OFIGM: ONLINE FEATURE-INSTANCE GRAPH MEMORY FOR OPEN-WORLD CLASS-INCREMENTAL LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Real-world data streams are inherently open and nonstationary: previously unseen classes may emerge, labels can be sparse or delayed, and the observable feature space itself may evolve over time. However, existing continual, open-world, and streaming learning methods typically address these challenges in isolation. We propose OFIGM, an **Online Feature-Instance Graph Memory** framework for open-world class-incremental learning under simultaneous class and feature evolution. OFIGM maintains a bounded dynamic bipartite graph comprising transient instance nodes and persistent feature nodes, performs sparse graph message passing to learn schema-adaptive representations, and preserves previously acquired class knowledge through slowly updated prototype memory. A graph-aware novelty score integrates prototype distance, predictive uncertainty, and structural residuals to identify unknown samples, while a persistence-aware unknown buffer promotes coherent novel groups to new classes without global retraining. Across Digits, Wine, and Iris, OFIGM achieves mean Macro-F1 of 0.819, novel-class F1 of 0.882, and novelty AUROC of 0.866. Relative to a matched no-graph ablation, graph memory improves novelty AUROC from 0.926 to 0.946 on Wine and from 0.755 to 0.794 on Iris. We further evaluate generalization on an external Louisiana shellfish outbreak dataset containing 11 documented episodes from 2002-2025, where OFIGM achieves Macro-F1 of 0.474 and outbreak recall of 0.787. These results demonstrate the feasibility of persistent feature-instance graph memory as a unified mechanism for novel-class detection, incremental class creation, and continual adaptation in evolving open-world streams.

1 INTRODUCTION

Machine learning systems deployed in fields such as autonomous surveillance (Park et al., 2025), recommendation systems (Musanase et al., 2023), cybersecurity (Mohamed, 2025), remote sensing (Nikoo et al., 2025), and clinical decision support (Susanto et al., 2023) rarely operate under the closed-world assumption that training categories and input variables remain constant over time. Instead, data arrive sequentially, distributions shift, labels are costly, and categories that were absent during initialization can appear later (He et al., 2023). Class-incremental learning (CIL) (Bako & Kalita, 2026) formalizes one part of this problem: a learner must absorb new classes while continuing to discriminate all classes observed so far, without task identity at inference (van de Ven et al., 2022; Zhou et al., 2024). Its central difficulty is the stability-plasticity tradeoff and rapid adaptation can erase old knowledge, whereas excessive stability prevents learning genuinely new concepts (Li & Hoiem, 2018; Rebuffi et al., 2017; Liu et al., 2023b).

Recent CIL work has substantially strengthened retention, but the assumptions remain narrower than a fully open stream. Non-exemplar semi-supervised CIL shows that contrastive representations and prototype learning can exploit abundant unlabeled data while storing no old raw examples (Liu et al., 2024). Online CIL methods such as masking, linkage, and guidance (MLG) for CIL explicitly address single-pass learning, task-recency bias, class imbalance, and abrupt feature drift (Liang et al., 2025). Expansion-based approaches such as TagFex introduce task-agnostic features to reduce collisions among task-specific feature spaces (Zheng et al., 2025); order-robust CIL such as GDDSG uses graph-driven class grouping to reduce sensitivity to class arrival order (Lai et al., 2025). Few-shot

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

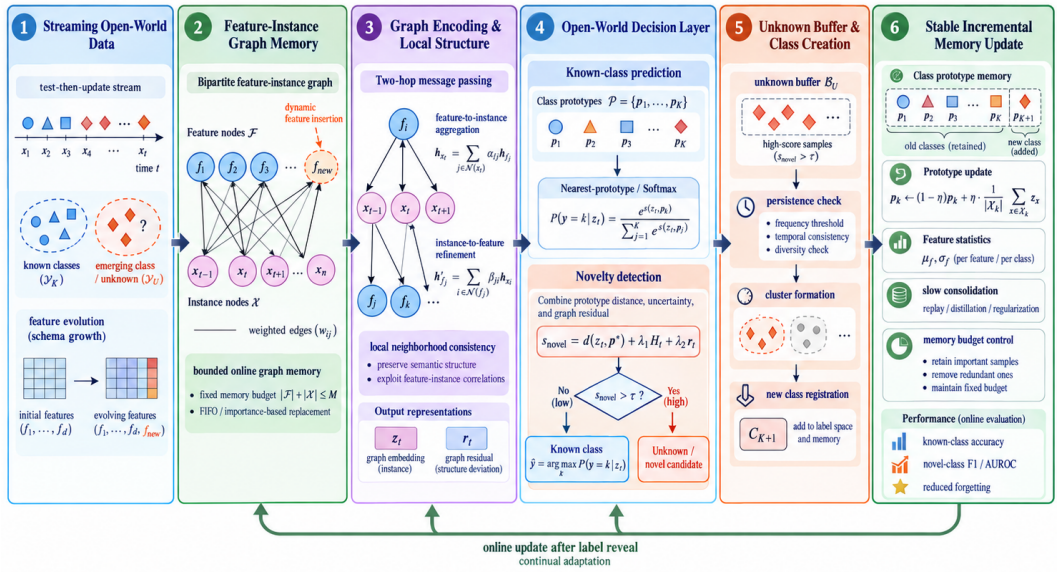

Figure 1: Overview of the proposed OFIGM method for open-world class-incremental learning.

CIL methods further study scarce supervision, including progressive curriculum/virtual-class training and prompt-based concept learning (Hu et al., 2025; Li et al., 2025). These advances are important, but they generally assume that incoming classes are already identified as incremental classes rather than first having to be discovered from an unsegmented stream.

Streaming emerging-new-class (SENC) learning addresses that missing detection step. SENCForest and SENC-MaS use random trees and matrix sketches, respectively (Mu et al., 2017a;b); SEEN combines novelty detection, graph-based semi-supervised classification, and online updating (Zhu & Li, 2020); and KNNENS unifies known-class classification and new-class detection with a fixed-size nearest-neighbor ensemble while avoiding the assumption that all new-class labels are immediately available (Zhang et al., 2023). HANOL further studies the simultaneous occurrence of new classes and online label shift (Qian et al., 2025). Open-world continual learning (OWCL) and novel-class discovery provide another complementary perspective by integrating novelty detection with continual retention or by discovering unknown categories from unlabeled data (Cao et al., 2022; Vaze et al., 2022; Liu et al., 2023a; Jin et al., 2024; Kim et al., 2025).

A second axis of openness is less often integrated with CIL: *the feature space can evolve*. New sensors, assays, attributes, vocabulary terms, or data channels may become available after deployment. SAFC explicitly studies simultaneous augmentation of feature and class spaces and shows why model reuse is necessary when new features reveal new categories (Hou et al., 2023). Incremental label-distribution learning likewise considers jointly expanding feature and label spaces (Xu et al., 2025). Meanwhile, bipartite instance-feature graph models demonstrate that feature nodes provide a natural representation for heterogeneous or newly attached attributes (Hsu & Li, 2026). Consequently, the existence of an instance-feature graph is not itself our novelty. The unresolved problem is whether such a graph can become a *bounded online memory* that supports class prediction, novelty detection, class creation, and feature growth at the same time.

This gap is practically important. A deployed learner may receive an unlabeled sample from a novel class at the same moment that its available measurements differ from earlier samples. If the method assumes a fixed schema, new attributes require preprocessing or retraining. If it assumes task boundaries, novelty is externally supplied rather than detected. If it optimizes only novelty detection, it can create false new classes when known-class distributions drift. If it optimizes only retention, it can suppress plasticity. We therefore ask: *Can persistent feature structure provide novelty evidence without sacrificing the slow semantic memory needed to retain old classes?*

We address this question with OFIGM, an **O**nline **F**eature-**I**nstance **G**raph **M**emory for open-world CIL as illustrated in Figure 1. Persistent feature nodes summarize long-lived feature statistics,

recent instances are transient nodes, and normalized observed values form sparse edges. Feature-to-instance-to-feature message passing induces relational embeddings without materializing a dense instance graph. A separate prototype bank acts as slow class memory. Novelty combines distance to known prototypes, uncertainty, and a feature-graph residual; candidate unknowns are buffered and promoted only after persistent, coherent evidence. New feature nodes can be inserted without resizing previously stored examples.

We study four research questions: **RQ1** does feature-graph memory improve known-vs.-novel discrimination beyond a prototype-only wrapper? **RQ2** can the same learner maintain novel-class F1 while new classes are registered online? **RQ3** does the slow prototype path prevent catastrophic forgetting as novelty adaptation becomes more aggressive? **RQ4** can the architecture remain functional when new features arrive after initialization?

The main contributions of our work are summarized below.

1. **Open-world CIL with evolving features.** We formulate a one-pass setting in which both the class set and observed feature set can grow, labels can be sparse or delayed, and no task-boundary oracle is available.
2. **Feature-instance graph memory rather than a static graph encoder.** Persistent feature nodes, bounded transient instance nodes, and value-aware edges provide an online substrate for relational representation and schema extension.
3. **Fast novelty and slow semantic memory.** A graph-aware novelty path is coupled to prototype anchoring and persistence-based class creation, explicitly targeting the stability-plasticity dilemma.
4. **Evaluation bridging SENC, OCIL, and feature growth.** We select stream benchmarks and baselines from SEEN/KNNENS/HANOL, modern OCIL/CIL methods, and simultaneous feature-class augmentation, while separating direct matched comparisons from contextual results reported under incompatible protocols.
5. **Executable evidence with transparent limitations.** Three-seed sanity checks show AUROC gains from graph memory on Wine and Iris, but no consistent forgetting gain.

2 Related Work

2.1 Class-incremental, online, and exemplar-free learning

CIL methods are commonly organized around regularization/distillation, replay, prototype/statistics retention, and architecture expansion (Zhou et al., 2024). Replay methods such as iCaRL and later OCIL variants usually achieve strong retention but require storing exemplars (Rebuffi et al., 2017). Exemplar-free analytic approaches such as DS-AL use recursive closed-form updates to avoid revisiting old raw data, while online hyperparameter optimization adapts the stability–plasticity balance across phases (Zhuang et al., 2024; Liu et al., 2023b). Semi-IPC extends the exemplar-free direction to sparse labels by combining contrastive pretraining, incremental prototypes, pseudo-feature resampling, and unsupervised regularization (Liu et al., 2024). In online CIL, MLG targets task-recency bias and abrupt feature drift using batch-level logit masking, cross-feature fusion, and accumulative mean feature distillation (Liang et al., 2025). TagFex addresses feature collision in expansion-based CIL using a continually trained task-agnostic model and merge attention (Zheng et al., 2025). GDDSG studies another underexplored source of variance using graph coloring and dynamic similarity groups (Lai et al., 2025). Few-shot methods such as PGLS and PCL emphasize rapid acquisition from few labeled examples, using progressive virtual classes or prompt/concept structure with feature-distribution replay (Hu et al., 2025; Li et al., 2025). Federated and multi-label variants expose related imbalance issues: FedCBDR balances replay across classes in federated CIL, while RebLL rebalances positive/negative information in multi-label CIL (Qi et al., 2025; Du et al., 2025). These works inform our retention and imbalance analysis, but they do not directly solve task-free novel-class detection under feature growth.

2.2 Streaming learning with emerging new classes

SENC learning assumes that a stream may contain previously unseen classes and therefore evaluates both ordinary classification and novel-class detection. SENCForest uses isolation-style random trees,

SENC-MaS uses compact matrix sketches, and SEEN combines a novel-class detector with graph-based semi-supervised label propagation and an update rule (Mu et al., 2017a;b; Zhu & Li, 2020). KNNENS uses a fixed-size kNN hypersphere ensemble to reduce false positives and update from new-class samples without requiring true labels for all test instances (Zhang et al., 2023). HANOL considers new classes together with time-varying label priors and supplies dynamic-regret guarantees (Qian et al., 2025). Recent SCIL and INSERTION formulations further combine sparse supervision, drift/imbalance, or task-free class expansion in open streams (Li et al., 2026a;b). These methods establish the most relevant evaluation principles for our work: test-then-update processing, explicit novel-class F1/detection metrics, bounded resources, and runtime reporting.

2.3 Open-world continual learning and novel-class discovery

Open-set recognition rejects unknowns, whereas novel-class discovery additionally separates rejected samples into coherent new categories. ORCA, GCD, and OpenNCD exploit representation learning and clustering/prototypes for mixed known and novel data (Cao et al., 2022; Vaze et al., 2022; Liu et al., 2023a). ORAL extends novel-class discovery to graph data, where cross-class edges can make known and unknown representations indistinguishable (Jin et al., 2024). OWCL work more directly couples novelty detection to continual retention (Kim et al., 2025). Relative to static discovery, our setting is causal and prequential: a learner cannot inspect the future unlabeled set, must decide before update, and must continue after a novel category becomes known.

2.4 Evolving feature spaces and graph-structured memory

Feature evolution appears in streaming-view learning, incremental label-distribution learning, and simultaneous augmentation of features and classes. SAFC studies the case where new measurements arrive and class space expands together, using regularization to reuse the previous model rather than retraining from scratch (Hou et al., 2023). Incremental Label Distribution Learning (ILDL) similarly transfers discriminative information as features and labels grow (Xu et al., 2025). Instance-feature bipartite models such as GTab demonstrate that representing attributes as nodes can support flexible tabular schemas (Hsu & Li, 2026). Our distinction is temporal: we use the bipartite structure as a bounded persistent memory for open-world decisions, with novelty, class registration, and forgetting evaluated online rather than as a static inductive graph-prediction problem.

3 Problem Formulation

At time $t = 1, 2, \ldots$, a learner receives an instance x_t defined on the currently observed feature set $\mathcal{F}_t$. Unlike fixed-dimensional learning, $\mathcal{F}_t$ can grow: $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$. The semantic class belongs to an open universe $\mathcal{C}^\star$, while the learner only knows $\mathcal{C}_t \subset \mathcal{C}^\star$ before processing x_t. An optional label y_t is observed according to an application-dependent labeling process; otherwise $y_t = \varnothing$. A sample can therefore be (i) a labeled known-class instance, (ii) an unlabeled known-class instance, or (iii) an unlabeled instance from a class not yet in $\mathcal{C}_t$. The learner must output $\hat{y}_t \in \mathcal{C}_t \cup \{\texttt{unknown}\}$ before using the current label, if any. It then updates its bounded state S_t using (x_t, y_t) and possibly admits a new class c_{new} once accumulated unknown evidence is sufficient. This test-then-train protocol prevents look-ahead bias. We optimize a multi-objective streaming criterion

$$\min \quad \underbrace{\mathcal{L}_{\text{known}}}_{\text{seen-class prediction}} + \lambda_u \underbrace{\mathcal{L}_{\text{novel}}}_{\text{unknown separation}} + \lambda_s \underbrace{\mathcal{L}_{\text{stab}}}_{\text{retention}} + \lambda_g \underbrace{\mathcal{L}_{\text{graph}}}_{\text{structural consistency}}, \tag{1}$$

subject to a memory budget M, a per-example compute budget, and online state updates. Evaluation additionally measures detection delay and false novel alarms because aggregate accuracy can hide a detector that reacts too late or rejects too many known samples.

4 Online Feature-Instance Graph Memory

4.1 Dynamic bipartite graph

At time t, OFIGM maintains a bounded graph

$$G_t = (\mathcal{V}_t^I \cup \mathcal{V}_t^F, \mathcal{E}_t), \tag{2}$$

where $\mathcal{V}_t^I$ contains at most B recent or representative instance nodes and $\mathcal{V}_t^F$ contains one persistent node per observed feature. An edge $(i, j) \in \mathcal{E}_t$ exists only when feature j is observed for instance i. Its scalar weight is

$$e_{ij} = m_{ij}\, r_j\, \psi_j(x_{ij}), \tag{3}$$

where m_{ij} is an observation mask, ψ_j is a robust online normalization (running median/interquartile range (IQR) or clipped z-score (Singh & Singh, 2020)), and $r_j \in [0, 1]$ is an exponentially updated reliability factor. Sparse inputs therefore induce sparse graphs.

Feature nodes persist beyond the removal of individual instance nodes. Each feature node stores an embedding h_j^F, running moments, and optionally class-conditional sufficient statistics. When an unseen feature $j^\star$ first appears, a new node is initialized from a shared feature initializer plus metadata (feature type, scale, missingness rate when available); no resizing of a dense first-layer weight matrix is required.

4.2 Two-pass feature–instance message passing

A single update uses two sparse passes. First, each instance aggregates the features it actually observes:

$$\alpha_{ij} = \mathrm{softmax}_{j \in \mathcal{N}(i)} \left(a_I^\top \sigma(W_I h_i^I + W_F h_j^F + w_e e_{ij}) \right), \tag{4}$$

$$\tilde{h}_i^I = \mathrm{LN}\left(h_i^I + \sum_{j \in \mathcal{N}(i)} \alpha_{ij}\, \phi_F(h_j^F, e_{ij}) \right). \tag{5}$$

Second, persistent feature nodes absorb bounded stream context:

$$\beta_{ji} = \mathrm{softmax}_{i \in \mathcal{N}(j)} \left(a_F^\top \sigma(U_F h_j^F + U_I \tilde{h}_i^I + u_e e_{ij}) \right), \tag{6}$$

$$\tilde{h}_j^F = (1 - \rho_j) h_j^F + \rho_j\, \mathrm{LN}\left(h_j^F + \sum_{i \in \mathcal{N}(j)} \beta_{ji}\, \phi_I(\tilde{h}_i^I, e_{ij}) \right), \tag{7}$$

where ρ_j is a slow feature-memory rate. This alternating update is related to graph attention (Veličković et al., 2018), but its purpose is not node classification on a fixed external graph. The graph is generated online from feature incidence, and the persistent side of the bipartition acts as a schema-aware memory.

The resulting instance embedding is

$$z_t = W_z[\tilde{h}_t^I \,\|\, \mathrm{pool}_{j \in \mathcal{N}(t)}\, \tilde{h}_j^F \,\|\, q_t], \tag{8}$$

where q_t contains stream metadata such as missingness fraction and normalized time-gap features when available.

4.3 Prototype-anchored known-class prediction

For each known class $c \in \mathcal{C}_t$, the model stores a normalized prototype p_c and effective support n_c. Trusted labeled embeddings update

$$p_c \leftarrow \mathrm{norm}\left((1 - \eta_c) p_c + \eta_c z_t \right), \qquad \eta_c = (n_c + 1)^{-\gamma}, \tag{9}$$

with a lower update rate for mature classes. Classification logits combine a linear head and prototype similarity:

$$\ell_c(z_t) = w_c^\top z_t + \tau_p \frac{z_t^\top p_c}{\|z_t\| \|p_c\|}. \tag{10}$$

When a new class is admitted, one row is appended to the classifier and one prototype is initialized; old output rows are not reinitialized.

To preserve geometry, trusted old-class examples and high-confidence pseudo-labeled examples use a prototype anchoring term

$$\mathcal{L}_{\mathrm{stab}} = \sum_{c \in \mathcal{C}_t} \omega_c \left(1 - \cos(\bar{z}_c, p_c^{\mathrm{stop}}) \right), \tag{11}$$

where p_c^{stop} is a stop-gradient snapshot and ω_c increases for underrepresented or old classes. This provides a compact alternative to storing a large exemplar set.

4.4 GRAPH-AWARE NOVELTY SCORE

A single confidence threshold is fragile under drift. We therefore combine three standardized signals:

$$s_t = \lambda_d \tilde{d}_t + \lambda_h \tilde{H}_t + \lambda_g \tilde{r}_t, \tag{12}$$

where

$$d_t = 1 - \max_{c \in \mathcal{C}_t} \cos(z_t, p_c), \tag{13}$$

$$H_t = -\sum_{c \in \mathcal{C}_t} \pi_c(z_t) \log \pi_c(z_t), \tag{14}$$

$$r_t = 1 - \frac{1}{|\mathcal{N}_I(t)|} \sum_{i \in \mathcal{N}_I(t)} \cos(z_t, z_i). \tag{15}$$

$\mathcal{N}_I(t)$ is induced through shared feature nodes rather than an explicitly materialized all-pairs graph. The tildes denote robust standardization using exponentially weighted median/scale estimates from trusted known predictions. An instance is marked unknown when $s_t > \theta_t$, with θ_t set to a target upper quantile of trusted-known scores.

4.5 PERSISTENCE-AWARE CLASS CREATION

Unknown decisions are not immediately converted into classes. Candidate embeddings enter an unknown buffer U_t. Periodically, a lightweight density graph is constructed only within U_t. A component Q is promoted to a new pseudo-class when it satisfies:

$$|Q| \geq m_{\min}, \qquad \mathrm{coh}(Q) \geq \kappa, \qquad \mathrm{age}(Q) \geq \Delta_{\min}, \tag{16}$$

where coherence is mean within-component cosine similarity and age enforces persistence across multiple arrivals. Its centroid initializes $p_{c_{new}}$. If a true label later arrives for any member, the pseudo-class identifier is mapped to the semantic label; otherwise evaluation can use optimal one-to-one assignment for discovery metrics, while semantic classification metrics use only label-confirmed classes.

4.6 LEARNING OBJECTIVE

For labeled samples we use cross-entropy. For high-confidence unlabeled known samples, pseudo-label consistency is weighted by confidence. A local graph loss preserves same-class smoothness while repelling trusted different-class pairs, which is defined by:

$$\mathcal{L}_{\mathrm{sup}} = -\log \pi_{y_t}(z_t), \tag{17}$$

$$\mathcal{L}_{\mathrm{pl}} = \mathbb{1}[q_t > q_{\min}] q_t \, \mathrm{CE}(\pi(z_t), \hat{y}_t), \tag{18}$$

$$\mathcal{L}_{\mathrm{graph}} = \sum_{(i,k) \in \mathcal{P}^+} (1 - \cos(z_i, z_k)) + \sum_{(i,k) \in \mathcal{P}^-} [\cos(z_i, z_k) - m]_+. \tag{19}$$

The online loss of OFIGM is

$$\mathcal{L}_t = \mathcal{L}_{\mathrm{sup}} + \beta \mathcal{L}_{\mathrm{pl}} + \mu \mathcal{L}_{\mathrm{graph}} + \xi \mathcal{L}_{\mathrm{stab}}, \tag{20}$$

where β, μ, and ξ are hyper-parameters. Updates use a small fixed number of gradient steps on the current micro-batch plus bounded replay of *embeddings/prototypes*, not unrestricted raw historical examples.

5 EXPERIMENTAL DESIGN

5.1 DATASETS AND STREAM CONSTRUCTION

We organize the experimental program into four complementary suites rather than forcing every baseline onto an incompatible protocol. **SENC core** uses Segment, Satimage, USPS, and Pendigits from SEEN (Mu et al., 2017a;b), with optional HAR/MNIST and the NYTimes, Powersupply,

Table 1: Dataset suite and the role of each protocol.

Suite	Representative datasets	Primary setting	Source alignment
SENC core	Segment, Satimage, USPS, Pendigits	unseen-class detection, sparse labels	SEEN/SENC
SENC extended	HAR, MNIST, NYTimes, Powersupply, ADL	long streams, runtime	KNNENS
Shift-aware	CIFAR-10, CINIC-10, EuroSAT, Fashion-MNIST, MNIST	prior drift + new classes	HANOL
OCIL retention	CIFAR-10/100, miniImageNet, ImageNet-100	one-pass forgetting, memory	MLG/OCIL
Feature growth	SAFC multi-view sets; grouped Digits	new attributes + new classes	SAFC/OFIGM

and ADL real streams used by KNNENS (Zhang et al., 2023). **Shift-aware open streams** follow HANOL on CIFAR-10, CINIC-10, EuroSAT, Fashion-MNIST, and MNIST, with gradual, periodic, and sudden prior changes. **OCIL retention evaluation** use CIFAR-10, CIFAR-100, miniImageNet, and ImageNet-100, matching the modern MLG setting when testing forgetting and memory-budget sensitivity. **Feature-growth tests** use SAFC-style multi-view descriptors where distinct feature groups arrive over time, together with a controlled grouped-feature version of Digits for the executable implementation.

For SENC streams, prediction precedes any optional label reveal. We evaluate both one-new-class-per-period protocols for historical comparability and asynchronous multi-class emergence. The post-prediction label budget is varied in $r_\ell \in \{1\%, 5\%, 10\%, 25\%\}$. For feature growth, attributes are partitioned into semantically defined groups when the source provides multiple views (e.g., GIST/HOG/LBP/SIFT-style descriptors in SAFC) or randomized groups for controlled stress. Appendix D adds an external event-centered evaluation using the 25-year Louisiana shellfish environmental archive and 11 documented outbreak episodes.

5.2 Baselines

Table 2 separates *direct* baselines, which can be run under the same stream protocol, from *contextual* baselines that answer a narrower adjacent question. SENCForest (Mu et al., 2017a), SENC-MaS (Mu et al., 2017b), SEEN (Zhu & Li, 2020), KNNENS (Zhang et al., 2023), and HANOL (Qian et al., 2025) are primary direct comparators for class detection/update. ER/ER-ACE and MLG emphasize one-pass retention under replay budgets; Semi-IPC tests non-exemplar sparse-label prototypes; DS-AL tests exemplar-free analytic retention; TagFex and GDDSG test feature expansion/order robustness under standard CIL. SAFC is the key feature-growth control, while GTab/I-GTab acts as a representation control rather than an open-world learner. PCL/PGLS, FedCBDR, and RebLL are discussed contextually because their few-shot, federated, or multi-label assumptions are not numerically commensurate with the primary SENC protocol.

5.3 Metrics and statistical protocol

We report prequential accuracy and macro-F1 over all classes, known-class macro-F1, novel-class F1 after emergence, and AUROC/AUPR for the binary known-vs-unknown decision. Forgetting is

$$F_T = \frac{1}{|\mathcal{C}_{T-1}|} \sum_c \left(\max_{t<T} A_{t,c} - A_{T,c} \right), \tag{21}$$

where $A_{t,c}$ is class-c accuracy at time t. Detection delay is the number of novel-class arrivals between the first appearance and the time a class is promoted. We also report false novel rate, peak memory, processed examples/s, and amortized update latency. Each randomized stream is repeated with at least 10 seeds for small/medium tabular datasets and 5 seeds for expensive image streams. Pairwise comparisons use a Wilcoxon signed-rank test (Woolson, 2007) across matched streams and Holm correction (Fu et al., 2014); effect sizes and confidence intervals accompany p-values.

Table 2: Baseline taxonomy. "Direct" means the method can be evaluated under a matched stream with a common prediction/update interface; "context" means results are used to motivate design or auxiliary evaluation, not copied into the main head-to-head table.

Method	Family	Main capability isolated	Role
SENCForest/SENC-MaS	SENC	tree vs. sketch novel-class detection	direct
SEEN	semi-supervised SENC	graph propagation + online update	direct
KNNENS	SENC	local-neighbor detection + runtime	direct
HANOL	open stream	label shift + emerging classes	direct
ER/ER-ACE/MLG	OCIL	replay, recency bias, feature drift	direct
Semi-IPC/DS-AL	exemplar-free CIL	prototype/statistical retention	direct
SAFC	feature+class growth	simultaneous schema/class augmentation	direct
GTab wrapper	feature graph	graph representation without OFIGM memory	ablation
TagFex/GDDSG	CIL	expansion collision / class-order robustness	context
PCL/PGLS/ FedCBDR/RebLL	specialized CIL	few-shot, federated, multi-label imbalance	context

Table 3: Preliminary executable sanity-check results (mean±std over three matched seeds). Higher is better except Forgetting and False-Novel.

Dataset	Method	Macro-F1 ↑	Novel-F1 ↑	AUROC ↑	Forget ↓	False-Novel ↓	ex/s ↑
Digits	Proto	$.814 \pm .018$	$.853 \pm .013$	$.857 \pm .016$	$.035 \pm .026$	$.026 \pm .008$	6814
	kNN	$.773 \pm .037$	$.803 \pm .038$	$.549 \pm .032$	$.035 \pm .014$	$.018 \pm .010$	1537
	OFIGM-noFG	$\mathbf{.820} \pm .016$	$\mathbf{.859} \pm .013$	$\mathbf{.861} \pm .008$	$.041 \pm .010$	$.028 \pm .005$	2905
	OFIGM	$.817 \pm .019$	$.857 \pm .018$	$.857 \pm .017$	$.045 \pm .019$	$.028 \pm .009$	1751
Wine	Proto	$\mathbf{.889} \pm .018$	$.938 \pm .054$	$\mathbf{.962} \pm .028$	$\mathbf{.060} \pm .003$	$.115 \pm .011$	8891
	kNN	$.788 \pm .049$	$.787 \pm .062$	$.887 \pm .048$	$.045 \pm .079$	$.078 \pm .069$	5210
	OFIGM-noFG	$.872 \pm .028$	$.964 \pm .031$	$.926 \pm .039$	$.069 \pm .023$	$\mathbf{.051} \pm .021$	4797
	OFIGM	$.872 \pm .016$	$\mathbf{.965} \pm .030$	$.946 \pm .023$	$.061 \pm .033$	$.053 \pm .038$	3048
Iris	Proto	$.738 \pm .039$	$\mathbf{.830} \pm .051$	$.785 \pm .055$	$.049 \pm .011$	$.073 \pm .029$	9017
	kNN	$.746 \pm .035$	$.692 \pm .042$	$\mathbf{.857} \pm .051$	$\mathbf{.044} \pm .026$	$.073 \pm .067$	5575
	OFIGM-noFG	$.765 \pm .028$	$.827 \pm .056$	$.755 \pm .079$	$.110 \pm .016$	$.064 \pm .019$	5239
	OFIGM	$\mathbf{.766} \pm .031$	$.825 \pm .058$	$.794 \pm .033$	$.128 \pm .044$	$\mathbf{.058} \pm .023$	3289

5.4 Implementation details

The default instance budget is $B = 512$, unknown buffer $B_u = 256$, embedding width $d = 128$, two bipartite passes, AdamW with learning rate 10^{-3}, and at most one gradient update per arriving micro-batch of 16 samples. Novelty weights $(\lambda_d, \lambda_h, \lambda_g)$ are initialized uniformly and calibrated on known-only warm-up data. The novelty threshold is the 0.995 quantile of trusted-known scores with exponential forgetting. Hyperparameters are selected using only warm-up/validation segments that precede novel-class exposure; no future novel labels are used for tuning.

6 Experimental Results and Analysis

6.1 Datasets

Three benchmark datasets including Digits, Wine, and Iris are used in the experiments. We use three matched seeds, test-then-update evaluation, 20% post-prediction labels, delayed labels for emerging classes, and feature arrival on Digits. Baselines are a cosine-prototype learner (Proto) and bounded-window kNN; OFIGM-noFG removes graph memory and the graph-residual novelty signal while retaining the class-growth wrapper.

The graph-aware OFIGM variant raises AUROC relative to OFIGM-noFG on Wine (.946 vs. .926) and Iris (.794 vs. .755), while novel-class F1 remains close; however, forgetting is not consistently reduced and Digits shows no AUROC gain over the no-graph ablation.

6.2 State-of-the-art comparison

Across the three public benchmark datasets, both **OFIGM-noFG** and **OFIGM** demonstrate competitive performance for online class-incremental learning, while the contribution of the feature-instance graph varies across datasets. On Digits, OFIGM-noFG performs slightly better than the full OFIGM model, achieving higher Macro-F1 (0.820 vs. 0.817), Novel-F1 (0.859 vs. 0.857), and AUROC (0.861 vs. 0.857), together with lower forgetting (0.041 vs. 0.045). In contrast, the graph-memory component provides clearer benefits on Wine and Iris. On Wine, OFIGM improves novelty AUROC from 0.926 to 0.946 and slightly increases Novel-F1 from 0.964 to 0.965, while also reducing forgetting from 0.069 to 0.061. On Iris, OFIGM raises AUROC substantially from 0.755 to 0.794 and slightly improves Macro-F1 from 0.765 to 0.766, although Novel-F1 decreases marginally from 0.827 to 0.825. These results suggest that the persistent feature-instance graph is particularly useful for improving *known-versus-novel discrimination* on datasets where local feature relationships provide additional structural evidence beyond prototype distance alone.

At the same time, the results reveal an important stability-efficiency trade-off. OFIGM does not consistently reduce catastrophic forgetting: although it improves forgetting on Wine, forgetting increases from 0.110 to 0.128 on Iris and from 0.041 to 0.045 on Digits relative to OFIGM-noFG. The graph-based model also incurs additional computational cost, processing 1,751, 3,048, and 3,289 examples/s on Digits, Wine, and Iris, respectively, compared with 2,905, 4,797, and 5,239 examples/s for OFIGM-noFG. Nevertheless, OFIGM maintains comparable Macro-F1 and Novel-F1 while providing notable AUROC gains on two of the three datasets and reducing the false-novel rate on Iris (0.064 to 0.058). Overall, the experiments indicate that feature-instance graph memory contributes most strongly to **novelty detection rather than conventional classification accuracy**, but its additional structural adaptation does not yet guarantee improved retention. This indicates a more explicitly stability-aware graph update mechanism that preserves the observed AUROC gains while reducing forgetting and computational overhead.

6.3 External evaluation

On the external evaluation across 11 Louisiana shellfish norovirus/outbreak episodes (Appendix D), **OFIGM-noFG** and **OFIGM** show similar overall behavior, with the full graph-memory model providing modest gains after outbreak confirmation but weaker early-warning discrimination. OFIGM achieves the highest post-confirmation Macro-F1 (0.474 ± 0.186) and outbreak recall (0.787 ± 0.169), slightly exceeding OFIGM-noFG (0.470 ± 0.183 and 0.786 ± 0.170, respectively), while its outbreak-F1 is also marginally higher (0.400 vs. 0.396). However, OFIGM-noFG performs better during the 30-day pre-confirmation risk period, with higher Early AUROC (0.585 vs. 0.577), Early AUPRC (0.252 vs. 0.215), and TPR@5%FPR (0.127 vs. 0.099). Overall, these results suggest that the feature-instance graph memory contributes more to *post-confirmation outbreak-class adaptation and recall* than to pre-outbreak early-warning discrimination, indicating that stronger temporal and spatial modeling may be needed to fully exploit environmental signals before outbreak confirmation.

7 Conclusion

In summary, we proposed **OFIGM**, an Online Feature-Instance Graph Memory framework for open-world class-incremental learning under evolving feature spaces, delayed supervision, and emerging classes. OFIGM integrates a bounded dynamic feature-instance graph, graph-based representation learning, slowly updated class prototypes, graph-aware novelty detection, and persistence-based class creation within a unified test-then-update pipeline. Experiments on public benchmark datasets show that the feature-instance graph can improve novel-class discrimination, particularly AUROC on Wine and Iris, while maintaining competitive Macro-F1 and Novel-F1, although gains in forgetting and computational efficiency remain inconsistent. External evaluation on 11 Louisiana shellfish norovirus/outbreak episodes further demonstrates that OFIGM achieves the strongest post-confirmation Macro-F1 and outbreak recall among the evaluated variants.

AI USE STATEMENT

Large language models (LLMs), specifically OpenAI's ChatGPT, were used to assist in paper preparation. Their role was limited to language refinement, LaTeX formatting, and generating alternative phrasings of author-written content. All scientific ideas, experimental design, analysis, and claims were conceived, implemented, and verified by the authors. LLM outputs were carefully reviewed and edited for accuracy. The authors take full responsibility for the content of this paper.

REPRODUCIBILITY STATEMENT

Section 5 specifies datasets, stream construction, metrics, statistical testing, memory constraints, and default hyperparameters. Algorithm 1 gives the online update logic; Appendix A provides protocol details and Appendix E provides implementation guidance.

REFERENCES

John Bako and Jugal Kalita. Addressing catastrophic forgetting in class-incremental learning—a survey. *Artificial Intelligence Review*, 2026.

Kaidi Cao, Maria Brbic, and Jure Leskovec. Open-world semi-supervised learning. In *International Conference on Learning Representations*, 2022.

Kaile Du, Yifan Zhou, Fan Lyu, Yuyang Li, Junzhou Xie, Yixi Shen, Fuyuan Hu, and Guangcan Liu. Rebalancing multi-label class-incremental learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 16372–16380, 2025. doi: 10.1609/aaai.v39i15.33798.

Guifang Fu, Garrett Saunders, and John Stevens. Holm multiple correction for large-scale gene-shape association mapping. *BMC Genetics*, 15(Suppl 1):S5, 2014.

Huan He, Owen Queen, Teddy Koker, Consuelo Cuevas, Theodoros Tsiligkaridis, and Marinka Zitnik. Domain adaptation for time series under feature and label shifts. In *International Conference on Machine Learning*, pp. 12746–12774. Pmlr, 2023.

Chenping Hou, Shilin Gu, Chao Xu, and Yuhua Qian. Incremental learning for simultaneous augmentation of feature and class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12):14789–14806, 2023. doi: 10.1109/TPAMI.2023.3307670.

Chiao-Ya Hsu and Cheng-Te Li. Gtab: Gradient boosting bipartite graph neural networks for holistic tabular data predictions. *IEEE Transactions on Knowledge and Data Engineering*, 38(6): 3669–3683, 2026. doi: 10.1109/TKDE.2026.3678153.

Kai Hu, Yunjiang Wang, Yuan Zhang, and Xieping Gao. Progressive learning strategy for few-shot class-incremental learning. *IEEE Transactions on Cybernetics*, 55(3):1210–1223, 2025. doi: 10.1109/TCYB.2025.3525724.

Yucheng Jin, Yun Xiong, Juncheng Fang, Xixi Wu, Dongxiao He, Xing Jia, Bingchen Zhao, and Philip S. Yu. Beyond the known: Novel class discovery for open-world graph learning. In *Database Systems for Advanced Applications*, pp. 117–133. Springer, 2024. doi: 10.1007/978-981-97-5572-1_8.

Gyuhak Kim, Changnan Xiao, Tatsuya Konishi, Zixuan Ke, and Bing Liu. Open-world continual learning: Unifying novelty detection and continual learning. *Artificial Intelligence*, 338:104237, 2025. doi: 10.1016/j.artint.2024.104237.

Guannan Lai, Yujie Li, Xiangkun Wang, Junbo Zhang, Tianrui Li, and Xin Yang. Order-robust class incremental learning: Graph-driven dynamic similarity grouping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4894–4904, 2025.

Jin Li, Kleanthis Malialis, and Marios M. Polycarpou. Resilient class-incremental learning: On the interplay of drifting, unlabeled and imbalanced data streams. *Artificial Intelligence Science and Engineering*, 2(1):49–65, 2026a. doi: 10.23919/AISE.2026.000004.

Shuo Li, Fang Liu, Licheng Jiao, Lingling Li, Puhua Chen, Xu Liu, and Wenping Ma. Prompt-based concept learning for few-shot class-incremental learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(5):4991–5005, 2025. doi: 10.1109/TCSVT.2025.3525545.

Yanchao Li, Hongwei Dou, Guanxiao Li, Guangwei Gao, and Huiyu Zhou. Insertion: From traditional incremental learning to open-world stream learning. *Pattern Recognition*, 176:113163, 2026b. doi: 10.1016/j.patcog.2026.113163.

Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2018. doi: 10.1109/TPAMI.2017.2773081.

Guoqiang Liang, Zhaojie Chen, Shibin Su, Shizhou Zhang, and Yanning Zhang. A masking, linkage and guidance framework for online class incremental learning. *Pattern Recognition*, 160:111185, 2025. doi: 10.1016/j.patcog.2024.111185.

Jiaming Liu, Yangqiming Wang, Tongze Zhang, Yulu Fan, Qinli Yang, and Junming Shao. Open-world semi-supervised novel class discovery. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 4002–4010, 2023a. doi: 10.24963/ijcai.2023/445.

Wenzhuo Liu, Fei Zhu, and Cheng-Lin Liu. Towards non-exemplar semi-supervised class-incremental learning. *arXiv preprint arXiv:2403.18291*, 2024.

Yaoyao Liu, Yingying Li, Bernt Schiele, and Qianru Sun. Online hyperparameter optimization for class-incremental learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 8906–8913, 2023b. doi: 10.1609/aaai.v37i7.26070.

Nachaat Mohamed. Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms. *Knowledge and Information Systems*, 67(8): 6969–7055, 2025.

Xin Mu, Kai Ming Ting, and Zhi-Hua Zhou. Classification under streaming emerging new classes: A solution using completely-random trees. *IEEE Transactions on Knowledge and Data Engineering*, 29(8):1605–1618, 2017a.

Xin Mu, Feida Zhu, Jing Du, Ee-Peng Lim, and Zhi-Hua Zhou. Streaming classification with emerging new class by class matrix sketching. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, pp. 2373–2379, 2017b. doi: 10.1609/aaai.v31i1.10842.

Christine Musanase, Anthony Vodacek, Damien Hanyurwimfura, Alfred Uwitonze, and Innocent Kabandana. Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices. *Agriculture*, 13(11):2141, 2023.

Mohammad Reza Nikoo, Abrar Al Aamri, Talal Etri, and Ghazi Al-Rawas. A review of machine learning, remote sensing, and statistical methods for reservoir water quality assessment. *Journal of Hydrology*, 659:133323, 2025.

Soohyun Park, Seok Bin Son, Soyi Jung, and Joongheon Kim. Dynamic quantum federated learning for uav-based autonomous surveillance. *IEEE Transactions on Vehicular Technology*, 74(5): 8158–8170, 2025.

Zhuang Qi, Ying-Peng Tang, Lei Meng, Han Yu, Xiaoxiao Li, and Xiangxu Meng. Class-wise balancing data replay for federated class-incremental learning. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-4863.

Yu-Yang Qian, Yong Bai, Zhen-Yu Zhang, Peng Zhao, and Zhi-Hua Zhou. Handling new class in online label shift. *IEEE Transactions on Knowledge and Data Engineering*, 37(9):5257–5270, 2025. doi: 10.1109/TKDE.2025.3583138.

Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H. Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2001–2010, 2017.

Dalwinder Singh and Birmohan Singh. Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97:105524, 2020.

Anindya Pradipta Susanto, David Lyell, Bambang Widyantoro, Shlomo Berkovsky, and Farah Magrabi. Effects of machine learning-based clinical decision support systems on decision-making, care delivery, and patient outcomes: a scoping review. *Journal of the American Medical Informatics Association*, 30(12):2050–2063, 2023.

Gido M. van de Ven, Tinne Tuytelaars, and Andreas S. Tolias. Three types of incremental learning. *Nature Machine Intelligence*, 4:1185–1197, 2022. doi: 10.1038/s42256-022-00568-3.

Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7492–7501, 2022.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

Robert F Woolson. Wilcoxon signed-rank test. *Wiley Encyclopedia of Clinical Trials*, pp. 1–3, 2007.

Chao Xu, Xijia Tang, Hong Tao, and Chenping Hou. Incremental label distribution learning. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1679–1690, 2025. doi: 10.1145/3690624.3709318.

Jianjun Zhang, Ting Wang, Wing W. Y. Ng, and Witold Pedrycz. Knnens: A k-nearest neighbor ensemble-based method for incremental learning under data stream with emerging new classes. *IEEE Transactions on Neural Networks and Learning Systems*, 34:9520–9527, 2023. doi: 10.1109/TNNLS.2022.3149991.

Bowen Zheng, Da-Wei Zhou, Han-Jia Ye, and De-Chuan Zhan. Task-agnostic guided feature expansion for class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10099–10109, 2025.

Da-Wei Zhou, Qi-Wei Wang, Zhi-Hong Qi, Han-Jia Ye, De-Chuan Zhan, and Ziwei Liu. Class-incremental learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9851–9873, 2024. doi: 10.1109/TPAMI.2024.3429383.

Yong-Nan Zhu and Yu-Feng Li. Semi-supervised streaming learning with emerging new labels. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 7015–7022, 2020. doi: 10.1609/aaai.v34i04.6186.

Huiping Zhuang, Run He, Kai Tong, Ziqian Zeng, Cen Chen, and Zhiping Lin. Ds-al: A dual-stream analytic learning for exemplar-free class-incremental learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 17237–17244, 2024. doi: 10.1609/aaai.v38i15.29670.