strategy must be accompanied by adequate changes to the map field and the online iCVI module's state.

The operations described in items 1–5 (swap, merge, split, compress, and prune-and-reassign) aim to reorganize the cluster structures, incrementally, for each sample presentation. Moreover, the first three operations aim to improve the partition quality as measured by the selected iCVI. If these strategies take place, then iCVI-TopoARTMAP variables must be changed accordingly; i.e., some or all the following items occur: 1) the online iCVI module updates its state to reflect changes; 2) the map field mapping matrix $\boldsymbol{W}^{ab}$ entries are modified accordingly; and/or 3) the module A variables are modified as per the compression operation. Finally, note that the swap, merge, and split operations depend on the nature of iCVI.

a) SS-based iCVIs: Since each prototype has local statistics associated with it in the original data range, swapping a category c from cluster i to cluster j corresponds to computing the following sequentially:

1) *Split:*

$$\boldsymbol{\mu}_i \leftarrow \frac{n_i}{n_i - n_c}\boldsymbol{\mu}_i - \frac{n_c}{n_i - n_c}\boldsymbol{\mu}_c \tag{30}$$

$$\text{CP}_i \leftarrow \text{CP}_i - \text{CP}_c - \frac{n_i n_c}{n_i - n_c}\|\boldsymbol{\mu}_c - \boldsymbol{\mu}_i\|_2^2 \tag{31}$$

$$n_i \leftarrow n_i - n_c. \tag{32}$$

2) *Merge:*[3]

$$\boldsymbol{\mu}_j \leftarrow \frac{n_j}{n_j + n_c}\boldsymbol{\mu}_j + \frac{n_c}{n_j + n_c}\boldsymbol{\mu}_c \tag{33}$$

$$\text{CP}_j \leftarrow \text{CP}_j + \text{CP}_c + \frac{n_j n_c}{n_j + n_c}\|\boldsymbol{\mu}_c - \boldsymbol{\mu}_i\|_2^2 \tag{34}$$

$$n_j \leftarrow n_j + n_c. \tag{35}$$

The merge operation is carried out like [30]. In particular, merging is accomplished using (33)–(35), where j and c would represent the entire clusters. Splitting a cluster i consists of removing category c that was mapped to it and creating a new cluster l. This is accomplished using (30)–(32). The new cluster l statistics are equal to the ones associated with category c: $n_l \leftarrow n_c$, $\boldsymbol{\mu}_l \leftarrow \boldsymbol{\mu}_c$, and $\text{CP}_l \leftarrow \text{CP}_c$.

b) iConn_index: Swapping, merging clusters i and j, and splitting cluster i entail changing the mapping of categories to clusters, while the CONN matrix remains the same.

E. Training

Algorithm 5 shows an online training procedure for iCVI-TopoARTMAP. It starts by presenting the input(s) to the network. If the current input(s) is/are the very first fed to the system, then the input(s) is/are used to initialize the network, i.e., module A, map field, and online iCVI module's state. For instance, the weight vector and local statistics of module A are initialized using (13), the CONN matrix is initialized as $[0]_{1\times 1}$, the inactivity of the first category is initialized as $\bar{a}_1 = 0$, and the map field mapping matrix $\boldsymbol{W}^{ab}$ is initialized as $[1]_{1\times 1}$. Next, the current value of iCVI is temporarily stored for future comparison.

The existence of a supervised class label $\boldsymbol{y}^{(\text{SL})}$ is verified by the system, and, in the affirmative case, then the $(\boldsymbol{x}, \boldsymbol{y}^{(\text{SL})})$ pair is used to train module A and proceed with the remaining training steps; in the negative case, then the number of clusters (k) detected by the system is verified, and if it is equal to one, then the map field vigilance test is disabled for module A training before proceeding with the remaining training steps. Conversely, if $k > 1$ (which is a necessary condition to compute iCVI values), then, at each iteration, the online iCVI module is engaged to generate a cluster label matrix $\boldsymbol{Y}$ using the selected iCVI representing the best placement(s) of the input sample (Section II-B), where an iCVI value is incrementally recomputed for the temporary

[3]In [16], the merging of compactnesses is not discussed nor a formula is given. Inspection of the code provided by the author indicates that in such work, the merge operation is computed as a weighted average (i.e., $\text{CP}_{\text{merge}} \leftarrow (n_1\text{CP}_1 + n_2\text{CP}_2)/(n_1 + n_2)$, where CP_i and n_i are the compactness and frequency associated with prototype $i \in \{1, 2\}$, respectively), which is not an exact computation like (34).

assignment of the presented input sample to each existing cluster. Note that not all the iCVI quantities are recomputed, solely the ones associated with the respective temporary cluster assignment.

In this manner, supervised labels have priority in the system, like the mixed-modality learning system presented in [31]. The capability of also handling supervised labels allows for the inclusion of external knowledge into the system at any time (e.g., initialization of the system using a priori knowledge and semi-supervised online learning mode—i.e., map field with parameter L_{type} set to *fixed*). In cases where supervised labels are provided, the online iCVI module's state is updated in accordance to such sample and label pairs. Note that the input(s) to the online iCVI module can be the raw input sample as well as data from module A and the map field.

The map field and module A components retain many of their original dynamics [14], [21], [22], [32]. Within module A, a copy of the presented sample $\boldsymbol{x}$ undergoes min–max normalization and complement coding, thus becoming the input to module A ($\boldsymbol{x}^a$), whereas its raw copy (non-normalized original sample $\boldsymbol{x}$) is used to update the summary statistics (Section II-A) of the resonant category (or initialize the local statistics of a new category). The raw input is also used for the SS-based iCVI computations performed via the online iCVI module. A sample $\boldsymbol{x}$ and its processed copy are discarded after being presented to the system. When a category resonates (i.e., satisfies both the module and map field resonance checks), then learning ensues as described in Sections II-A and II-C. Following sample assignment, the online iCVI module updates its state as necessary.

After the presentation of each sample, the system attempts to perform enabled postprocessing heuristics (Section II-D). Note that when performing these operations, iCVI-TopoARTMAP internal variables must be updated accordingly. Moreover, the iCVI-specific quantities associated with the current partition are cached [11], [30] to speed up the learning process in operations related to iCVIs. The last step for a given sample presentation consists of comparing the iCVI values at the beginning and the end of the current iteration and updating the iCVI tracker variable: if it is worse, then $\upsilon \leftarrow \upsilon + 1$; otherwise, $\upsilon \leftarrow \max(0, \upsilon - 1)$.

Algorithm 5 iCVI-TopoARTMAP Learning

Input : Streaming data $\boldsymbol{X}$, class label $y^{(SL)}$, and iCVI-TopoARTMAP hyperparameters (Table I).
Output : Trained iCVI-TopoARTMAP.

```
/* Notation */
t: time (proxy for number of samples).
k_t: number of clusters at time t.
t ← 0.
for x ∈ X do
    t ← t + 1
    if t = 1 then
        Initialize iCVI-TopoARTMAP.
    else
        Get current iCVI value (iCVI_start).
        if y^(SL) ≠ {∅} then
            Y ← y^(SL).
        else
            if k_t > 1 then
                Present x to the online iCVI module to
                generate the label matrix Y.
                if EN_MT_icvi then
                    Call the iCVI-based match tracking
                    routine (Section II-B).
            else
                Disable the map field's vigilance test.
        Run the dynamics of module A (Section II-A).
        Run the dynamics of map field (Section II-C).
        Update the iCVI module state (Section II-B).
        Postprocessing (Section II-D)
        if EN_merge then
            Call the merge strategy routine.
        if EN_split then
            Call the split strategy routine.
        if EN_swap then
            Call the swap strategy routine.
        if EN_compress then
            Call the compress strategy routine.
        if EN_prune_reassign then
            Call the prune-and-reassign strategy routine.
        Get current iCVI value (iCVI_end).
        if iCVI_start is better than iCVI_end then
            υ ← υ + 1.
        else
            υ ← max(0, υ − 1).
```

III. Experiments

A. Datasets

In this work, experiments were carried out with 11 synthetic and 3 real-world datasets (Table II).

1) Synthetic Data: This type of data is often used in the literature [47], [48]. The benchmark datasets listed in Table II comprise a mix of properties, such as a different number of clusters, degrees of overlap, and class imbalance. Subsets of this collection of datasets have been used in several recent data stream studies, such as [6], [9], [10], [16], [17], and [49].

2) Real-World Image Data: The FEI dataset [50] consists of the faces of 200 individuals (14 images per individual) taken under a controlled environment with different illumination settings. Individuals vary in age, poses, and facial expressions.

3) Real-World Sensor Data: Sensor datasets are frequently used in the literature given their streaming nature [48]. Moreover, sensor networks are a typical use case for the data stream algorithms [47]. In this work, experiments were carried out with the gas sensor array under dynamic gas mixtures (GSA) [51] and the occupancy detection (OD) [52] datasets, which have been used in recent data stream studies [6], [53].

TABLE II
DATASETS' CHARACTERISTICS: N, d, AND k REPRESENT THE NUMBER OF SAMPLES, ATTRIBUTES, AND CLUSTERS, RESPECTIVELY

	Data set	N	d	k
Synthetic	DS0[a] [16]	1600	2	7
	DS1[b] [6]	1100	2	11
	DS2[b] [6]	1100	2	11
	A1[c] [54], [55]	3000	2	20
	A2[c] [54], [55]	5250	2	35
	A3[c] [54], [55]	7500	2	50
	S1[c] [54], [56]	5000	2	15
	S2[c] [54], [56]	5000	2	15
	S3[c] [54], [56]	5000	2	15
	S4[c] [54], [56]	5000	2	15
	Unbalance[c] [54], [57]	6500	2	8
Real-world	FEI[d] [50]	2800	$640 \times 480 \times 3$	200
	GSA [51], [58] (subset)	300	8	3
	OD (training) [52], [58]	8143	5	2
	OD (test 1) [52], [58]	2665	5	2
	OD (test 2) [52], [58]	9752	5	2

The data set names *DS1* and *DS2* are used to avoid duplicate nomenclature: the names in the corresponding references are S2 and S3, respectively. The data set name *DS0* was used for consistency in this work.
[a] DS0 is similar to the data set used in [16]: https://github.com/iskmeans/iskm.
[b] Data sets generated in-house. They are similar to the ones used in [6].
[c] Clustering basic benchmark: http://cs.uef.fi/sipu/datasets/
[d] Available at https://fei.edu.br/~cet/facedatabase.html

All aforementioned datasets were used for benchmark-only purposes within the scope of this research article.

B. Experimental Setup

1) Synthetic Data: Input ordering tends to have a dramatic effect on the performance of online agglomerative clustering algorithms such as ART [59]. Therefore, the unsupervised experiments were conducted in class-incremental order [60] (i.e., cluster-by-cluster) and random order of sample presentation. Moreover, for each ordering, experiments were repeated twice per dataset. The semi-supervised learning experiments were conducted five times per dataset with random ordering.

2) Real-World Image Data: Recently, the ART-based systems have been applied to process embeddings generated by deep neural networks, such as clustering in [30] and object detection in [61]. For online unsupervised and semi-supervised learning purposes with the FEI dataset, this work uses a standard face recognition pipeline [42], [62], which, in its most succinct form, consists of a face detector (MTCNN[4] [63] pretrained model), a feature extractor (VGGFace2[5] [42] pretrained model providing a 2 048-D feature vector for a given face), and a classifier (e.g., iCVI-TopoARTMAP model). We extended the cropped area of the face bounding boxes by 30% like [42] and set the confidence threshold to 0.99. However, unlike [42], ℓ_2 normalization was not used on the embeddings prior to presentation to any algorithm used here (Section III-C). The face detector failed in detecting some faces; these were disregarded. Hence, the total number of faces was 2 774. Experiments were carried out like the synthetic datasets in terms of orderings for sample presentation and number of repetitions.

3) Real-World Sensor Data: Like in [6], the experiments with the GSA dataset were conducted using the first 5 min of readings from 8 sensors for the gas mixture of ethylene and CO, wherein batches of 100 samples were averaged as they arrive, thereby yielding the processed subset with the properties listed in Table II. Note that in this work, the label (gas mixture concentration) of the processed subset is defined as the majority of labels per batch because there are transitions that are not multiple of 100. Unsupervised and semi-supervised experiments were conducted once and five times per dataset, respectively, wherein samples were presented in their natural temporal order.

In all the semi-supervised learning experiments carried out in this work, a single sample from each class was randomly chosen for training (initialization in the case of iCVI-TopoARTMAP). The remaining samples were used for testing and for iCVI-TopoARTMAP semi-supervised learning.

The evaluation protocol consisted of presenting—at the end of the stream—the entire data to each method, recording their predictions, and then measuring their performance. The algorithms (Section III-C) were evaluated using either the adjusted rand index (ARI)[6] [64] or classification accuracy (ACC), which correspond to the cases of unsupervised and semi-supervised learning, respectively.

C. Algorithms, Implementation, and Reproducibility

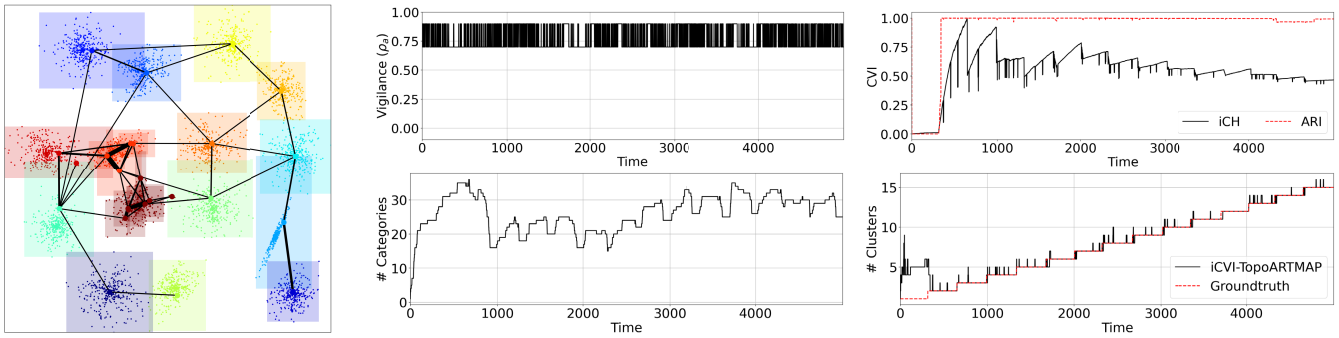
The iCVI-TopoARTMAP in unsupervised learning mode was compared with sequential k-means (skm) [65], the iXB-based method of incremental sequential k-means (iskm) [16], and the following ART-based clustering algorithms: dual vigilance fuzzy ART (DVFA) [66], topoFA [32]—only a single module of topoFA was used, and developmental resonance network (DRN) [36].

Note that except for DRN, none of the previously mentioned ART-based models can handle use cases in which the data ranges are unknown, i.e., they require a priori knowledge of the data maximum and minimum statistics to perform min–max normalization and subsequent complement coding [25], [33], [34], [36], [67]. Therefore, to benchmark iCVI-TopoARTMAP against a larger pool of state-of-the-art ART-based models, we equipped DVFA and topoFA with online normalization of inputs, online complement coding, and weight vector rescaling [33], [34] (Section II-A).

The performance of iCVI-TopoARTMAP in the semi-supervised mode was compared with the prediction of a 1-nearest neighbor (1-NN) classifier [39]. The latter is commonly used in face recognition tasks with deep feature descriptors [41]. Since the FEI face image dataset is used in this work, 1-NN was chosen for consistency across experiments.

[4] Available at https://github.com/ipazc/mtcnn
[5] Available at https://github.com/rcmalli/keras-vggface

[6] Most clustering algorithms used in this work do not have a parameter that explicitly controls the number of clusters: these are dynamically created. The ARI measures the agreement between some partition and ground truth, thus enabling the comparison of the algorithms' partitions even if they comprise different numbers of clusters: larger values of ARI indicate better partitions, wherein ARI = 1 corresponds to a perfect agreement.



(a) (b)

Fig. 2. Experiment with iCVI-TopoARTMAP and the S1 dataset (presented in a class-incremental order) as per the parameterization associated with the results reported in Table III. (a) Color-coded partition and footprints. The connections shown between categories represent the CONN matrix (thicker lines represent stronger connections)—see CONNvis [35]. Although categories might be connected, the clusters in iCVI-TopoARTMAP are determined by the map field. (b) Tracking of iCVI-TopoARTMAP: vigilance parameter of module A (ρ_a), number of categories, number of clusters, iCH (values were normalized for visualization purposes), and ARI values over time.

It used the Euclidean distance for the synthetic and sensor datasets and cosine distance for the FEI dataset.

The iCVI-TopoARTMAP, DVFA, topoFA, 1-NN, and skm python code are provided at Guise AI's GitHub repository.[7] Some of the iCVI-TopoARTMAP components are based on the *iCVI-toolbox for MATLAB*[8] [11]. The code for iskm is an in-house python port of the MATLAB code from[9] [16]. The DRN code is from[10] [68]. ARI, and ACC computations as well as the min–max normalizations were made using *scikit-learn* [69]. Finally, the statistical analysis was carried out in R with the scmamp package [70].

D. Results

All iCVI-TopoARTMAP results reported here were obtained using iCH, as per its good performance and execution time. The results for other iCVIs are included in the supplementary material.

1) Synthetic Data: To illustrate iCVI-TopoARTMAP learning over time, Fig. 2 depicts this model's dynamics when clustering the S1 dataset and the obtained partition. It can be observed that: 1) the iCVI-based match tracking is constantly triggered as module A's vigilance parameter toggles from the baseline 0.7 (ρ_a) to 0.9 ($\rho_{\text{MT}_{\text{icvi}}}$) and 2) there are several moments in which the compression strategy reduces the number of categories. Moreover, Fig. 2 also shows that iCVI-TopoARTMAP achieved and maintained high ARI values across the majority of the data stream (roughly from the emergence of the second true cluster onward)—the evaluation consisted of representing each sample up to time t, predicting their classes, and computing the ARI.

Table III shows that iCVI-TopoARTMAP outperformed all other ART-based methods in the online clustering experiments across different orderings. Table IV shows that regarding the test accuracy, iCVI-TopoARTMAP in the semi-supervised mode outperformed 1-NN in all the experiments.

[7]Available at https://github.com/GuiseAI/iCVI-TopoARTMAP
[8]Available at https://github.com/ACIL-Group/iCVI-toolbox
[9]Available at https://github.com/iskmeans/iskm
[10]Available at https://github.com/Uehwan/Incremental-Learning

TABLE III

RESULTS OF THE UNSUPERVISED LEARNING EXPERIMENTS. THE BEST PERFORMANCES (ARI) ARE REPORTED IN BOLD

data set	skm	iskm	DRN	DVFA[a]	TopoFA[a]	iCVI-TopoARTMAP
Class-incremental order (synthetic and real-world image data)						
DS0	0.53994	0.98716	0.83399	0.71852	0.34860	**0.99445**
DS1	0.37661	0.99619	0.96604	0.78869	0.67632	**0.99799**
DS2	0.52242	**0.92669**	0.71054	0.54921	0.30016	0.91490
A1	0.26948	**0.99615**	0.76198	0.69884	0.31689	0.97814
A2	0.25548	0.78441	0.73529	0.60794	0.07504	**0.95815**
A3	0.18374	0.76944	0.87301	0.76014	0.12401	**0.97288**
S1	0.39743	0.94199	0.89842	0.77433	0.32500	**0.98904**
S2	0.49663	0.91958	0.74788	0.66580	0.00005	**0.93652**
S3	0.47896	0.64609	0.51413	0.48952	0.00002	**0.68493**
S4	0.33220	0.57024	0.42331	0.43790	0.00005	**0.59559**
Unbalance	0.45895	0.99946	0.99792	0.87213	0.56368	**1.00000**
FEI	0.08122	0.38504	0.94686	0.79921	0.78344	**0.97714**
Random order (synthetic and real-world image data)						
DS0	0.91180	0.00000	0.77163	0.91529	0.42156	**0.98475**
DS1	0.83764	0.00000	0.89405	0.94046	0.18396	**1.00000**
DS2	0.77097	0.00000	0.60359	0.69257	0.19787	**0.91606**
A1	0.74796	0.00000	0.62747	0.67903	0.05151	**0.97188**
A2	0.73635	0.00000	0.71265	0.56909	0.01109	**0.98155**
A3	0.75070	0.00000	0.62983	0.60796	0.00000	**0.96932**
S1	0.79337	0.00000	0.56103	0.85935	0.40998	**0.98572**
S2	0.74552	0.29119	0.49512	0.78415	0.13253	**0.93414**
S3	0.67352	0.18556	0.38160	0.55469	0.00015	**0.72696**
S4	0.54613	0.42382	0.36442	0.51277	0.00002	**0.62772**
Unbalance	0.65965	0.06419	0.99051	0.99939	**1.00000**	**1.00000**
FEI	0.42588	0.50503	0.90784	0.74847	0.88473	**0.97841**
Chronological order (real-world sensor data)						
GSA	0.40086	0.49222	**0.57251**	0.49231	0.48843	0.54767
OD (training)	0.78661	0.83697	0.81153	0.60871	0.35721	**0.87349**
OD (test 1)	0.74027	0.77989	0.67332	0.48951	**0.88311**	0.83819
OD (test 2)	0.45267	0.61305	0.52455	0.42760	0.27947	**0.73284**

Bold values correspond to (i) peak average *ARI* over 2 runs for the class-incremental and random order experiments and (ii) peak *ARI* over 1 run for the chronological order experiments.
[a] Augmented versions of these algorithms (see Section III-C).

A statistical analysis was carried out following the procedures described in [70]. For the unsupervised learning experiments, iCVI-TopoARTMAP was separately compared with two groups (ART- and non-ART-based methods), wherein first the Iman–Davenport's test (omnibus test) was performed and, if there was an evidence of statistical difference among the algorithms, the Bergmann and Hommel's test (*post hoc* test) was conducted. For the semi-supervised learning experiments, iCVI-TopoARTMAP and 1-NN test performances were compared using the Wilcoxon Signed-Rank test. The analysis indicates that under a 0.1 significance level (the p-values are provided within the supplementary material), iCVI-TopoARTMAP: 1) statistically outperforms skm,

TABLE IV

RESULTS OF THE SEMI-SUPERVISED LEARNING EXPERIMENTS. THE BEST PERFORMANCES (ACC) ARE REPORTED IN BOLD

data set	1-NN		iCVI-TopoARTMAP	
	train	test	train	test
Random order (synthetic and real-world image data)				
DS0	**1.0000**	0.9803	**1.0000**	**0.9938**
DS1	**1.0000**	0.9996	**1.0000**	**1.0000**
DS2	**1.0000**	0.9188	**1.0000**	**0.9611**
A1	**1.0000**	0.9567	**1.0000**	**0.9920**
A2	**1.0000**	0.9430	**1.0000**	**0.9923**
A3	**1.0000**	0.9426	0.9920	**0.9929**
S1	**1.0000**	0.9815	0.9867	**0.9931**
S2	**1.0000**	0.9346	0.9733	**0.9681**
S3	**1.0000**	0.7628	0.8933	**0.8238**
S4	**1.0000**	0.6868	0.8800	**0.7438**
Unbalance	**1.0000**	0.9997	**1.0000**	**1.0000**
FEI	**1.0000**	**0.9907**	**1.0000**	0.9903
Chronological order (real-world sensor data)				
GSA	**1.0000**	0.6875	**1.0000**	**0.7279**
OD (training)	**1.0000**	0.9273	**1.0000**	**0.9516**
OD (test 1)	**1.0000**	0.9072	0.9000	**0.9179**
OD (test 2)	**1.0000**	**0.7967**	0.9000	0.7282

Bold values correspond to peak average ACC over 5 runs.
While the 1-NN only predicts, the iCVI-TopoARTMAP continuously learns, unsupervisedly, with each incoming test sample.

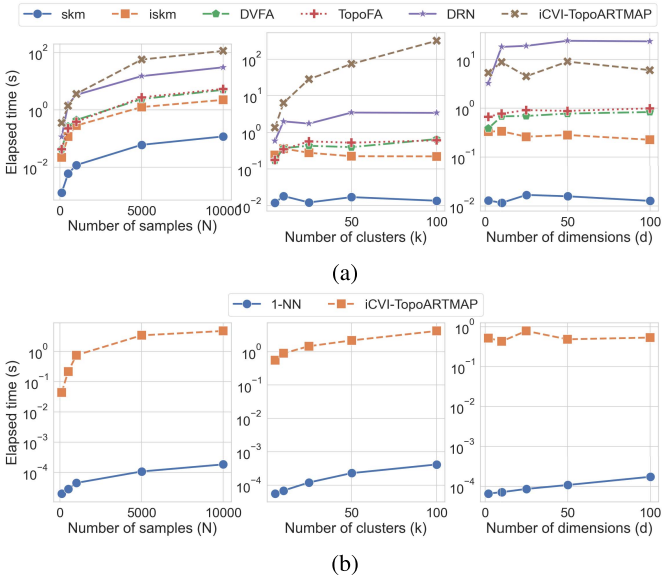


Fig. 3. Elapsed time of algorithms when varying the number of samples ($N \in [100, 500, 1,000, 5\,000, 10\,000]$, $k = 5$, $d = 2$), clusters ($N = 1\,000$, $k \in [5, 10, 25, 50, 100]$, $d = 2$), and dimensions ($N = 1\,000$, $k = 5$, $d \in [2, 10, 25, 50, 100]$) of a toy synthetic dataset generated using scikit-learn. (a) Main vigilance parameter of all the ART models, the number of clusters of skm, and the forgetting factor of iskm were set to 0.9, the ground truth, and 0.99, respectively. All the iCVI-TopoARTMAP postprocessing strategies are enabled and $\rho_c = 0$. (b) $\rho_a = 0$ in iCVI-TopoARTMAP and all the postprocessing strategies are disabled. iCVI-TopoARTMAP uses iCH in both (a) and (b).

is equivalent to iskm, and outperforms all the ART-based methods for the class-incremental order of presentation and 2) statistically outperforms all other methods for random order presentation. Regarding the semi-supervised learning experiments, iCVI-TopoARTMAP statistically outperforms 1-NN at a 0.05 significance level (p-value = 0.001673).

2) Real-World Image Data: Table III shows that iCVI-TopoARTMAP outperformed all the methods across both orderings in the online clustering experiments. Table IV shows that 1-NN outperformed iCVI-TopoARTMAP in the semi-supervised mode by a tiny margin.

3) Real-World Sensor Data: Table III shows that iCVI-TopoARTMAP outperformed the non-ART-based methods in all the datasets and the ART-based methods in two out of four sensor datasets (it still ranked second for the GSA and OD (test 1) datasets). Table IV shows that regarding the test accuracy, iCVI-TopoARTMAP outperformed 1-NN in most of the experiments with sensor data.

E. Discussion

1) Computational Time: The tradeoff for the often superior performance of iCVI-TopoARTMAP is its larger computational cost compared with the other algorithms (Fig. 3). Its execution time may significantly vary depending on the hyperparameterization, particularly with respect to the chosen iCVI and the enabled postprocessing strategies. Fig. 3 shows that skm is faster than iskm and that DVFA and topoFA were the fastest among the ART models, followed by DRN [Fig. 3(a) shows that the latter tended to be slower than iCVI-TopoARTMAP when the number of dimensions increased]. Fig. 3(b) shows that as expected, 1-NN is faster than iCVI-TopoARTMAP, because its computations simply consist of finding the closest neighbor and there is only one prototype per class (see Section III-B).

2) Known Versus Unknown Data Range: The vigilance parameter of an ART network constrains the maximum hyperbox size of its categories [14]. When the dataset range is unknown a priori (but tracked), it continuously expands with each incoming sample. Therefore, categories generated at the beginning of the data stream are smaller than the ones generated later in time for the same vigilance value, which may cause category proliferation. This phenomenon is clearly observed in Fig. 4, which depicts the outcome of a class-incremental presentation of the DSO dataset to iCVI-TopoARTMAP using iCH. In Fig. 4(a), the categories representing the top two clusters (presented first) are much smaller than the ones representing the bottom five clusters (presented last). Fig. 4(b) shows that the compression strategy (Section II-D) can mitigate this issue and reduce category proliferation.

To investigate the implication of the a priori knowledge of the data ranges, experiments were conducted under the same setup described in Section III-B, but using the data ranges for min–max normalization of samples for all the algorithms. The results obtained (detailed in the supplementary material) indicate that: 1) iCVI-TopoARTMAP in the unsupervised mode still outperformed all other clustering methods for most synthetic and FEI datasets; 2) iCVI-TopoARTMAP in the semi-supervised mode still outperformed 1-NN for all the synthetic datasets, while the latter yielded superior (but comparable) performance for the FEI dataset; and 3) the performance degraded for many algorithms across the experiments

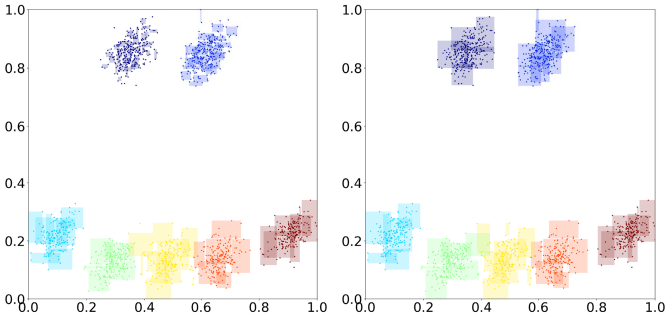


Fig. 4. Effect of the data range expansion on the category size. The iCVI-TopoARTMAP categories (color-coded by cluster) are depicted when the DS0 dataset range is unknown and the compression strategy is (a) disabled and (b) enabled ($\rho_c = \rho_a$).

with the sensor datasets: in the unsupervised experiments, iCVI-TopoARTMAP and DRN yielded the best performance for GSA and OD (training), respectively, whereas skm for OD (test 1) and (test 2); in the semi-supervised experiments, iCVI-TopoARTMAP yielded the best performance for GSA and OD (test 1), whereas 1-NN for OD (training) and OD (test 2).

3) Robustness to Vigilance Parameter Setting: The vigilance parameter is critical to the performance of the ART-based models, and most of them (including the ones in this work) require a manual setting [25]. Parameterization of the online clustering algorithms is a challenging and data-dependent task [16]. Nonetheless, our experiments with the synthetic datasets indicate that iCVI-TopoARTMAP coupled with iCH, iWB, or iPBM yielded superior performance for a wider range of vigilance values when compared with the other ART models. This combination seemed particularly robust in the case that the data range is unknown, as depicted in Fig. 5 (see the supplementary material for additional graphs).

4) Order Dependence and Order Indifference: Order dependence is widely known to be important in incremental learning (see [25], [59] and the references cited within), and thus, for many online agglomerative clustering algorithms, the order in which samples are presented can lead to substantial performance variations. Naturally, ordering affects these algorithms differently, but for many of them, moving toward an orderly presentation typically tends to facilitate learning. For instance, in scenarios in which the data ranges are known a priori, fuzzy ART, DVFA, and topoFA tend to perform better when samples are sorted using visual assessment of cluster tendency [71] and worse for random presentation [45], [59], [66], [72].

The results obtained in this work suggest that regardless of the a priori knowledge of data ranges, skm tends to perform poorly for class-incremental and better for random presentation (this is in agreement with [6], [16]), whereas the opposite was observed for iskm. In turn, the performances of DVFA and topoFA were better for class-incremental presentation in the case of the known data range (thus corroborating the findings in [45], [59], [66], [72]); however, no conclusive relationship was observed for the case of unknown data range. DRN performed better for class-incremental presentation compared with random for both known and unknown data ranges. Finally, iCVI-TopoARTMAP using iCH was

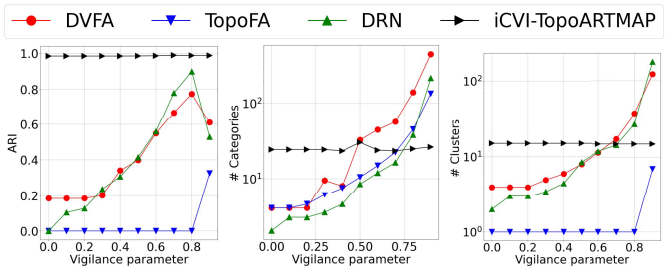


Fig. 5. Performances, number of categories, and number of clusters of the ART-based models as a function of their main vigilance parameter for the S1 dataset. Samples are presented in a class-incremental order and the data range is unknown. The iCVI-TopoARTMAP model uses iCH.

TABLE V

EXPECTED ABSOLUTE PERFORMANCE DIFFERENCE BETWEEN RANDOM AND CLASS-INCREMENTAL PRESENTATIONS. THE SMALLEST DIFFERENCES ARE REPORTED IN BOLD

data range	skm	iskm	DRN	DVFA	TopoFA	iCVI-TopoARTMAP
Unknown	0.3505	0.7244	0.1225	0.1020	0.1563	**0.0106**
Known	0.3544	0.6243	0.0856	0.1151	0.3298	**0.0173**

minimally affected by ordering effects for the majority of experiments with synthetic and image datasets: a high degree of order indifference was observed for this model, as it yielded consistent performance (either superior or comparable to other methods) for both random and class-incremental orderings irrespective of the data range knowledge. In fact, Table V shows that the expected absolute performance difference between these two orders of presentation across the aforementioned datasets (i.e., $\mathbb{E}[|p_{r,d} - p_{\text{ci},d}|]$, where $p_{r,d}$ and $p_{\text{ci},d}$ are the ARI values for random and class-incremental presentations of dataset d, respectively) is smaller for the ART models compared with the non-ART models, wherein iskm and iCVI-TopoARTMAP using iCH yielded the largest and smallest differences, respectively.

IV. CONCLUSION

This article presented iCVI-TopoARTMAP, a novel model for online unsupervised and semi-supervised learning. It is equipped with several features, such as iCVI-based match tracking, online label generation driven by a user-selected iCVI, storage of additional local footprints (e.g., frequency, mean, compactness, and connectivity matrix), online normalization and complement coding of samples, online weight vector rescaling, as well as heuristics for swapping categories between clusters, merging and splitting clusters, reducing the number of categories, and pruning and reassigning satellite categories. In addition, because of the map field component of the ARTMAP architecture, iCVI-TopoARTMAP is capable of incremental multiprototype-based representation of clusters.

In most experiments, iCVI-TopoARTMAP in the unsupervised mode yielded performance superior or comparable to the state-of-the-art iCVI-based and ART-based clustering algorithms, even after augmenting some of the latter to enable their application in use cases with unknown data ranges. Moreover, while the performances of the competitor algorithms were affected in different levels when samples were presented randomly or in a class-incremental fashion (a dramatic decrease in

performance was observed for some of these algorithms when faced with particular ordering instances), iCVI-TopoARTMAP remained very resilient. In addition, iCVI-TopoARTMAP in a semi-supervised mode often yielded a superior or comparable test performance to the prediction of a 1-NN classifier.

REFERENCES

[1] R. Xu and D. C. Wunsch II, *Clustering*. Hoboken, NJ, USA: Wiley, 2009.

[2] G. W. Milligan and M. C. Cooper, "An examination of procedures for determining the number of clusters in a data set," *Psychometrika*, vol. 50, no. 2, pp. 159–179, 1985.

[3] L. Vendramin, R. J. G. B. Campello, and E. R. Hruschka, "Relative clustering validity criteria: A comparative overview," *Stat. Anal. Data Mining Data Sci. J.*, vol. 3, no. 4, pp. 209–235, 2010.

[4] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognit.*, vol. 46, no. 1, pp. 243–256, 2013.

[5] J. Hämäläinen, S. Jauhiainen, and T. Kärkkäinen, "Comparison of internal clustering validation indices for prototype-based clustering," *Algorithms*, vol. 10, no. 3, p. 105, Sep. 2017.

[6] M. Moshtaghi, J. C. Bezdek, S. M. Erfani, C. Leckie, and J. Bailey, "Online cluster validity indices for performance monitoring of streaming data clustering," *Int. J. Intell. Syst.*, vol. 34, no. 4, pp. 541–563, Apr. 2019.

[7] J. C. Bezdek and J. M. Keller, "Streaming data analysis: Clustering or classification?" *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 1, pp. 91–102, Jan. 2021.

[8] W. Wu, J. M. Keller, J. Dale, and J. C. Bezdek, "StreamSoNG: A soft streaming classification approach," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 3, pp. 700–709, Jun. 2022.

[9] O. A. Ibrahim, J. M. Keller, and J. C. Bezdek, "Evaluating evolving structure in streaming data with modified Dunn's indices," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 5, no. 2, pp. 262–273, Apr. 2021.

[10] O. A. Ibrahim, J. M. Keller, and M. Popescu, "A new incremental cluster validity index for streaming clustering analysis," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jun. 2019, pp. 1–8.

[11] L. E. Brito da Silva, N. M. Melton, and D. C. Wunsch, "Incremental cluster validity indices for online learning of hard partitions: Extensions and comparative study," *IEEE Access*, vol. 8, pp. 22025–22047, 2020.

[12] R. Xu, J. Xu, and D. C. Wunsch, "A comparison study of validity indices on swarm-intelligence-based clustering," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 4, pp. 1243–1256, Aug. 2012.

[13] L. E. Brito da Silva and D. C. Wunsch, "Validity index-based vigilance test in adaptive resonance theory neural networks," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–8.

[14] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Netw.*, vol. 4, no. 6, pp. 759–771, 1991.

[15] C. Smith and D. Wunsch, "Particle swarm optimization in an adaptive resonance framework," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2015, pp. 1–4.

[16] M. Chenaghlou, "Data stream clustering and anomaly detection," Ph.D. dissertation, School Comput. Inf. Syst., Univ. Melbourne, Melbourne, VIC, Australia, 2019.

[17] O. A. Ibrahim, J. M. Keller, and J. C. Bezdek, "Analysis of streaming clustering using an incremental validity index," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2018, pp. 1–8.

[18] E. Lughofer, "Extensions of vector quantization for incremental clustering," *Pattern Recognit.*, vol. 41, no. 3, pp. 995–1011, Mar. 2008.

[19] M.-S. Yang and K.-L. Wu, "A new validity index for fuzzy clustering," in *Proc. 10th IEEE Int. Conf. Fuzzy Syst.*, vol. 1, Dec. 2001, pp. 89–92.

[20] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network," *Neural Netw.*, vol. 4, no. 5, pp. 565–588, 1991.

[21] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Trans. Neural Netw.*, vol. 3, no. 5, pp. 698–713, Sep. 1992.

[22] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "A fuzzy ARTMAP nonparametric probability estimator for nonstationary pattern recognition problems," *IEEE Trans. Neural Netw.*, vol. 6, no. 6, pp. 1330–1336, Nov. 1995.

[23] T. Kasuba, "Simplified fuzzy ARTMAP," *AI Expert*, vol. 8, no. 11, pp. 18–25, Nov. 1993.

[24] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Comput. Vis. Graph. Image Process.*, vol. 37, no. 1, pp. 54–115, 1987.

[25] L. E. Brito da Silva, I. Elnabarawy, and D. C. Wunsch, "A survey of adaptive resonance theory neural network models for engineering applications," *Neural Netw.*, vol. 120, pp. 167–203, Dec. 2019.

[26] R. Xu and D. C. Wunsch, II, "BARTMAP: A viable structure for biclustering," *Neural Netw.*, vol. 24, no. 7, pp. 709–716, Sep. 2011.

[27] S. Kim, "Novel approaches to clustering, biclustering algorithms based on adaptive resonance theory and intelligent control," Ph.D. dissertation, Dept. Elect. Comput. Eng., Missouri Univ. Sci. Technol., Rolla, MO, USA, 2016.

[28] R. Yelugam, L. E. Brito da Silva, and D. C. Wunsch, "TopoBARTMAP: Biclustering ARTMAP with or without topological methods in a blood cancer case study," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[29] G. Bartfai, "Hierarchical clustering with ART neural networks," in *Proc. IEEE Int. Conf. Neural Netw. (ICNN)*, vol. 2, Jun. 1994, pp. 940–944.

[30] L. E. Brito da Silva, N. Rayapati, and D. C. Wunsch, "iCVI-ARTMAP: Using incremental cluster validity indices and adaptive resonance theory reset mechanism to accelerate validation and achieve multiprototype unsupervised representations," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 30, 2022, doi: 10.1109/TNNLS.2022.3160381.

[31] J. Seiffertt and D. C. Wunsch II, *Unified Computational Intelligence for Complex Systems* (Evolutionary Learning and Optimization), vol. 6. Berlin, Germany: Springer, 2010.

[32] M. Tscherepanow, *A Topology Learning Hierarchical ART Network*. Berlin, Germany: Springer, 2010, pp. 157–167.

[33] J. A. Swope, "ARTdECOS, adaptive evolving connectionist model and application to heart rate variability," *Evolving Syst.*, vol. 3, no. 2, pp. 95–109, Jun. 2012.

[34] L. Meng, A.-H. Tan, and D. C. Wunsch II, *Adaptive Resonance Theory in Social Media Data Clustering: Roles, Methodologies, and Applications*. Cham, Switzerland: Springer, 2019.

[35] K. Taşdemir and E. Merényi, "Exploiting data topology in visualization and clustering of self-organizing maps," *IEEE Trans. Neural Netw.*, vol. 20, no. 4, pp. 549–562, Apr. 2009.

[36] G.-M. Park, J.-W. Choi, and J.-H. Kim, "Developmental resonance network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1278–1284, Apr. 2019.

[37] M. Carnein and H. Trautmann, "Optimizing data stream representation: An extensive survey on stream clustering algorithms," *Bus. Inf. Syst. Eng.*, vol. 61, no. 3, pp. 277–297, Jun. 2019.

[38] C. P. Lim and R. F. Harrison, *ART-Based Autonomous Learning Systems: Part I Architectures and Algorithms*. Berlin, Germany: Physica-Verlag, 2000, pp. 133–166.

[39] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Hoboken, NJ, USA: Wiley, 2000.

[40] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3645–3649.

[41] M. Wang and W. Deng, "Deep face recognition: A survey," 2018, *arXiv:1804.06655*.

[42] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 67–74.

[43] G. A. Carpenter and N. Markuzon, "ARTMAP-IC and medical diagnosis: Instance counting and inconsistent cases," *Neural Netw.*, vol. 11, no. 2, pp. 323–336, Mar. 1998.

[44] I. Škrjanc, J. A. Iglesias, A. Sanchis, D. Leite, E. Lughofer, and F. Gomide, "Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: A survey," *Inf. Sci.*, vol. 490, pp. 344–368, Jul. 2019.

[45] L. E. Brito da Silva, I. Elnabarawy, and D. C. Wunsch, "Distributed dual vigilance fuzzy adaptive resonance theory learns online, retrieves arbitrarily-shaped clusters, and mitigates order dependence," *Neural Netw.*, vol. 121, pp. 208–228, Jan. 2020.

[46] D. de Araújo, A. D. Neto, J. Melo, and A. Martins, "Clustering using elements of information theory," in *Artificial Neural Networks ICANN*, K. Diamantaras, W. Duch, and L. S. Iliadis, Eds. Berlin, Germany: Springer, 2010, pp. 397–406.

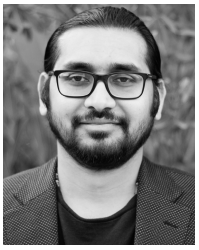
[47] J. A. Silva, E. R. Faria, R. C. Barros, E. R. Hruschka, A. C. P. L. F. D. Carvalho, and J. Gama, "Data stream clustering: A survey," *ACM Comput. Surv.*, vol. 46, no. 1, pp. 1–31, Jul. 2013.

[48] A. Zubaroğlu and V. Atalay, "Data stream clustering: A review," *Artif. Intell. Rev.*, vol. 54, no. 2, pp. 1201–1236, Feb. 2021.
[49] C. G. Bezerra, B. S. J. Costa, L. A. Guedes, and P. P. Angelov, "An evolving approach to data streams clustering based on typicality and eccentricity data analytics," *Inf. Sci.*, vol. 518, pp. 13–28, May 2020.
[50] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image Vis. Comput.*, vol. 28, no. 6, pp. 902–913, 2010.
[51] J. Fonollosa, S. Sheik, R. Huerta, and S. Marco, "Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring," *Sens. Actuators B, Chem.*, vol. 215, pp. 618–629, Aug. 2015.
[52] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models," *Energy Buildings*, vol. 112, pp. 28–39, Jan. 2016.
[53] G. Casalino, G. Castellano, and C. Mencar, "Data stream classification by dynamic incremental semi-supervised fuzzy clustering," *Int. J. Artif. Intell. Tools*, vol. 28, no. 8, Dec. 2019, Art. no. 1960009.
[54] P. Fränti and S. Sieranoja, "K-means properties on six clustering benchmark datasets," *Appl. Intell.*, vol. 48, no. 12, pp. 4743–4759, 2018.
[55] I. Kärkkäinen and P. Fränti, "Dynamic local search algorithm for the clustering problem," Dept. Comput. Sci., Univ. Joensuu, Joensuu, Finland, Tech. Rep., A-2002-6, 2002.
[56] P. Fränti and O. Virmajoki, "Iterative shrinking method for clustering problems," *Pattern Recognit.*, vol. 39, no. 5, pp. 761–765, May 2006.
[57] M. Rezaei and P. Franti, "Set matching measures for external cluster validity," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 8, pp. 2173–2186, Aug. 2016.
[58] D. Dua and C. Graff. (2017). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml
[59] L. E. Brito da Silva and D. C. Wunsch, "A study on exploiting VAT to mitigate ordering effects in fuzzy ART," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 2351–2358.
[60] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "ICaRL: Incremental classifier and representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2001–2010.
[61] A. P. Brna, R. C. Brown, P. M. Connolly, S. B. Simons, R. E. Shimizu, and M. Aguilar-Simon, "Uncertainty-based modulation for lifelong learning," *Neural Netw.*, vol. 120, pp. 129–142, Dec. 2019.
[62] B. Prihasto et al., "A survey of deep face recognition in the wild," in *Proc. Int. Conf. Orange Technol. (ICOT)*, Dec. 2016, pp. 76–79.
[63] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.
[64] L. Hubert and P. Arabie, "Comparing partitions," *J. Classification*, vol. 2, no. 1, pp. 193–218, Jan. 1985.
[65] J. Macqueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1, no. 233, Nov. 1967, pp. 281–297.
[66] L. E. Brito da Silva, I. Elnabarawy, and D. C. Wunsch, "Dual vigilance fuzzy adaptive resonance theory," *Neural Netw.*, vol. 109, pp. 1–5, Jan. 2019.
[67] M. Tscherepanow, "Incremental on-line clustering with a topology-learning hierarchical ART neural network using hyperspherical categories," in *Proc. Ind. Conf. Data Mining (ICDM)*, P. Perner, Ed., 2012, pp. 22–34.
[68] I.-U. Yoon, U.-H. Kim, H. Myung, and J.-H. Kim, "S-DRN: Stabilized developmental resonance network," in *Robot Intelligence Technology and Applications*, J. Kim, B. Englot, H.-W. Park, H.-L. Choi, H. Myung, J. Kim, and J.-H. Kim, Eds. Cham, Switzerland: Springer, 2022, pp. 431–442.
[69] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2012.
[70] B. Calvo and G. Santafé, "scmamp: Statistical comparison of multiple algorithms in multiple problems," *R J.*, vol. 8, no. 1, pp. 248–256, 2016.
[71] J. C. Bezdek and R. J. Hathaway, "VAT: A tool for visual assessment of (cluster) tendency," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, vol. 3, May 2002, pp. 2225–2230.
[72] I. Elnabarawy, D. C. Wunsch, and L. E. Brito da Silva, "Dual vigilance hypersphere adaptive resonance theory," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2019, pp. 2425–2532.

Leonardo Enzo Brito da Silva (Member, IEEE) received the B.S. degree in electrical engineering and the M.S. degree in electrical and computer engineering from the Universidade Federal do Rio Grande do Norte, Natal, Brazil, in 2011 and 2013, respectively, and the Ph.D. degree in computer engineering from the Missouri University of Science and Technology (Missouri S&T), Rolla, MO, USA, in 2019.

He is currently a Machine Learning Researcher with Guise AI Inc., Rolla. His research interests include unsupervised learning, incremental learning, neural networks, information theory, evolutionary computation, and computer vision.

Dr. Brito da Silva received the 2019 Missouri S&T Dean's Ph.D. Scholar Award and the 2021 International Neural Networks Society Doctoral Dissertation Award.



Nagasharath Rayapati received the bachelor's degree in electrical and electronics engineering from Acharya Nagarjuna University, Guntur, India, in 2008.

He worked as a Data Architect with American Express, Phoenix, AZ, USA, building data lake; and a Data Engineer with Cerner, Kansas, MO, USA, helping build revenue cycle applications associated with patient-centric data. He is currently working as the Chief Executive Officer with Guise AI Inc., Rolla, MO, USA. His research interests include unsupervised learning, incremental learning, neural networks, evolutionary algorithms, and computer vision.

Mr. Rayapati serves on the Industrial Advisory Board of the International Neural Networks Society and as a member for the Forbes Technology Council, which publishes his articles related to business and policy implications of artificial intelligence/machine learning.

Donald C. Wunsch, II (Fellow, IEEE) received the B.S. degree in applied mathematics from the University of New Mexico, Albuquerque, NM, USA, in 1984, the Jesuit Honors Program, Seattle University, Seattle, WA, USA, the M.S. degree in applied mathematics and the Ph.D. degree in electrical engineering from the University of Washington, Seattle, in 1987 and 1991, respectively, and the M.B.A. degree from Washington University, St. Louis, MO, USA, in 2006.

He is the Mary K. Finley Missouri Distinguished Professor with the Missouri University of Science and Technology (Missouri S&T), Rolla, MO, USA, and the Director of the Kummer Institute Center for Artificial Intelligence and Autonomous Systems and the Applied Computational Intelligence Laboratory, Rolla. Earlier employers were: Texas Tech, Lubbock, TX, USA; Boeing, Seattle; Rockwell International, Albuquerque, NM, USA; and International Laser Systems, Albuquerque. His research interests include real-time learning, unsupervised learning, and reinforcement learning.

Dr. Wunsch produced 23 Ph.D. recipients in Computer Engineering, Electrical Engineering, Systems Engineering, and Computer Science. From 2021 to 2022, he served as a Program Director at the National Science Foundation. He served as the International Neural Networks Society (INNS) President, the IJCNN General Chair, and on the St. Patrick's School Board, IEEE Neural Networks Council, INNS Board, and the University of Missouri Bioinformatics Consortium, Chaired the Missouri S&T Information Technology and Computing Committee, and the Student Design and Experiential Learning Center Board. His awards include INNS Fellow, NSF CAREER, INNS Gabor, INNS Ada Lovelace, and IEEE Neural Networks Pioneer.