From Neural Collapse to Label-Limited Evolving Streams: Geometry-Constrained Learning under Dynamic Class Imbalance

Hongliang Wang[1], **Hongyuan Liu**[1,2], **Qirui Hao**[1], **Yun Sing Koh**[3], **Qinli Yang**[1] and **Junming Shao**[1*]
[1]University of Electronic Science and Technology of China
[2]University of Bristol
[3]University of Auckland
{honglwang, hongyuanliu, qiruihao}@std.uestc.edu.cn, y.koh@auckland.ac.nz,
{qinli.yang, junmshao}@uestc.edu.cn

Abstract

Learning from label-limited streams presents significant challenges, particularly when coupled with concept drift and dynamic class imbalance. Existing works often struggle to maintain a discriminative feature space under these constraints, biasing decision boundaries toward majority classes or outdated concepts. To address this, we propose a novel framework named **N**eural collapse **I**nspired **L**abel-limited **E**volving stream learning (NILE). Instead of utilizing learnable classifiers, NILE exploits Neural Collapse (NC) geometry to explicitly construct a Simplex Equiangular Tight Frame (ETF) as a fixed classifier, ensuring maximal inter-class separability to guide feature discriminability. To maintain this discrimination with limited labels, NILE employs hybrid active learning to prioritize uncertain and minority samples, and an NC-adapted semi-supervised mechanism to enhance representation learning. NILE further updates the classifier by dynamically adjusting the ETF structure, enabling continual adaptation to drifting concepts. Extensive experiments show that NILE can effectively guide features toward the NC state, significantly outperforming state-of-the-art baselines in nonstationary environments.

1 Introduction

Data streams are ubiquitous in real-world applications, ranging from network intrusion detection and financial transaction monitoring to real-time sensor analysis [Zhao *et al.*, 2022]. A defining characteristic of evolving data streams is concept drift, where the underlying statistical distribution changes over time, requiring models to continuously adapt to new concepts while making immediate decisions [Gunasekara *et al.*, 2024; Zhao and Koh, 2020]. Simultaneously, the high-speed throughput and annotation costs of data typically render streams label-limited [Fahy *et al.*, 2022], often accompanied by evolving class ratios that manifest as dynamic class imbalance [Aguiar *et al.*, 2024; Zhou *et al.*, 2023]. Within this complex scenario, models must adapt to concept changes with limited labels while avoiding decision boundaries dominated by majority classes, thereby significantly increasing the difficulty of stream learning [Liu *et al.*, 2021].

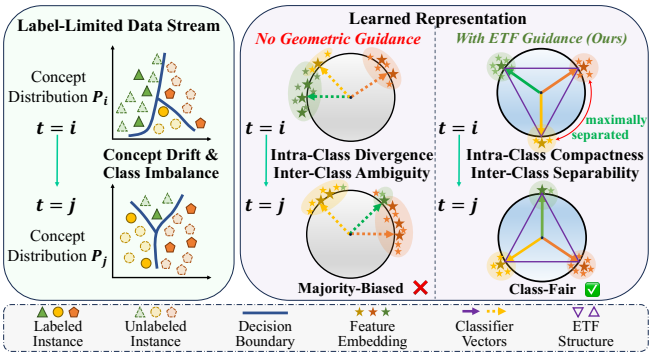


Figure 1: The challenges in representation learning for label-limited evolving streams with concept drift and dynamic class imbalance.

Recent efforts to simultaneously address these challenges primarily fall into active learning (AL) and semi-supervised learning (SSL). AL aims to select the most informative samples for annotation to alleviate imbalance and track drift [Zhang *et al.*, 2022; Li *et al.*, 2024]; however, it only utilizes the queried labeled data to update the model, resulting in a waste of unlabeled distributional information. In contrast, SSL attempts to mine knowledge from unlabeled data to improve adaptability, but it cannot actively query the informative instances, making performance sensitive to the labeled data quality [Abdi *et al.*, 2024; Vafaie *et al.*, 2020]. Given their complementarity, semi-supervised active learning (SSAL) is regarded as a promising solution [Jiao *et al.*, 2025]. However, existing SSAL methods on data streams are restricted to binary classification scenarios [Guo *et al.*, 2024]. To our knowledge, there is currently no effective SSAL algorithm capable of handling limited labels, concept drift, and dynamic multi-class imbalance simultaneously.

Meanwhile, traditional label-limited stream learning methods capable of handling class imbalance are often built on shallow models and lack powerful representation learning.

*Corresponding author.

Recent study suggests that neural networks can be valuable for label-limited stream learning [Wang *et al.*, 2026c], but maintaining discriminative embeddings under non-stationary and imbalance is challenging. As illustrated in Fig. 1, without appropriate geometric guidance, learned features tend to degenerate into intra-class divergence and inter-class ambiguity, leading to suboptimal, majority-biased solutions.

Recent discovery of Neural Collapse (NC) provides key insights into these problems [Papyan *et al.*, 2020]. It indicates that networks trained on balanced data converge toward an optimal geometric structure: a Simplex Equiangular Tight Frame (ETF). In an ETF, class prototypes have equal norms and maximal equal pairwise angles, maximizing intra-class compactness and inter-class separability [Yan *et al.*, 2024]. This naturally leads to the hypothesis that guiding feature representations toward such symmetric, class-fair NC structure, while dynamically adjusting it to adapt to drifts, would benefit label-limited streams with dynamic class imbalance.

To this end, we propose a novel SSAL method, named NILE, for inducing Neural Collapse in imbalanced data streams with limited labels. The core idea is employing a neural network with explicitly constructed ETF vectors to classify imbalanced streams, and dynamically adjusting the ETF structure to adapt to concept drift. Specifically, we propose a hybrid active learning strategy adapted to ETF that prioritizes uncertain and minority samples under limited budgets. We further introduce an NC-based semi-supervised learning method utilizing high-confidence unlabeled data. Moreover, a data-driven adjustment mechanism ensures the ETF remains aligned with class prototypes under new concepts. The main contributions of this paper are as follows:

- We are the first to introduce NC into label-limited data stream learning with dynamic class imbalance, proposing a dynamic adjustment method for the ETF classifier to handle concept drift.
- We propose a SSAL method adapted to NC geometry, which focuses on uncertain and minority class samples while effectively utilizing unlabeled data.
- Comprehensive experimental analysis validates the effectiveness of NILE and demonstrates its ability to guide feature representations to converge toward the NC state.

2 Related Work

Label-Limited Data Stream Learning. To mine knowledge from label-limited streams, SSL and AL have garnered significant attention, which are detailed in surveys [Gomes *et al.*, 2022] and [Cacciarelli and Kulahci, 2024], respectively. SSL aims to leverage both a small amount of labeled data and a large volume of unlabeled data for model training [Jiang *et al.*, 2025; Wu *et al.*, 2026]. Existing methodologies include self-training [Tanha *et al.*, 2022; Wang *et al.*, 2026b], co-training [Wang and Li, 2018], clustering-based [Din *et al.*, 2022; Liao *et al.*, 2023], graph-based [Yu *et al.*, 2025], and neural network-based methods [Wang *et al.*, 2026a; Wang *et al.*, 2026c]. However, SSL cannot guarantee that the available labeled data is the most valuable. Conversely, AL optimizes model learning by selecting the most informative samples for annotation using strategies like uncertainty sampling [Žliobaitė *et al.*, 2014], density-based strategies [Malialis *et al.*, 2022], query-by-committee [Krawczyk and Cano, 2019], and hybrid strategies [Liu *et al.*, 2023a]. Nevertheless, AL often neglects the potential information in the unqueried data. To exploit the complementary strengths of both paradigms, [Jiao *et al.*, 2025] integrated smoothness-based querying with semi-supervised regularization. However, it is restricted to binary classification and fails to address class imbalance, compromising performance in scenarios where accurate recognition of minority classes is critical.

Class-Imbalanced Data Stream Learning. Early research primarily targeted binary classification scenarios, typically assuming only one majority class and one minority class [Wang *et al.*, 2015]. To handle multi-class imbalance challenges, numerous Supervised Learning (SL) [Boiko Ferreira *et al.*, 2019; Loezer *et al.*, 2020] and AL algorithms [Li *et al.*, 2024] have been proposed. Within SL, MUOB and MOOB [Wang *et al.*, 2016] utilize online bagging-based resampling, while KUE [Cano and Krawczyk, 2020] and ROSE [Cano and Krawczyk, 2022] enhance robustness through dynamic ensemble weighting and self-adjusting bagging mechanisms to handle imbalance and drift. Regarding AL, CALMID [Liu *et al.*, 2021] targets minority classes via asymmetric margin thresholds, whereas LEPID [Wu *et al.*, 2024] increases query probabilities for instances near minority micro-clusters. Apart from SL and AL, only a few studies addressed multi-class imbalance within the SSL scenario [Abdi *et al.*, 2024; Vafaie *et al.*, 2020]. For instance, IOE [Vafaie *et al.*, 2020] combines prediction probabilities or nearest neighbors to assign pseudo-labels, thereby performing self-training learning. A broader review for class-imbalanced data stream algorithms is found in survey [Aguiar *et al.*, 2024]. Despite these advances, to our knowledge, there is currently no effective SSAL algorithm simultaneously addresses concept drift, label scarcity, and dynamic multi-class imbalance.

Neural Collapse. The NC phenomenon reveals a specific geometric structure exhibited by deep nets during the terminal phase of training: last-layer features converge to class prototypes, and the prototypes align with classifier weights to form a Simplex ETF [Papyan *et al.*, 2020]. Theoretical studies indicate that NC represents the global optimal feature representation under losses such as Cross-Entropy [Zhu *et al.*, 2021; Hong and Ling, 2024] and MSE [Zhou *et al.*, 2022], yielding maximal inter-class separation and minimal intra-class variation. This geometry has been exploited for out-of-distribution detection [Liu and Qin, 2025], knowledge distillation [Zhang *et al.*, 2025], SSL [Hu *et al.*, 2024; Xiao *et al.*, 2024], and continual learning [Seo *et al.*, 2024]. For class imbalance, prior work addresses minority collapse by fixing an ETF classifier to regularize the feature space and improve tail-class generalization [Yang *et al.*, 2022]. However, the behavior of NC in evolving data streams and its drift-adaptive mechanisms have not yet been effectively investigated.

3 Preliminaries

3.1 Problem Setup

We define a label-limited evolving data stream as $S = \{(x_t, y_t)\}_{t=1}^{\infty}$, where the instance $x_t \in \mathbb{R}^n$ contains n-

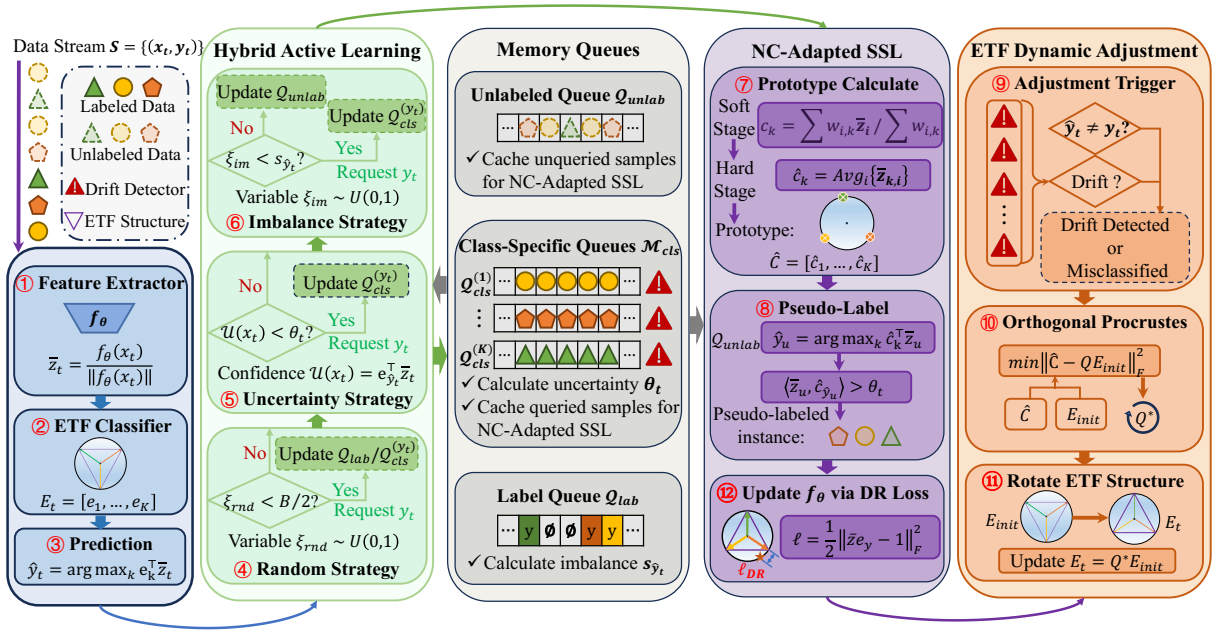


Figure 2: The workflow of the stream learning phase. For each incoming instance, predictions are obtained using the ETF classifier following feature extraction. The Hybrid Active Learning integrates random, uncertainty, and class imbalance-aware sampling to selectively query labels. Corresponding queues are updated based on the query results while the detector is employed within each Class-Specific Queue to identify concept drift. NC-adapted SSL is proposed to leverage labeled and unlabeled data by calculating soft weights for unlabeled samples using the ETF and performing hard assignments to refine class prototypes. Pseudo-labels for unlabeled samples are determined by the nearest class prototype and filtered via the uncertainty threshold followed by network updates using the Dot-Regression Loss. Notably, the process checks whether to trigger the ETF Dynamic Adjustment module before the network update where the ETF structure is realigned according to the class prototypes if drift or misclassification occurs.

dimensional features, and the label $y_t \in \{1, \ldots, K\}$ belongs to one of K classes. Ground truth labels are accessible only for the initial N samples and subsequent actively queried instances that satisfy the labeling budget $B \in [0, 1]$; the remaining instances are denoted as $(x_t, \emptyset)$. The data stream exhibits concept drift, where $P_t(x, y) \neq P_{t+1}(x, y)$, and dynamic class imbalance. Specifically, the class prior distribution is skewed and evolves over time: $\min_k P_t(y = k) \ll \max_k P_t(y = k)$ and $P_t(y) \neq P_{t+1}(y)$. The objective of this paper is to update the model online at time t using limited supervision, aiming to accurately predict the label of x_{t+1} and adapt to distribution evolution.

3.2 Properties of Neural Collapse

Definition 1 (Simplex Equiangular Tight Frame). Given a set of vectors $\mathbf{E} = \{\mathbf{e}_k\}_{k=1}^K \in \mathbb{R}^d$, where $K \leq d + 1$, it is termed a Simplex ETF if its matrix form satisfies the following condition:

$$\mathbf{E} = \sqrt{\frac{K}{K-1}}\mathbf{U}\left(\mathbf{I}_K - \frac{1}{K}\mathbf{1}_K\mathbf{1}_K^\top\right), \tag{1}$$

where $\mathbf{U} \in \mathbb{R}^{d\times K}$ satisfies $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_K$, and $\mathbf{1}_K$ is an all-ones vector. This structure ensures that the dot product between any distinct pair of vectors is a fixed value $-\frac{1}{K-1}$, thereby achieving maximal class separability. Accordingly, the NC phenomenon is characterized by the following key properties:

(NC1) Variability Collapse: Feature representations $\mathbf{z}_{k,i}$ from the same class k progressively converge to their class prototype $\boldsymbol{\mu}_k = \text{Avg}_i\{\mathbf{z}_{k,i}\}$, causing the within-class covariance matrix Σ_W to approach zero:

$$\Sigma_W = \underset{k,i}{\text{Avg}}\left\{(\mathbf{z}_{k,i} - \boldsymbol{\mu}_k)(\mathbf{z}_{k,i} - \boldsymbol{\mu}_k)^\top\right\} \to \mathbf{0}. \tag{2}$$

(NC2) Convergence to Simplex ETF: The normalized class prototypes $\bar{\boldsymbol{\mu}}_k$, centered by the global feature mean $\boldsymbol{\mu}_G = \text{Avg}_k\{\boldsymbol{\mu}_k\}$, converge to the vertices of a Simplex ETF as described in Definition 1:

$$\forall k \neq j, \langle\bar{\boldsymbol{\mu}}_k, \bar{\boldsymbol{\mu}}_j\rangle \to -\frac{1}{K-1}, \text{where } \bar{\boldsymbol{\mu}}_k = \frac{\boldsymbol{\mu}_k - \boldsymbol{\mu}_G}{\|\boldsymbol{\mu}_k - \boldsymbol{\mu}_G\|}. \tag{3}$$

(NC3) Self-Duality: The weight vectors of the linear classifier $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$ align with the centered class prototypes, collapsing into the identical simplex structure:

$$\frac{\mathbf{w}_k}{\|\mathbf{w}_k\|} \approx \bar{\boldsymbol{\mu}}_k, \forall k \in \{1, \ldots, K\}. \tag{4}$$

4 Methodology

The core of NILE involves leveraging the geometric properties of NC to maintain a discriminative and fair feature space, while subsequently conducting active learning and semi-supervised learning based on NC principles. The proposed method first utilizes initial labeled data to initialize the feature extractor, the ETF classifier, and three types of queues: a

Label Queue, K Class-Specific Labeled Queues, and an Unlabeled Queue. During the subsequent stream learning phase, we sequentially execute Hybrid Active Learning, NC-adapted semi-supervised learning, and ETF Dynamic Adjustment, as detailed in Fig. 2. Pseudo-code are shown in Supplementary Material[1].

4.1 Initialization

Before data stream learning commences, we utilize an initial labeled dataset $\mathcal{D}_{init}$ to pre-train the feature extractor $f_\theta : \mathcal{X} \to \mathbb{R}^d$ and construct a fixed Simplex ETF classifier $\mathbf{E}_{init} = [\mathbf{e}_1, \ldots, \mathbf{e}_K]$ to explicitly induce Neural Collapse. To guide the alignment of features with the ETF structure, we employ Dot-Regression Loss for optimization [Yang *et al.*, 2022]. Unlike loss functions containing repulsive terms, this objective function leverages the optimality of the fixed ETF and retains only the gradient assumption that pulls features toward their class prototypes, thereby achieving feature collapse more precisely. It is defined as follows:

$$\mathcal{L} = \frac{1}{2|\mathcal{D}_{init}|} \sum_{(x_i, y_i) \in \mathcal{D}_{init}} \left(\mathbf{e}_{y_i}^\top \bar{\mathbf{z}}_i - 1\right)^2, \tag{5}$$

where $\bar{\mathbf{z}}_i = \frac{f_\theta(x_i)}{\|f_\theta(x_i)\|}$. In this metric space, the predicted class $\hat{y}$ of a sample is determined by the maximum cosine similarity between the normalized feature and the ETF classifier: $\hat{y} = \arg\max_k \mathbf{e}_k^\top \bar{\mathbf{z}}$. Upon completion of pre-training, three key queues are initialized based on a sliding window mechanism to adapt to the streaming environment: 1) **Label Queue** $\mathcal{Q}_{lab} = \{q_i\}_{i=1}^{S_L}$, where $q_i \in \{1, \ldots, K\} \cup \{\emptyset\}$. This queue stores label values of historical data (unlabeled data is represented by a placeholder $\emptyset$) and is used to estimate the class prior distribution $P_t(y)$ in real-time to quantify the degree of class imbalance. 2) **Class-Specific Labeled Queues** $\mathcal{M}_{cls} = \{\mathcal{Q}_{cls}^{(k)}\}_{k=1}^{K}$. For the k-th class, $\mathcal{Q}_{cls}^{(k)} = \{x_j \mid y_j = k\}$ stores the features of samples belonging to the same class with a capacity of $\lceil S_L/K \rceil$. An independent concept drift detector $ADWIN^{(k)}$ is maintained for each class to identify concept drift by monitoring the classification accuracy indicator $\mathbb{I}(\hat{y} = y)$ of samples in that class. 3) **Unlabeled Queue** $\mathcal{Q}_{unlab} = \{x_m\}_{m=1}^{S_{ul}}$, with a capacity set to $S_{ul} = S_L \cdot \lceil (1-B)/B \rceil$. This queue is used to cache features of unqueried samples, aiming to provide data support for subsequent semi-supervised learning.

4.2 Hybrid Active Learning

Based on the Simplex ETF geometric space and memory queues constructed in the initialization phase, we propose a Hybrid Active Learning strategy. This strategy aims to efficiently select the most valuable samples for model updates under a limited labeling budget B. To balance the unbiased estimation of data distribution, the refinement of decision boundaries, and the coverage of long-tail classes, we sequentially execute random sampling, uncertainty sampling, and imbalance-aware sampling.

[1]https://github.com/WangHLzZ/NILE

First, to ensure basic coverage of the current data stream distribution, we perform random sampling. For each newly arriving data stream instance x_t, a random variable $\xi_{rnd} \sim U(0,1)$ is generated. If $\xi_{rnd} < B/2$, its label y_t is directly queried, and y_t is pushed into $\mathcal{Q}_{lab}$ and $\mathcal{Q}_{cls}^{(y_t)}$; otherwise, it enters the uncertainty sampling stage.

Uncertainty sampling aims to capture hard-to-classify samples located near the decision boundary, which is typically a sensitive region for the occurrence of concept drift and majority class dominance. Based on the property of NC, we define the confidence score of a sample as the cosine similarity between its normalized feature $\bar{\mathbf{z}}_t$ and the predicted class ETF prototype $\mathbf{e}_{\hat{y}_t}$: $u(x_t) = \mathbf{e}_{\hat{y}_t}^\top \bar{\mathbf{z}}_t$. A lower confidence score implies that the sample deviates from the class prototype, indicating higher uncertainty. Given the evolving nature of data streams, the threshold for uncertainty sampling needs to be dynamically adjusted. Therefore, we adopt a data-driven method to dynamically calculate the threshold based on the prediction performance in the labeled data queue $\mathcal{M}_{cls}$. Specifically, samples in the queue are divided into a correctly predicted set $\mathcal{U}_{corr}$ and an incorrectly predicted set $\mathcal{U}_{inc}$. We define the Empirical Cumulative Distribution Function (ECDF) as $\widehat{F}(u) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(u_i \le u)$ and its α-quantile function as $\widehat{F}^{-1}(\alpha) = \inf\{u \in \mathbb{R} : \widehat{F}(u) \ge \alpha\}$. The dynamic uncertainty sampling threshold θ_t is set as:

$$\theta_t = \max\left(\widehat{F}_{\mathcal{U}_{corr}}^{-1}(0.05), \widehat{F}_{\mathcal{U}_{inc}}^{-1}(0.95)\right). \tag{6}$$

If $u(x_t) < \theta_t$, the label y_t is queried, and x_t is pushed into $\mathcal{Q}_{cls}^{(y_t)}$; otherwise, it enters imbalance-aware sampling.

Finally, to alleviate the minority collapse problem in imbalanced streams, we introduce an imbalance-aware sampling strategy. This strategy calculates the prior ratio p_k of each class in real-time based on the statistics of non-placeholder data in the label queue $\mathcal{Q}_{lab}$. To prioritize sampling of minority classes, we define the normalized class frequency $\hat{p}_k = \frac{p_k - p_{min}}{1/K - p_{min}}$, where p_{min} is the current minimum class ratio. For classes with $p_k \le 1/K$, their sampling probability s_k is determined by the following exponential decay function:

$$s_k = 1 - \frac{e^{K\hat{p}_k} - 1}{e^K - 1}. \tag{7}$$

This function ensures that the sampling probability for the most minority class is 1, while the sampling probability for classes approaching the average level tends to 0. For majority classes ($p_k > 1/K$), we directly set $s_k = 0$. Similar to random sampling, for instance x_t, a random variable $\xi_{im} \sim U(0,1)$ is generated. If $\xi_{im} < s_{\hat{y}_t}$, its label y_t is directly queried, and x_t is pushed into $\mathcal{Q}_{cls}^{(y_t)}$; otherwise, the sample is treated as unlabeled data and pushed into $\mathcal{Q}_{unlab}$.

4.3 NC-Adapted Semi-Supervised Learning

Based on the labeled samples selected by hybrid active learning and the unlabeled data stream in $\mathcal{Q}_{unlab}$, we further propose a semi-supervised learning mechanism adapted to NC characteristics to enhance representation learning on label-limited streams. Inspired by the prototype calibration idea in

ETF-SSL [Hu *et al.*, 2024], we design a Soft-Hard Two-Stage class prototype estimation strategy, utilizing the currently maintained data queues to dynamically reconstruct class prototypes for guiding model updates.

In the soft stage, we utilize the ETF classifier as a geometric anchor to calculate the soft assignment weights for unlabeled data. Specifically, for an unlabeled sample $x_u \in \mathcal{Q}_{unlab}$, its weight belonging to the k-th class is determined by the rectified inner product of its normalized feature $\bar{\mathbf{z}}_u$ and the ETF vertex $\mathbf{e}_k$, i.e., $w_{u,k} = \max(0, \mathbf{e}_k^\top \bar{\mathbf{z}}_u)$, where the negative correlation part is truncated to eliminate interference. For a labeled sample $x_l \in \mathcal{Q}_{cls}$, its weight is directly determined by the ground truth label, i.e., $w_{l,k} = \mathbb{I}(y_l = k)$. Based on this, the class prototype $\mathbf{c}_k$ in the soft stage is calculated as the weighted average of all samples:

$$\mathbf{c}_k = \frac{\sum_{x_i \in \mathcal{Q}_{all}} w_{i,k} \bar{\mathbf{z}}_i}{\sum_{x_i \in \mathcal{Q}_{all}} w_{i,k}}, \tag{8}$$

where $\mathcal{Q}_{all} = \mathcal{M}_{cls} \cup \mathcal{Q}_{unlab}$. In the hard stage, we use $\mathbf{c}_k$ to pre-classify unlabeled data to obtain hard pseudo-labels $\hat{y}_u = \arg\max_k \langle \bar{\mathbf{z}}_u, \mathbf{c}_k \rangle$. Subsequently, based on the updated hard label distribution, class prototypes are calculated again to obtain the final refined prototype $\hat{\mathbf{c}}_k$. This process employs iterative correction, enabling class prototypes to more accurately reflect the true distribution structure of the current data stream.

To ensure the robustness of model training, we only select unlabeled samples with high confidence in the final prototype space to participate in gradient updates. Specifically, a sample is accepted only if its similarity with the corresponding prototype exceeds the previously defined uncertainty threshold θ_t (i.e., satisfying $\langle \bar{\mathbf{z}}_u, \hat{\mathbf{c}}_{\hat{y}_u} \rangle > \theta_t$). The final optimization objective is composed of the supervised loss from labeled data and the semi-supervised loss from high-confidence unlabeled data, both employing Dot-Regression Loss to explicitly induce the NC structure:

$$\mathcal{L}_{total} = \frac{1}{N_L} \sum_{x_l \in \mathcal{M}_{cls}} \ell_{DR}(\bar{\mathbf{z}}_l, \mathbf{e}_{y_l}) + \frac{1}{N'_U} \sum_{x_u \in \mathcal{D}'} \ell_{DR}(\bar{\mathbf{z}}_u, \mathbf{e}_{\hat{y}_u}), \tag{9}$$

where $\mathcal{D}'$ is the screened unlabeled dataset, and $\ell_{DR}(\mathbf{z}, \mathbf{e}) = \frac{1}{2}\|\mathbf{z}^\top \mathbf{e} - 1\|^2$. In this way, the model not only leverages unlabeled data to stabilize intra-class compactness in the feature space but also achieves rapid adaptation to evolving distributions through the guidance of dynamic prototypes.

4.4 ETF Classification Dynamic Adjustment

When drift occurs, the geometric structure of the feature space deviates from the NC state, causing a mismatch between the ETF classifier and the current feature distribution. To quickly restore the NC state of the post-drift distribution while maintaining the optimal geometric properties of the ETF classifier, we introduce a dynamic ETF alignment mechanism to explicitly adapt to concept drift. Based on the refined class prototypes $\hat{\mathbf{C}} = [\hat{\mathbf{c}}_1, \ldots, \hat{\mathbf{c}}_K]$ calculated in the previous stage, our goal is to find an optimal rotation matrix $\mathbf{Q} \in \mathbb{R}^{d \times d}$ such that the distance between the transformed ETF structure $\mathbf{Q}\mathbf{E}_{init}$ and the current prototype $\hat{\mathbf{C}}$ is minimized. This objective is equivalent to solving the Orthogonal Procrustes problem:

$$\min_{\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}} \|\hat{\mathbf{C}} - \mathbf{Q}\mathbf{E}_{init}\|_F^2 \iff \max_{\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}} \mathrm{Tr}(\mathbf{Q}^\top \hat{\mathbf{C}} \mathbf{E}_{init}^\top). \tag{10}$$

Let $\mathbf{M} = \hat{\mathbf{C}}\mathbf{E}_{init}^\top$. Performing Singular Value Decomposition (SVD) on it yields $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, where $\mathrm{rank}(\mathbf{M}) = r$. The general solution to this problem includes arbitrary rotational degrees of freedom in the null space, formally expressed as:

$$\mathbf{Q} = \mathbf{U} \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \mathbf{V}^\top, \tag{11}$$

where $\mathbf{R} \in \mathbb{R}^{(d-r)\times(d-r)}$ is an arbitrary orthogonal matrix. To ensure the smoothness of the ETF classifier update, we select the optimal solution $\mathbf{Q}^*$ closest to $\mathbf{E}_{init}$ according to the principle of minimal movement. Finally, whenever $ADWIN$ detects drift or an already queried x_t is misclassified, we dynamically update the classifier to $\mathbf{E}_t = \mathbf{Q}^*\mathbf{E}_{init}$, forcing the feature extractor to collapse towards the corrected optimal geometric center in subsequent iterations.

Dataset	Ins.	At.	Cl.	Imb. Ratio	Drift
SYN-GGG	100k	10	5	2/1/5/4/3→1/5/4/3/2 →5/4/3/2/1→4/3/2/1/5	grad., grad., grad.
SYN-SSS	100k	10	5	5/4/3/2/1→4/3/2/1/5 →3/2/1/5/4→2/1/5/4/3	sud., sud., sud.
SYN-GIGG	100k	10	5	2/1/5/4/3→1/5/4/3/2 →5/4/3/2/1→4/3/2/1/5	grad., inc., grad., grad.
SYN-SISS	100k	10	5	5/4/3/2/1→4/3/2/1/5 →3/2/1/5/4→2/1/5/4/3	sud., inc., sud., sud.
SYN-GSG	100k	20	5	4/3/2/1/5→10/5/5/4/1 →1/2/5/2/5→2/2/5/5/1	grad., sud., grad.
SYN-SGS	100k	20	5	1/2/5/2/5→10/5/5/4/1 →4/3/2/1/5→5/4/1/10/5	sud., grad., sud.
SYN-GISG	100k	20	5	4/3/2/1/5→10/5/5/4/1 →1/2/5/2/5→2/2/5/5/1	grad., inc., sud., grad.
SYN-SIGS	100k	20	5	1/2/5/2/5→10/5/5/4/1 →4/3/2/1/5→5/4/1/10/5	sud., inc., grad., sud.
Spam	9324	499	2	3/1	—
GSAD	13910	128	6	2/2/1/1/2/1	—
Weather	18159	8	2	2/1	—
CovType	581012	54	7	77/103/13/1/3/6/7	—
Asfault	8566	64	5	12/5/25/1/1	—
HAR	10299	561	6	1/1/1/1/1/1	—

Table 1: Characteristic of synthetic and real-world datasets.

5 Experiments

5.1 Settings

Datasets. To evaluate the algorithm performance comprehensively, we selected 8 synthetic data streams and 6 widely used real-world benchmarks (SPAM[2], GAS[2], WEATHER[2], CovType[2], Asfault[3], HAR[4]) for experiments. The synthetic data streams were generated by the MOA framework[5], each with a length of 100k. They were designed to undergo concept drift (gradual, sudden, and incremental) and mutations in class imbalance rates every 25k instances to test the robustness of the model in non-stationary environments. Detailed statistical characteristics of all data streams (number of instances, attributes, classes, etc.) are summarized in Table 1.

[2]https://github.com/canoalberto/imbalanced-streams/

[3]https://sites.google.com/view/uspdsrepository

[4]https://archive.ics.uci.edu/datasets

[5]https://moa.cms.waikato.ac.nz/

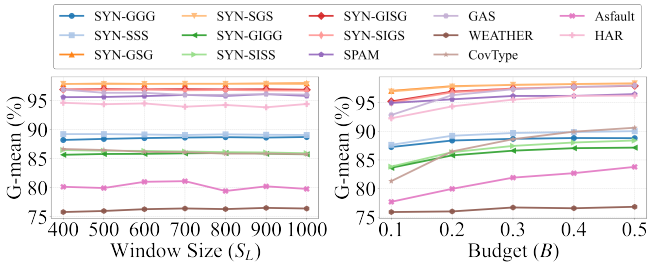


Figure 3: Effect of parameter variation to NILE on all data streams.

Evaluation Metrics. The experiments adopt the standard Test-Then-Train evaluation method. Given the class imbalance in data streams, we primarily employ G-mean, F1-score, Recall, and Precision as performance metrics. Due to space limitations, we mainly report the experimental results of G-mean in the main text; detailed analyses of other metrics are provided in the Supplementary Material.

Baselines. We compared NILE with three categories of state-of-the-art imbalanced data stream classification or representation learning algorithms: 1) Fully supervised algorithms (MOOB [Wang *et al.*, 2016], KUE [Cano and Krawczyk, 2020], ARF_{RE} [Boiko Ferreira *et al.*, 2019], ROSE [Cano and Krawczyk, 2022]); 2) Active learning algorithms (CALMID [Liu *et al.*, 2021], MicFoal [Liu *et al.*, 2023b], LEPID [Wu *et al.*, 2024]); 3) SSL algorithms ($\text{IOE}_{\text{Prob.}}$, IOE_{1NN} [Vafaie *et al.*, 2020], CEMC [Wang *et al.*, 2026c]). Detailed descriptions are listed in the Supplementary Material.

Implementation Details. All comparison algorithms are open-source: fully supervised baselines, CALMID, and MicFoal were integrated into the Java-based MOA framework by [Aguiar *et al.*, 2024], while the remaining algorithms were implemented based on Python code provided by the original authors. Regarding label settings, fully supervised algorithms use 100% labels, semi-supervised algorithms randomly sample 20% labels, and the labeling budget for both active learning and NILE is uniformly set to $B = 0.2$. All algorithms use the first 500 samples for initialization. For NILE, we set the label queue capacity $S_L = 500$ to ensure a fair comparison with CALMID and MicFoal. We constructed a three-layer MLP network using PyTorch for representation learning and utilized the Adam optimizer for weight updates. Further reproduction details can be found in our open-source code[1].

5.2 Parameter Sensitivity Analysis

To evaluate the stability and robustness of NILE under different hyperparameter configurations, we analyzed the impact of label queue capacity S_L and labeling budget B on model performance (G-mean). Regarding the label queue capacity S_L, this parameter jointly determines the capacities of the label queue, class-specific labeled queues, and the unlabeled queue, thereby collectively constraining the data scale for class prior estimation, prototype construction, and semi-supervised learning. As shown in Fig. 3 (left), when S_L varies within the range $[400, 1000]$, the G-mean curves of NILE on the vast majority of datasets remain stable without significant fluctuations. This indicates that NILE is insensitive to this parameter. For the labeling budget B, its value range is set to $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. As illustrated in Fig. 3 (right), with the increase of B, the classification performance on all datasets exhibits an expected upward trend, characterized by significant improvement followed by a plateau. This is attributed to the hybrid active learning strategy, which can efficiently select more informative samples under a larger label budget and further assist in building a more precise Simplex ETF structure and decision boundary.

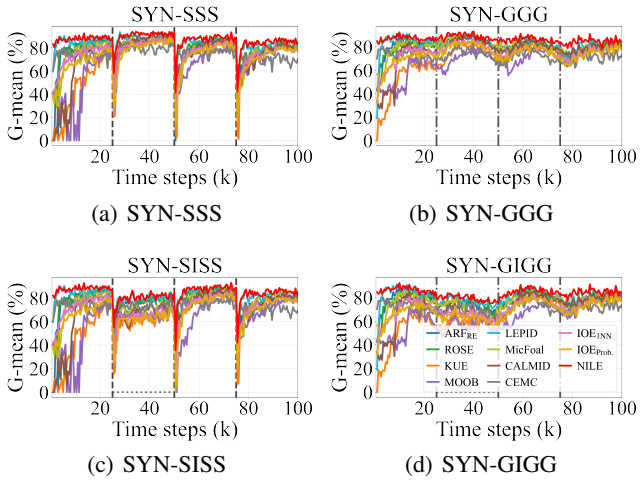


Figure 4: G-mean curves of all comparison algorithms.

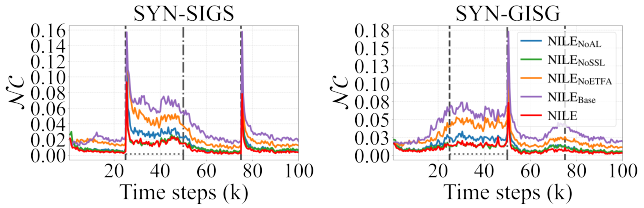
5.3 Comparative Experiments

Table 2 reports the G-mean performance of each algorithm on 14 datasets. NILE achieves the best results on 13 of them, with an average G-mean of 90.57%, significantly outperforming the best-performing baselines ARF_{RE}, LEPID, and CEMC. Compared with fully supervised algorithms, NILE achieves an average performance improvement of 7.17% with only a 20% labeling budget. This is primarily attributed to the Simplex ETF structure explicitly inducing NC in the feature space, effectively addressing the deficiencies of traditional methods in feature representation learning. Compared to AL and SSL baselines, NILE overcomes the insufficiency of single strategies in utilizing distribution information through the closed loop of hybrid active learning and NC-adapted SSL calibration, demonstrating remarkable robustness in complex dynamic class imbalance and concept drift scenarios.

To intuitively evaluate the dynamic adaptation capability of algorithms in non-stationary environments, Fig. 4 displays the G-mean curves of all comparison algorithms under different drift types. Evidently, NILE achieves optimal results in every stable concept stage, proving its effectiveness in addressing class imbalance. In sudden drift scenarios (Fig. 4(a)), comparison algorithms generally experience significant performance decline and recovery lag. Conversely, NILE can rapidly perceive distribution changes and guide the ETF classifier to complete realignment, achieving the fastest recovery with minimal performance retraction. Unlike sudden drift, in the face of gradual drift (Fig. 4(b)), the magnitude of performance decline for all algorithms decreases, but the affected duration is prolonged. However, the performance curve of

Dataset	Supervised Learning Algorithms				Active Learning Algorithms			Semi-Supervised Algorithms			NILE
	MOOB	KUE	ARF_{RE}	ROSE	CALMID	MicFoal	LEPID	$IOE_{Prob.}$	IOE_{1NN}	CEMC	
SYN-GGG	71.42	76.93	85.67	82.61	80.99	84.66	86.53	76.35	77.37	78.22	**88.40**
SYN-SSS	73.34	78.96	86.40	83.91	81.85	85.01	88.37	75.90	79.26	80.75	**89.23**
SYN-GSG	86.21	82.63	95.79	93.92	93.66	95.63	95.39	87.05	89.19	91.32	**97.90**
SYN-SGS	86.93	83.31	95.41	93.54	94.81	95.31	96.22	88.54	88.92	92.65	**97.84**
SYN-GIGG	66.56	72.48	83.71	80.22	76.92	81.63	84.39	71.67	73.23	76.79	**85.80**
SYN-SISS	68.75	73.22	84.58	81.45	77.80	82.01	86.00	70.94	74.41	78.97	**86.35**
SYN-GISG	77.79	76.79	95.26	93.00	91.01	93.03	94.83	83.81	85.70	91.22	**96.95**
SYN-SIGS	81.77	77.71	94.74	92.72	92.26	92.66	95.55	84.39	85.09	92.36	**96.83**
Spam	89.32	80.01	94.16	94.27	87.08	89.21	89.51	80.59	82.92	94.03	**95.60**
GSAD	55.73	45.32	90.00	84.11	72.18	89.06	90.24	63.20	78.88	94.89	**96.33**
Weather	71.62	63.70	75.41	75.02	69.29	65.09	70.60	65.86	68.76	65.42	**75.99**
CovType	76.77	70.83	**87.35**	87.96	85.81	86.54	79.13	71.64	77.01	81.70	86.49
Asfault	69.13	14.94	78.54	78.24	65.06	66.92	55.50	66.43	70.62	71.16	**79.94**
HAR	8.24	8.39	20.54	14.75	7.87	21.94	87.27	51.91	62.18	93.55	**94.37**
Avg.	70.26	64.66	83.40	81.12	76.90	80.62	85.68	74.16	78.11	84.50	**90.57**
Avg. Rank	9.21	10.29	3.00	4.50	6.93	5.00	3.71	9.00	7.43	5.71	**1.21**

Table 2: G-mean results (%) of all comparison algorithms.



(a) SYN-SIGS (b) SYN-GISG

Figure 5: NC metric of NILE and its variants.

Dataset	$NILE_{NoETFA}$	$NILE_{NoAL}$	$NILE_{NoSSL}$	$NILE_{Base}$	NILE
SYN-GGG	87.17	87.31	87.99	85.97	**88.40**
SYN-SSS	88.03	87.93	88.48	86.59	**89.23**
SYN-GSG	96.61	97.07	97.42	94.01	**97.90**
SYN-SGS	96.56	97.02	97.51	94.34	**97.84**
SYN-GIGG	83.38	84.05	85.02	81.83	**85.80**
SYN-SISS	83.97	84.65	85.18	81.60	**86.35**
SYN-GISG	94.60	95.54	96.73	91.53	**96.95**
SYN-SIGS	94.39	95.41	96.67	91.73	**96.83**

Table 3: G-mean results (%) of NILE and its variants.

NILE remains highly smooth and stable throughout the evolution cycle, effectively suppressing sawtooth fluctuations during model adaptation to new concepts. When facing incremental drift (steps 25k to 50k in Fig. 4(c) and 4(d)), even though the concept changes continuously, NILE still maintains optimal performance over the long term.

5.4 Ablation Studies

To verify the contribution of each component, we compared four variants: $NILE_{NoETFA}$ which removes dynamic ETF adjustment, $NILE_{NoAL}$ which uses random sampling instead of hybrid active learning, $NILE_{NoSSL}$ which removes semi-supervised learning, and the baseline $NILE_{Base}$ which removes all the above modules. Meanwhile, to quantify the evolution of the geometric structure, we calculate the NC metric within a sliding window. Given the feature representation $\bar{\mathbf{z}}_{k,i}$ of the i-th sample in the window, where k is its true label, we calculate the mean of the cosine similarity between the feature representation and the ETF classifier vectors of other classes within the window: $\text{Avg}_{i,k\neq k'}\{\langle \bar{\mathbf{z}}_{k,i}, \bar{\boldsymbol{e}}_{k'}\rangle + \frac{1}{K-1}\}$. A value closer to 0 indicates that the feature space conforms more closely to the ideal Neural Collapse structure.

The results in Table 3 show that NILE achieves the best performance on all datasets, while all variants exhibit varying degrees of performance degradation, confirming the necessity of each component. Among them, the performance decline of $NILE_{NoETFA}$ is the most significant. Fig. 5 further reveals that after drift occurs, the NC metric of $NILE_{NoETFA}$ oscillates violently and moves away from the 0-axis, while NILE always stays close to the 0-axis, confirming the critical role of the dynamic alignment mechanism in maintaining the optimal feature-classifier match in non-stationary environments. $NILE_{NoAL}$ follows, indicating that hybrid active learning is superior to random strategies in capturing imbalanced boundary samples. The disadvantage of $NILE_{NoSSL}$ validates the gain of unlabeled data for precise decision boundaries. The performance of $NILE_{Base}$ is the worst, highlighting the necessity of synergy among the modules.

6 Conclusion

In this paper, we analyze the challenges of learning from label-limited streams exhibiting dynamic class imbalance and concept drift. We argue that maintaining a geometrically optimal feature space is essential for these challenges. To this end, we propose NILE, a framework incorporating NC properties into stream learning. By integrating the ETF classifier with dynamic adaptation, NILE effectively mitigates geometric mismatch during drifts. Our hybrid active learning strategy efficiently identifies critical samples, while NC-adapted semi-supervised learning leverages unlabeled data to enhance feature representation learning. Moreover, the dynamic ETF adjustment mechanism ensures classifier geometric structure realignment with new concepts. Extensive evaluations confirm that NILE significantly outperforms state-of-the-art methods, successfully guiding representations toward the ideal NC state under label-limited data stream scenarios with dynamic class imbalance and concept drift.

Acknowledgments

Hongliang Wang was supported by the China Scholarship Council (202506070066). Junming Shao was supported by the National Natural Science Foundation of China (62376054), Sichuan Science and Technology Program (2025YFHZ0236, DQ202411), Shenzhen Science and Technology Program (JCYJ20230807120008016), AVIC United Technology Center for Intelligent Decision-making and Coordinated Control Mechanism Model Research grant (1725-GSYS02000000001-ZB-Z013-0), Guangdong Basic and Applied Basic Research Foundation (2024A1515011634), .

References

[Abdi *et al.*, 2024] Yousef Abdi, Mohammad Asadpour, and Mohammad-Reza Feizi-Derakhshi. An ensemble-based semi-supervised learning approach for non-stationary imbalanced data streams with label scarcity. *Applied Soft Computing*, 167:112353, 2024.

[Aguiar *et al.*, 2024] Gabriel Aguiar, Bartosz Krawczyk, and Alberto Cano. A survey on learning from imbalanced data streams: taxonomy, challenges, empirical study, and reproducible experimental framework. *Machine Learning*, 113(7):4165–4243, 2024.

[Boiko Ferreira *et al.*, 2019] Luis Eduardo Boiko Ferreira, Heitor Murilo Gomes, Albert Bifet, and Luiz S. Oliveira. Adaptive random forests with resampling for imbalanced data streams. In *IJCNN*, pages 1–6, 2019.

[Cacciarelli and Kulahci, 2024] Davide Cacciarelli and Murat Kulahci. Active learning for data streams: a survey. *Machine Learning*, 113(1):185–239, 2024.

[Cano and Krawczyk, 2020] Alberto Cano and Bartosz Krawczyk. Kappa updated ensemble for drifting data stream mining. *Machine Learning*, 109(1):175–218, 2020.

[Cano and Krawczyk, 2022] Alberto Cano and Bartosz Krawczyk. Rose: robust online self-adjusting ensemble for continual learning on imbalanced drifting data streams. *Machine Learning*, 111(7):2561–2599, 2022.

[Din *et al.*, 2022] Salah Ud Din, Jay Kumar, Junming Shao, Cobbinah Bernard Mawuli, and Waldiodio David Ndiaye. Learning high-dimensional evolving data streams with limited labels. *IEEE Transactions on Cybernetics*, 52(11):11373–11384, 2022.

[Fahy *et al.*, 2022] Conor Fahy, Shengxiang Yang, and Mario Gongora. Scarcity of labels in non-stationary data streams: A survey. *ACM Comput. Surv.*, 55(2):1–39, 2022.

[Gomes *et al.*, 2022] Heitor Murilo Gomes, Maciej Grzenda, Rodrigo Mello, Jesse Read, Minh Huong Le Nguyen, and Albert Bifet. A survey on semi-supervised learning for delayed partially labelled data streams. *ACM Comput. Surv.*, 55(4):1–42, 2022.

[Gunasekara *et al.*, 2024] Nuwan Gunasekara, Bernhard Pfahringer, Heitor Murilo Gomes, Albert Bifet, and Yun Sing Koh. Recurrent concept drifts on data streams. In *IJCAI*, pages 8029–8037, 2024.

[Guo *et al.*, 2024] Yinan Guo, Jiayang Pu, Botao Jiao, Yanyan Peng, Dini Wang, and Shengxiang Yang. Online semi-supervised active learning ensemble classification for evolving imbalanced data streams. *Applied Soft Computing*, 155:111452, 2024.

[Hong and Ling, 2024] Wanli Hong and Shuyang Ling. Neural collapse for unconstrained feature model under cross-entropy loss with imbalanced data. *Journal of Machine Learning Research*, 25(192):1–48, 2024.

[Hu *et al.*, 2024] Zhanxuan Hu, Yichen Wang, Hailong Ning, Yonghang Tai, and Feiping Nie. Neural collapse inspired semi-supervised learning with fixed classifier. *Information Sciences*, 667:120469, 2024.

[Jiang *et al.*, 2025] Jun Jiang, Bin Wang, Quan Tang, Guoxiang Zhong, Xuhao Tang, and Joel J. P. C. Rodrigues. Incremental semi-supervised learning for data streams classification in internet of things. *IEEE Transactions on Network and Service Management*, 22(3):2489–2501, 2025.

[Jiao *et al.*, 2025] Botao Jiao, Heitor Murilo Gomes, Bing Xue, Yinan Guo, and Mengjie Zhang. A semi-supervised active learning neural network for data streams with concept drift. *IEEE Computational Intelligence Magazine*, 20(1):18–33, 2025.

[Krawczyk and Cano, 2019] Bartosz Krawczyk and Alberto Cano. Adaptive ensemble active learning for drifting data stream mining. In *IJCAI*, pages 2763–2771, 2019.

[Li *et al.*, 2024] Ang Li, Meng Han, Dongliang Mu, Zhihui Gao, and Shujuan Liu. Online active learning method for multi-class imbalanced data stream. *Knowledge and Information Systems*, 66(4):2355–2391, 2024.

[Liao *et al.*, 2023] Guobo Liao, Peng Zhang, Hongpeng Yin, Xuanhong Deng, Yanxia Li, Han Zhou, and Dandan Zhao. A novel semi-supervised classification approach for evolving data streams. *Expert Systems with Applications*, 215:119273, 2023.

[Liu and Qin, 2025] Litian Liu and Yao Qin. Detecting out-of-distribution through the lens of neural collapse. In *CVPR*, pages 15424–15433, 2025.

[Liu *et al.*, 2021] Weike Liu, Hang Zhang, Zhaoyun Ding, Qingbao Liu, and Cheng Zhu. A comprehensive active learning method for multiclass imbalanced data streams with concept drift. *Knowledge-Based Systems*, 215:106778, 2021.

[Liu *et al.*, 2023a] Sanmin Liu, Shan Xue, Jia Wu, Chuan Zhou, Jian Yang, Zhao Li, and Jie Cao. Online active learning for drifting data streams. *IEEE Transactions on Neural Networks and Learning Systems*, 34(1):186–200, 2023.

[Liu *et al.*, 2023b] Weike Liu, Cheng Zhu, Zhaoyun Ding, Hang Zhang, and Qingbao Liu. Multiclass imbalanced and concept drift network traffic classification framework based on online active learning. *Engineering Applications of Artificial Intelligence*, 117:105607, 2023.

[Loezer *et al.*, 2020] Lucas Loezer, Fabrício Enembreck, Jean Paul Barddal, and Alceu de Souza Britto. Cost-sensitive learning for imbalanced data streams. In *SAC*, pages 498–504, 2020.

[Malialis *et al.*, 2022] Kleanthis Malialis, Christos G. Panayiotou, and Marios M. Polycarpou. Nonstationary data stream classification with online active learning and siamese neural networks. *Neurocomputing*, 512:235–252, 2022.

[Papyan *et al.*, 2020] Vardan Papyan, X. Y. Han, and David L. Donoho. Prevalence of neural collapse during the terminal phase of deep learning training. *Proceedings of the National Academy of Sciences*, 117(40):24652–24663, 2020.

[Seo *et al.*, 2024] Minhyuk Seo, Hyunseo Koh, Wonje Jeung, Minjae Lee, San Kim, Hankook Lee, Sungjun Cho, Sungik Choi, Hyunwoo Kim, and Jonghyun Choi. Learning equi-angular representations for online continual learning. In *CVPR*, pages 23933–23942, 2024.

[Tanha *et al.*, 2022] Jafar Tanha, Negin Samadi, Yousef Abdi, and Nazila Razzaghi-Asl. Cpssds: Conformal prediction for semi-supervised classification on data streams. *Information Sciences*, 584:212–234, 2022.

[Vafaie *et al.*, 2020] Parsa Vafaie, Herna Viktor, and Wojtek Michalowski. Multi-class imbalanced semi-supervised learning from streams through online ensembles. In *ICDMW*, pages 867–874, 2020.

[Wang and Li, 2018] Yi Wang and Tao Li. Improving semi-supervised co-forest algorithm in evolving data streams. *Applied Intelligence*, 48(10):3248–3262, 2018.

[Wang *et al.*, 2015] Shuo Wang, Leandro L. Minku, and Xin Yao. Resampling-based ensemble methods for online class imbalance learning. *IEEE Transactions on Knowledge and Data Engineering*, 27(5):1356–1368, 2015.

[Wang *et al.*, 2016] Shuo Wang, Leandro L. Minku, and Xin Yao. Dealing with multiple classes in online class imbalance learning. In *IJCAI*, pages 2118–2124, 2016.

[Wang *et al.*, 2026a] Hongliang Wang, Liangxv Pan, Zhonglin Wu, Lei Liu, Haifeng Peng, Yulin Tao, Qinli Yang, and Junming Shao. A deep metric framework for reliable semi-supervised learning on evolving data streams. *Applied Intelligence*, 56(2):49, 2026.

[Wang *et al.*, 2026b] Hongliang Wang, Zhonglin Wu, Jinxia Guo, Wei Han, Lei Liu, Qinli Yang, and Junming Shao. Exploiting reliable evolving micro-clusters for robust semi-supervised learning on data streams. *Information Sciences*, 735:123069, 2026.

[Wang *et al.*, 2026c] Hongliang Wang, Zhonglin Wu, Jinxia Guo, Qirui Hao, Xinyu Liu, Hongyuan Liu, Qinli Yang, and Junming Shao. Learning contrastive evolving micro-clusters for robust semi-supervised data stream classification. *IEEE Transactions on Cybernetics*, 2026.

[Wu *et al.*, 2024] Zhonglin Wu, Hongliang Wang, Jingxia Guo, Qinli Yang, and Junming Shao. Learning evolving prototypes for imbalanced data stream classification with limited labels. *Information Sciences*, 679:120979, 2024.

[Wu *et al.*, 2026] Zhonglin Wu, Hongliang Wang, Tongze Zhang, Hongyuan Liu, Jinxia Guo, Qinli Yang, and Junming Shao. Region-based deep metric learning for tackling class overlap in online semi-supervised data stream classification. *Information Fusion*, 130:104126, 2026.

[Xiao *et al.*, 2024] Ruixuan Xiao, Lei Feng, Kai Tang, Junbo Zhao, Yixuan Li, Gang Chen, and Haobo Wang. Targeted representation alignment for open-world semi-supervised learning. In *CVPR*, pages 23072–23082, 2024.

[Yan *et al.*, 2024] Hongren Yan, Yuhua Qian, Furong Peng, Jiachen Luo, Zheqing Zhu, and Feijiang Li. Neural collapse to multiple centers for imbalanced data. In *NeurIPS*, volume 37, pages 65583–65617, 2024.

[Yang *et al.*, 2022] Yibo Yang, Shixiang Chen, Xiangtai Li, Liang Xie, Zhouchen Lin, and Dacheng Tao. Inducing neural collapse in imbalanced learning: Do we really need a learnable classifier at the end of deep neural network? In *NeurIPS*, volume 35, pages 37991–38002, 2022.

[Yu *et al.*, 2025] Hang Yu, Jiahao Wen, Yiping Sun, Xiao Wei, and Jie Lu. Ca-gnn: A competence-aware graph neural network for semi-supervised learning on streaming data. *IEEE Transactions on Cybernetics*, 55(2):684–697, 2025.

[Zhang *et al.*, 2022] Hang Zhang, Weike Liu, and Qingbao Liu. Reinforcement online active learning ensemble for drifting imbalanced data streams. *IEEE Transactions on Knowledge and Data Engineering*, 34(8):3971–3983, 2022.

[Zhang *et al.*, 2025] Shuoxi Zhang, Zijian Song, and Kun He. Neural collapse inspired knowledge distillation. *AAAI*, 39(21):22542–22550, 2025.

[Zhao and Koh, 2020] Di Zhao and Yun Sing Koh. Feature drift detection in evolving data streams. In *DEXA*, 2020.

[Zhao *et al.*, 2022] Di Zhao, Yun Sing Koh, and Philippe Fournier-Viger. Measuring drift severity by tree structure classifiers. In *IJCNN*, pages 1–8, 2022.

[Zhou *et al.*, 2022] Jinxin Zhou, Xiao Li, Tianyu Ding, Chong You, Qing Qu, and Zhihui Zhu. On the optimization landscape of neural collapse under MSE loss: Global optimality with unconstrained features. In *ICML*, volume 162, pages 27179–27202, 2022.

[Zhou *et al.*, 2023] Han Zhou, Hongpeng Yin, Xuanhong Deng, and Yuyu Huang. Online harmonizing gradient descent for imbalanced data streams one-pass classification. In *IJCAI*, pages 2468–2475, 2023.

[Zhu *et al.*, 2021] Zhihui Zhu, Tianyu Ding, Jinxin Zhou, Xiao Li, Chong You, Jeremias Sulam, and Qing Qu. A geometric analysis of neural collapse with unconstrained features. In *NeurIPS*, volume 34, pages 29820–29834, 2021.

[Žliobaitė *et al.*, 2014] Indrė Žliobaitė, Albert Bifet, Bernhard Pfahringer, and Geoffrey Holmes. Active learning with drifting streaming data. *IEEE Transactions on Neural Networks and Learning Systems*, 25(1):27–39, 2014.