RAN–IDS: Adaptive Continual Learning for Intrusion Detection in Open RAN

Nadia Niknami* and Jie Wu†
*Computer Science Department, Villanova University, USA
†Department of Computer and Information Sciences, Temple University, USA

***Abstract*—Open Radio Access Networks (O-RAN) introduce dynamic control-plane policies, slice reconfigurations, and multi-vendor functional splits that create continual shifts in network behavior. Static or centralized intrusion detection systems (IDS) cannot adapt to this drift and typically suffer from catastrophic forgetting when updated in real time. To address these limitations, we propose an adaptive IDS that aligns with the O-RAN architecture through a hierarchical teacher–student design and slice-aware Continual Learning (CL). A high-capacity teacher model operating at the non-RT RIC distills knowledge to lightweight student models deployed at the near-RT RIC and DU–CU interfaces, enabling resource-efficient, real-time detection across heterogeneous O-RAN nodes. We further incorporate federated CL (FCL) to allow collaborative model adaptation without sharing raw traffic. Experimental results show that our approach maintains high accuracy under evolving traffic distributions, preserves past knowledge, and significantly improves detection of unseen attacks while operating efficiently within realistic O-RAN constraints.**

***Index Terms*—O-RAN Security, Intrusion Detection System (IDS), Continual Learning, Concept Drift, Novelty Detection.**

I. Introduction

Open Radio Access Networks (O-RAN)[1] disaggregate the Radio Unit (RU), Distributed Unit (DU), and Centralized Unit (CU) and introduce multi-vendor, software-defined, and cloud-native operation. While this architecture improves flexibility and innovation, it also creates a highly dynamic and heterogeneous attack surface. Frequent RAN Intelligent Controller (RIC) policy changes, DU–CU functional updates, slice reconfigurations, and vendor-specific variations continuously shift control and user-plane behaviors, rendering static Intrusion Detection System (IDS) [2], [3] ineffective. These evolving conditions introduce significant challenges for IDS design. Effective IDS deployment in O-RAN must therefore adapt to distributed data sources and varying levels of granularity across interfaces. The near-RT RIC provides a rich stream of control-plane signaling and policy updates, making it a prime location for rapid detection of anomalies. The DU–CU interface offers visibility into fronthaul and midhaul traffic, where changes in baseband processing and radio scheduling introduce further drift. These diverse monitoring points require an IDS capable of continuous, real-time adaptation.

Existing O-RAN IDS approaches lack mechanisms for continual adaptation and do not address drift across heterogeneous RAN interfaces, nor do they leverage hierarchical knowledge transfer. To address these challenges, we propose RAN–IDS, an adaptive intrusion detection framework that integrates Continual Learning (CL)[4] and hierarchical knowledge transfer tailored to O-RAN. Our approach integrates unsupervised novelty detection with CL to handle drift without catastrophic forgetting. This hybrid strategy achieves three goals:

1) Detect novel and evolving threats across heterogeneous O-RAN interfaces
2) Preserve historical knowledge through CL
3) Adapt efficiently to real-time changes in traffic, slices, and control-plane behavior

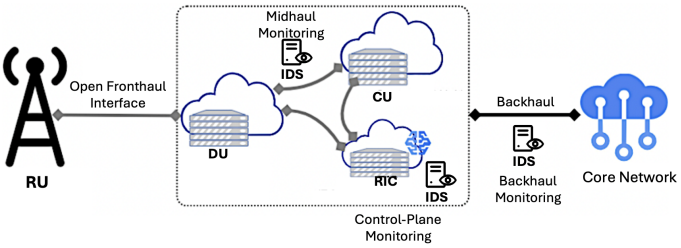

Fig. 1. High-level IDS placement across RIC, midhaul, and backhaul in the O-RAN architecture.

Fig. 1 illustrates where IDS should be positioned within the O-RAN architecture to capture the distinct forms of drift introduced by functional splits and control-plane changes. The RU, DU, CU, and RIC exchange traffic across fronthaul, midhaul, and backhaul interfaces, each carrying different signaling and data patterns. Because these interfaces experience drift for different reasons (such as RIC policy updates, DU–CU functional changes, and slice reconfigurations) RAN–IDS places lightweight monitors at three strategic locations: (i) near the near-RT RIC for control-plane supervision, (ii) along the DU–CU midhaul for data-plane visibility, and (iii) at the backhaul to inspect traffic heading toward the core network. This placement provides complementary visibility across O-RAN components and forms the foundation for the adaptive mechanisms developed in the remainder of the paper.

In particular, slice-aware CL allows the IDS to adapt to drift arising from slice scaling, reassignment, and changing service requirements. Unlike prior O-RAN IDS solutions, RAN–IDS jointly supports continual adaptation, interface-specific detection, and hierarchical long-term memory through a unified teacher–student design.

The key contributions of this paper are as follows:

979-8-3315-7111-5/26/$31.00 ©2026 IEEE

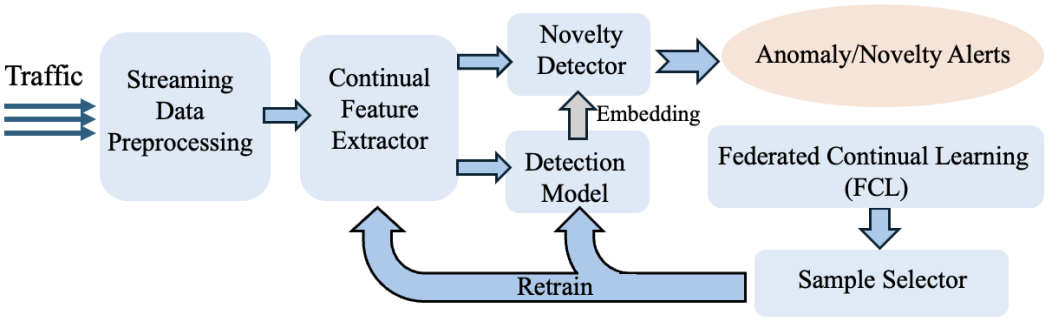

Fig. 2. High-level architecture of the proposed adaptive RAN–IDS.

- We introduce a hierarchical teacher–student IDS architecture tailored to the O-RAN functional splits (RIC, DU, CU), enabling resource-efficient, interface-aware IDS.
- We design a slice-aware CL mechanism that adapts to control-plane and user-plane drift caused by RIC policies, slicing, and vendor reconfigurations.
- We integrate federated CL across O-RAN nodes, enabling collaborative adaptation without sharing raw traffic data.
- We demonstrate that these mechanisms jointly mitigate catastrophic forgetting, enable zero-day detection in drifting environments, and significantly outperform conventional CL baselines

II. Proposed Framework

The proposed RAN–IDS framework is organized into two coordinated layers. The first layer consists of the *local student pipeline*, deployed at near-RT O-RAN components such as the RIC and DU–CU interfaces. This pipeline performs real-time IDS using CL and novelty detection [5] modules that operate directly on streaming traffic. The second layer consists of a *global teacher model* located at the non-RT RIC. The teacher maintains a higher-capacity representation of long-term network behavior and periodically distills updated knowledge to the lightweight student models. This two-layer design ensures that RAN–IDS can provide fast, resource-efficient inference at distributed nodes while benefiting from global, drift-aware adaptation without requiring raw traffic sharing. The remainder of this section describes the components of the local student pipeline and explains how the teacher layer supports global adaptation through Federated CL (FCL).

The framework uses interconnected modules that form a continuous pipeline to detect anomalies and adapt to drift in dynamic O-RAN environments. Fig. 2 presents the overall RAN–IDS architecture. The pipeline begins with raw traffic, which is processed by the *Streaming Data Preprocessing* module. This component organizes incoming packets or flows into a continuous stream suitable for incremental learning, performing basic cleaning, normalization, and temporal buffering without relying on predefined training or testing splits. Next, the data is forwarded to the *Continual Feature Extractor*, which produces compact representations suitable for both classification and anomaly detection. These representations are provided to two downstream modules: the *Detection Model* and the *Novelty Detector*.

To enable adaptation over time, the framework incorporates FCL, which aggregates model updates from distributed O-RAN nodes such as the DU, CU, and RIC without sharing raw traffic data. This is essential in O-RAN environments where data is vendor-specific, privacy-sensitive, and highly heterogeneous. Each node contributes updates that reflect local drift including changes driven by slicing, scheduling, or configuration, while maintaining global consistency across the network. FCL outputs are used by the *Sample Selector*, which builds representative replay batches by combining new data with strategically selected historical samples. These batches are passed as a retrain to the Detection Model, enabling the detection model to incrementally adapt while mitigating catastrophic forgetting. This architecture provides scalable, privacy-preserving, and drift-aware IDS tailored to the distributed, multi-vendor, and rapidly evolving nature of O-RAN networks.

A. Architecture and Deployment

The IDS integrates naturally into key O-RAN components where both control-plane and user-plane behavior can be monitored:

- Near-RT RIC: Monitors control-plane signaling, policy updates, and RAN management functions. CL operates here to adapt detection models in real time using federated inputs from distributed O-RAN nodes.
- DU–CU Midhaul / Transport Interface: Observes data-plane traffic and radio-related behaviors subject to rapid drift. Implements replay-based CL to retain historical threat patterns and provides feedback to improve global model adaptation within the FCL process.

In RAN–IDS, multiple components operate in a coordinated pipeline to ensure adaptive, accurate, and drift-resilient IDS. Each module plays a distinct role, contributing to both internal learning processes and external anomaly detection. The key components are outlined below:

- Continual Feature Extractor: This module converts raw streaming traffic into compact flow-level representations that support both detection and CL. It adapts over time as traffic patterns evolve across different O-RAN interfaces and slices.

- Detection Model: The detection model captures previously learned behaviors and produces predictions based on historical patterns. It operates primarily as an internal component, generating latent embeddings and confidence scores that are consumed by the novelty detector. The model is incrementally updated through replay-driven retraining to preserve stability while accommodating new behaviors.
- Novelty Detector: This module identifies traffic instances that deviate significantly from known patterns. It analyzes embeddings or confidence distributions produced by the detection model to detect unseen or zero-day behaviors. The novelty detector generates the system's primary external alerts and is responsible for identifying slice-specific drift as slice configurations and traffic characteristics change over time.

In summary, the detection model preserves prior knowledge and maintains stable classification, while the novelty detector identifies emerging threats that deviate from previously learned patterns. Together, they enable the IDS to detect both known and unseen attacks under the evolving conditions of O-RAN environments.

B. Feature Extractor

Raw traffic is processed by a lightweight feature extraction module that derives flow-level metadata using early packet headers. This approach ensures low computational cost, compatibility with encrypted traffic, and efficient support for continual adaptation.

C. Novelty Detector

The novelty detector evaluates deviations between current samples and historical behavior using internal outputs from the detection model, including:

- Latent feature embeddings,
- Confidence or probability distributions, and
- Residual-based measures (e.g., reconstruction error in autoencoder variants).

These signals allow the detector to identify zero-day attacks and concept drift without requiring labeled attack data. Novelty detection is performed per slice and per interface, enabling sensitivity to slice lifecycle events such as scaling, reassignment, or reconfiguration.

D. Sample Selector and Feedback Loop

Representative samples—derived from incoming traffic, federated updates, or confirmed false alarms—are maintained in a replay buffer using a strategic sample selection process. This buffer supports periodic retraining of the detection model and feature extractor, enabling the system to learn legitimate new behaviors while minimizing catastrophic forgetting.

E. Hierarchical Adaptive IDS for O-RAN

To support distributed deployment across the O-RAN architecture, we extend the continual-learning IDS into a hierarchical teacher–student configuration. A high-capacity teacher model running at the non-RT RIC aggregates global knowledge and adapts to long-term drift. Lightweight student models deployed at the near-RT RIC and DU–CU interfaces perform real-time inference under strict latency and resource constraints. The teacher periodically distills updated knowledge to these students, while FCL aggregates updates from distributed nodes, enabling privacy-preserving, resource-aware, and drift-resilient IDS across the RAN.

III. Evaluation

The experiments are conducted using the 5G-NIDD dataset [6], which contains benign and malicious traffic generated by real mobile devices connected to the 5G Test Network (5GTN) using two base stations. The malicious traffic corresponds to multiple DDoS attack scenarios. Following the CL setting, the dataset is partitioned into nine sequential tasks, each representing one attack session. During evaluation, the IDS receives traffic incrementally in a streaming manner without access to ground-truth labels. After completing each task, the model is updated using the replay-driven CL procedure described in Section II. Performance is measured across all tasks to assess how well the IDS adapts to drift while retaining previously acquired knowledge.

We compare the proposed approach against five well-established CL methods:

- AOC-IDS [7]: An IDS-specific CL framework that augments its memory buffer with pseudo-labeled samples and applies label perturbation to enhance robustness.
- Experience Replay (ER) [8]: Maintains a replay buffer using reservoir sampling and mixes past samples with new data during training to reduce forgetting.
- Averaged Gradient Episodic Memory (AGEM) [9]: Stores gradient information from previous tasks and projects new gradients to prevent interference with past knowledge.
- Learning without Forgetting (LwF) [10]: Uses knowledge distillation to preserve the output behavior on earlier tasks while learning new ones.
- Elastic Weight Consolidation (EWC) [11]: Estimates parameter importance via the Fisher Information Matrix and restricts updates to weights critical for previous tasks.

We additionally include a static non-CL model that is trained once and never updated. This baseline quantifies the degradation caused by concept drift and highlights the need for continual adaptation.

To quantify model performance, we report the F1 score due to the imbalance between benign and attack flows. Following standard CL practice [12], [5], we construct an evaluation matrix R_{ij}, where R_{ij} is the F1 score achieved when the model is trained up to task i and evaluated on task j.

Three metrics are derived from this matrix:

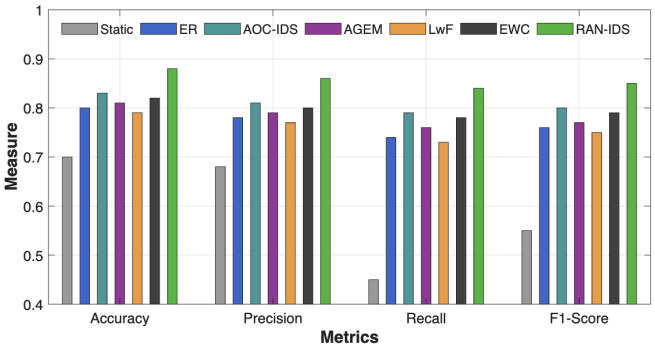

Fig. 3. Overall Performance on 5G-NIDD.

a) Average Performance (AVG):

$$\text{AVG}_{F_1} = \frac{1}{m}\sum_{i=1}^{m} R_{ii}, \tag{1}$$

which measures performance on the current task immediately after it is learned.

b) Forward Transfer (FwdTrans):

$$\text{FwdTrans}_{F_1} = \frac{\sum_{i<j} R_{ij}}{m(m-1)/2}, \tag{2}$$

which quantifies the model's ability to generalize to unseen (future) tasks and reflects the capability to detect zero-day attacks.

c) Backward Transfer (BwdTrans):

$$\text{BwdTrans}_{F_1} = \frac{\sum_{i=1}^{m-1}(R_{mi} - R_{ii})}{m(m-1)/2}, \tag{3}$$

where R_{mi} is the performance on a past task i after the final task is learned. Positive values indicate knowledge retention or backward improvement, whereas negative values indicate catastrophic forgetting. Overall, AVG measures performance on seen attacks, FwdTrans captures zero-day generalization, and BwdTrans evaluates forgetting. Higher values across all metrics indicate stronger CL performance.

A. Comparative Results on 5G–NIDD

We compare RAN–IDS with five CL baselines along with a static model trained only once. All methods are evaluated on the same sequence of nine attack sessions from 5G-NIDD, where the static model serves as a non-adaptive reference to illustrate concept drift. Fig. 3 summarizes accuracy, precision, recall, and F1 score. RAN–IDS consistently outperforms all baselines across every metric. The static model and ER show moderate accuracy but poor recall, indicating missed attacks as drift increases. Regularization-based approaches (LwF, EWC, AGEM) achieve more balanced results but still degrade over time because they cannot fully retain prior knowledge while adapting to new traffic patterns. AOC-IDS performs strongly but remains limited by its single-model design and lack of hierarchical adaptation.

In contrast, RAN–IDS achieves the highest accuracy (0.88), precision (0.86), recall (0.84), and F1 (0.85), driven by its continual-learning pipeline, slice-aware adaptation, and teacher–student architecture. These features allow RAN–IDS to maintain robust detection under dynamic O-RAN conditions, making it the most reliable and adaptive IDS among the evaluated methods.

TABLE I
ZERO–DAY DETECTION PERFORMANCE ON 5G–NIDD

Method	F1 (%)	Recall (%)	Precision (%)
PCA	58	50	68
OC–SVM	57	46	74
Isolation Forest	63	62	63
Autoencoder (AE)	67	64	71
RAN–IDS (Ours)	**80**	**77**	**83**

B. Zero–Day Detection Performance

To evaluate zero-day performance, we compare RAN–IDS against four widely used novelty-detection baselines: Principal Component Analysis (PCA), One-Class SVM (OC–SVM), Isolation Forest (IF), and Autoencoders (AE). All baselines are trained once on initial benign data and tested on later tasks without adaptation, measuring pure generalization in the absence of continual learning.

Table I shows distinct differences between classical novelty detectors and RAN–IDS. PCA and OC–SVM exhibit low F1 scores (57–58%) due to poor recall, indicating limited ability to identify unseen attacks. Isolation Forest performs moderately better, but its precision and recall remain insufficient for reliable detection. Although Autoencoders yield the highest baseline performance (67% F1), they still fail to generalize effectively to attack patterns not represented in the initial training data. In contrast, RAN–IDS achieves 80% F1, 77% recall, and 83% precision, substantially outperforming all baselines. These gains stem from its continual-learning pipeline and adaptive novelty-detection mechanisms, which allow it to incorporate new behaviors while avoiding catastrophic forgetting. Overall, the results demonstrate that classical detectors cannot handle the evolving nature of O-RAN traffic, whereas RAN–IDS offers robust and reliable zero-day threat detection.

C. Ablation Study

To quantify the contribution of each component in RAN–IDS, we conduct a series of ablation experiments. Table II summarizes the high-level impact of CL mechanisms. Removing CL ("w/o CL") leads to a substantial drop in AVG (from 82.4% to 68.7%) and FwdTrans (73.1% to 58.4%), indicating that the model cannot adapt to new attack patterns when trained in a static fashion. Naïve fine-tuning achieves slightly better performance on new tasks (60.3% FwdTrans), but exhibits the most severe catastrophic forgetting, as shown by its highly negative BwdTrans (–18.9). This confirms that incremental training without memory or consolidation rapidly erodes previously learned knowledge.

TABLE II
ABLATION STUDY: IMPACT OF CONTINUAL LEARNING MECHANISMS

Method	AVG (%)	FwdTrans (%)	BwdTrans (%)
RAN–IDS (Full)	**82.4**	**73.1**	**+1.8**
w/o CL	68.7	58.4	–12.6
Naïve Fine-Tuning	72.1	60.3	–18.9

TABLE III
ABLATION STUDY: IMPACT OF REPLAY AND SAMPLE SELECTION

Method	AVG (%)	FwdTrans (%)	BwdTrans (%)
RAN–IDS (Full)	**82.4**	**73.1**	**+1.8**
w/o Sample Selector	78.9	67.8	+0.4
w/o Replay	74.2	63.5	–9.7

Table III further investigates the internal components of RAN–IDS. Removing the replay buffer ("w/o Replay") significantly reduces AVG (82.4% to 74.2%) and produces a strongly negative BwdTrans (–9.7), demonstrating that replay is essential for retaining past patterns. Disabling the sample selector ("w/o Sample Selector") also degrades performance, especially FwdTrans (73.1% to 67.8%), showing that selecting representative samples improves adaptation to new traffic distributions. Overall, the full RAN–IDS architecture consistently achieves the highest AVG, FwdTrans, and BwdTrans, confirming that replay, sample-aware selection, and continual adaptation work jointly to provide stable and robust intrusion detection in dynamic O–RAN environments.

D. Task-wise F1 Performance across Sequential Tasks

Task-wise continual learning performance is illustrated in Fig. 4, which reports the F1 score of all methods across the nine sequential 5G–NIDD tasks. The static model rapidly collapses after the first two tasks, reflecting its inability to accommodate distribution shift. ER slows this degradation through replay, but still exhibits a steady decline as new tasks overwrite earlier knowledge. Regularization-based approaches such as AGEM, LwF, and EWC demonstrate more stable trajectories, yet all show progressive forgetting as the task sequence introduces drift that cannot be fully mitigated by parameter constraints alone. AOC-IDS performs better than the standard CL baselines, preserving higher F1 scores on later tasks due to its IDS-oriented design, but it still suffers from measurable performance erosion. In contrast, RAN–IDS maintains consistently high F1 scores throughout all nine tasks, showing only minimal decline as the task sequence progresses. This stability highlights the strength of its continual-learning pipeline, replay-driven consolidation, and sample-aware adaptation, enabling the model to incorporate new attack patterns without compromising previously learned behaviors. Overall, the figure demonstrates that RAN–IDS delivers the most robust and drift-resistant performance among all evaluated methods.

IV. CONCLUSION

This paper presented RAN–IDS, an adaptive intrusion detection framework designed for the dynamic and distributed

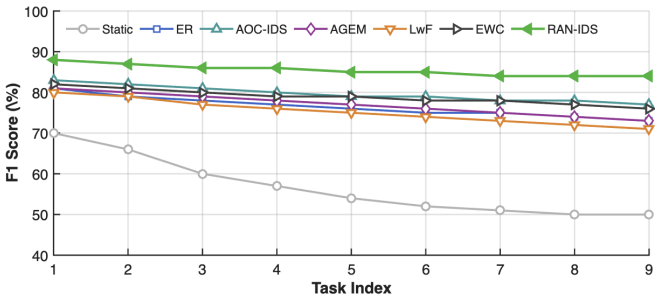

Fig. 4. Task-wise F1 performance across the nine sequential 5G–NIDD tasks for all evaluated methods.

characteristics of O–RAN. By integrating CL, replay-based consolidation, and lightweight novelty detection, RAN–IDS effectively adapts to concept drift while retaining knowledge of previously learned behaviors. Experiments on the 5G–NIDD dataset show that it outperforms established continual-learning baselines and classical novelty detectors, while maintaining low overhead suitable for near–real-time deployment.

As future work, we aim to extend RAN–IDS with a closed-loop mitigation module that automatically adjusts RIC policies or slice configurations based on detected threats

REFERENCES

[1] M. Liyanage, A. Braeken, S. Shahabuddin, and P. Ranaweera, "Open ran security: Challenges and opportunities," *Journal of Network and Computer Applications*, vol. 214, p. 103621, 2023.
[2] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac, and P. Faruki, "Network intrusion detection for iot security based on learning techniques," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2671–2701, 2019.
[3] S. Walling and S. Lodh, "A survey on intrusion detection systems: Types, datasets, machine learning methods for nids and challenges," in *13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2022, pp. 1–7.
[4] D. Rolnick, A. Ahuja, J. Schwarz, T. Lillicrap, and G. Wayne, "Experience replay for continual learning," *Advances in neural information processing systems*, vol. 32, 2019.
[5] S. Fuhrman, O. Gungor, and T. Rosing, "Cnd-ids: Continual novelty detection for intrusion detection systems," *arXiv preprint arXiv:2502.14094*, 2025.
[6] S. Samarakoon, Y. Siriwardhana, P. Porambage, M. Liyanage, S.-Y. Chang, J. Kim, J. Kim, and M. Ylianttila, "5g-nidd: A comprehensive network intrusion detection dataset generated over 5g wireless network," *arXiv preprint arXiv:2212.01298*, 2022.
[7] X. Zhang, R. Zhao, Z. Jiang, Z. Sun, Y. Ding, E. C. Ngai, and S.-H. Yang, "Aoc-ids: Autonomous online framework with contrastive learning for intrusion detection," in *IEEE INFOCOM-IEEE Conference on Computer Communications*, 2024, pp. 581–590.
[8] R. Aljundi, M. Lin, B. Goujaud, and Y. Bengio, "Online continual learning with no task boundaries," *arXiv preprint arXiv:1903.08671*, vol. 3, p. 2, 2019.
[9] A. Chaudhry, M. Ranzato, M. Rohrbach, and M. Elhoseiny, "Efficient lifelong learning with a-gem," *arXiv preprint arXiv:1812.00420*, 2018.
[10] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2935–2947, 2017.
[11] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the national academy of sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.
[12] N. Díaz-Rodríguez, V. Lomonaco, D. Filliat, and D. Maltoni, "Don't forget, there is more than forgetting: new metrics for continual learning," *arXiv preprint arXiv:1810.13166*, 2018.