UNIVERSITY OF WEST ATTICA

eISSN:3136-7658

Applied Manufacturing Intelligence Journal

Vol. 5, No. 3, July, 2026



Distributed Edge Model Update Scheduling for SwinV2 Adapter Models Based on Wireless Bandwidth and Drift Indicators

Tadas Mikalauskas[1,*], Deividas Remeika[1], Orestas Grigon[1]

[1] *Department of Computer Science, SMK College of Applied Sciences, Vilnius, LT-03107, Lithuania*

[2] *Faculty of Technical Sciences, Telšiai College, Telšiai, LT-87132, Lithuania*

Received 27 April 2026; Revised 16 August 2026; Accepted 20 August 2026; Published 23 August 2026

Abstract

Distributed visual models require timely adaptation when scene distributions change, while fluctuating wireless capacity limits the availability of local updates. This paper develops a participant scheduling framework for SwinV2 Adapter models that combines bandwidth, queue delay, feature drift, validation loss deviation, gradient informativeness, and bounded freshness credit. A coordinator ranks local adapter updates under a communication budget and aggregates the admitted parameters through utility weighted contributions. The evaluation examines accuracy and communication efficiency across 24 edge nodes, four wireless regimes, and five drift categories. In the experimental comparison, final aggregate F1 is 89.3% for the proposed scheduler and 87.7% for periodic participation. Mean completion time decreases from 18.3 to 11.2 s, while the 95th percentile is 19.6 s. The energy distributions have median values of 42.5 and 62.0 J for the proposed and periodic policies, respectively. The proposed payload is 42.8% of the periodic reference, with median staleness of 1.9 rounds and a 95th percentile of 3.8 rounds. Component ablation results separate the roles of bandwidth feasibility and statistical urgency and identify their distinct effects on update latency and adaptation accuracy. The framework provides an interpretable formulation for assessing how communication allocation affects model freshness, participant representation, and adaptation quality in distributed wireless edge systems.

Keywords: Bandwidth Drift Utility, Edge Model Updating, SwinV2 Adapter, Wireless Scheduling, Feature Distribution Shift

1. Introduction

Wireless edge vision architectures place inference near distributed cameras to meet the latency and data locality requirements of traffic monitoring, industrial inspection, and mobile perception [1]. Communication constrained learning frameworks extend this architecture to model maintenance, where uplink variability affects the arrival and usefulness of locally trained updates [2]. Hierarchical window attention provides a computationally structured representation for vision transformers operating across multiple image resolutions [3]. SwinV2 introduces mechanisms for stable scaling and resolution transfer that provide a foundation for adapting visual representations to changing deployment conditions [4]. Resource aware edge learning research identifies the interaction between uplink capacity, device energy, and local data heterogeneity as a central constraint on distributed updating [5]. Parameter efficient adaptation addresses part of this burden by training compact modules attached to a largely frozen backbone [6]. The resulting update path changes the scheduling problem: communication can occur more frequently, while the coordinator

* Corresponding author: tadas.mika@smk.lt

https://doi.org/10.68406/amij.2026.vol5iss3np5:65-78

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

must still decide which local correction should receive priority within the available airtime. A camera observing a stable scene may have sufficient bandwidth for repeated uploads, whereas another camera encountering an emerging event may experience prolonged queueing. The utility of either update depends on its timing, its representation of the current scene, and the opportunities displaced by its transmission. These dependencies become consequential when the application requires continuous service and shares the wireless medium with inference traffic. Adapter payload reduction therefore motivates a scheduling mechanism that relates communication feasibility to the changing statistical value of each node's observations.

The central engineering difficulty is that model drift and wireless availability rarely align. A camera facing a flooded street, a conveyor line with new defects, or a roadside sensor in rain may generate the most valuable update signal during periods of wireless channel congestion. Existing edge update systems often separate communication scheduling from model state assessment, using either channel quality or local loss as the primary trigger [7]. This separation can produce competing scheduling errors. A bandwidth-oriented scheduler repeatedly favors nodes with stable, high-throughput links even when their data remain redundant [8]. A drift-oriented scheduler may select nodes with severe distribution shift whose uploads exceed the aggregation deadline [9]. Federated and split learning studies have reported similar tensions between statistical heterogeneity and resource heterogeneity [10]. Adapter-based update schemes reduce the payload while retaining the need to decide whose update should be transmitted, delayed, or skipped in each round [11]. The problem is particularly relevant to edge systems operating continuously across successive deployment intervals. A scheduler that initially maintains balanced access may develop selection bias over successive operating intervals, because nodes with favorable communication conditions receive frequent updates, while nodes with constrained links accumulate staleness. The scheduling rule therefore requires both an update freshness record and a measure of current link feasibility.

This paper studies distributed edge model update scheduling for a SwinV2 Adapter model by jointly considering wireless bandwidth and drift indicators. The objective is to control the timing and source of compact adapter uploads under changing link and scene conditions. The contribution lies in coupling communication feasibility, statistical urgency, and bounded freshness within an adapter-update scheduling rule; the SwinV2 backbone and adapter concept serve as its architectural foundation. The proposed method uses a bandwidth-drift utility score that merges instantaneous uplink bandwidth, short queue delay, local feature drift, validation loss deviation, and adapter gradient informativeness. The scheduler then chooses a subset of nodes under a per-round payload budget and applies weighted adapter aggregation. The design uses scalar monitoring signals and compact statistics that edge agents derive from locally computed summaries. A second design constraint is interpretability. Deployment monitoring requires explicit reasons for node deferral during congestion and for the admission of nodes with limited bandwidth. The proposed utility score supports inspection of scheduling decisions because each term corresponds to an observable communication or model-state indicator with a defined contribution to node priority. The selected nodes are evaluated as contributors to a shared adapter state, with communication cost assessed at admission and statistical urgency assessed from recent local observations. This connection allows the scheduling objective to express both the timeliness of transmission and the relevance of the underlying scene change.

The paper develops this scheduling problem from the practical constraints of distributed camera systems through an explicit control formulation and an evaluation of its competing accuracy and communication objectives. Section 2 reviews participant selection, event triggered synchronization, representation drift indicators, and parameter efficient adaptation, identifying the unresolved connection between link availability and local update value. Section 3 describes the edge monitoring architecture and the SwinV2 Adapter update path before defining the utility score, communication budget, aggregation weights, and bounded freshness credit. This formulation makes the inputs to each scheduling decision explicit and relates them to quantities available at the node or coordinator. Section 4 organizes the evaluation around the configuration of heterogeneous nodes, the behavior of accuracy and resource metrics, and the response to component removal and changing operating conditions. The analysis distinguishes convergence from final accuracy, mean delay from its upper tail, and participant coverage from frequency-based fairness. It also examines how coefficient choice affects the relative influence of bandwidth and drift, allowing the observed tradeoffs to be traced to individual scheduling terms. Section 5 concludes the paper by presenting the scope of the proposed method, limitations of the evaluation, and research directions for physical deployments and online controller adaptation.

2. Related Work

2.1. Edge Model Update Scheduling

Edge model update scheduling has been studied through periodic synchronization, event-triggered uploading, resource-aware participant selection, and hierarchical aggregation. Periodic synchronization assigns equal priority to all nodes in each round, which can reduce communication efficiency when scene changes affect a subset of nodes [12]. Event-triggered uploading activates nodes when local signals cross a threshold, reducing communication demand; mixed wireless conditions complicate threshold calibration [13]. Resource-aware participant selection improves deadline

compliance by considering bandwidth, device availability, or energy, and it is therefore well aligned with engineering deployment constraints [14]. Hierarchical aggregation moves part of the update process closer to the edge cluster and can reduce wide-area traffic, although it may introduce extra coordination delay when cluster states differ [15]. Participant selection directly changes the statistical composition and timeliness of the model update stream. A practical scheduler operates under partial observability, using current measurements to estimate transmission feasibility and update value before local training is complete. For this reason, many deployed rules prefer scalar indicators that can be monitored with limited computational overhead and compared across nodes. Scalar indicators may express competing objectives, so prioritizing one indicator can affect performance associated with another. The proposed formulation accounts for conflicts among indicators during routine edge updating.

2.2. Drift Indicators and Adapter Based Vision Updating

Data drift in edge vision appears through illumination changes, weather, sensor pose, object density, background clutter, and label distribution movement. Drift indicators are commonly derived from feature statistics, confidence distributions, validation loss, reconstruction residuals, or disagreement among local predictors [16]. For transformer-based vision systems, intermediate feature shift is often more informative than raw pixel differences because the learned representation already filters many irrelevant variations [17]. Wang et al. [18] introduced TR-Adapter as a parameter-efficient transfer learning method for video question answering. Distributed adapter updates require careful selection because drift severity and transmission feasibility can favor different nodes within a round [19]. The research gap concerns joint evaluation of wireless bandwidth and drift information within a common scheduling criterion. Adapter tuning also changes the meaning of update value. A full-model update may contain broad representation change, while an adapter update usually carries a compact correction that is more local to the current domain. This compactness supports frequent updating and increases the cost of selecting redundant clients. A scheduler should therefore assess both deviation from the reference distribution and the potential contribution of the compact adapter step to the shared model.

3. Proposed Bandwidth Drift Scheduling Framework

The proposed framework comprises edge monitoring, participant scheduling, and adapter aggregation. Each edge node runs a SwinV2 backbone with lightweight adapters inserted after the attention and feed-forward blocks. During local adaptation, the backbone remains frozen, while recent image batches update the adapter parameters. A coordinator receives compact monitoring messages containing estimated uplink bandwidth, queue delay, feature drift, validation loss deviation, and gradient informativeness. Figure 1 illustrates the relationship between the distributed nodes, monitoring modules, participant selection mechanism, and adapter synchronization path. Raw images and labels remain at their originating nodes. This arrangement supports data locality, while privacy protection also depends on the information exposed by monitoring statistics and transmitted adapter parameters. Monitoring, local adaptation, and adapter transmission are separate operations with distinct computational and communication costs.

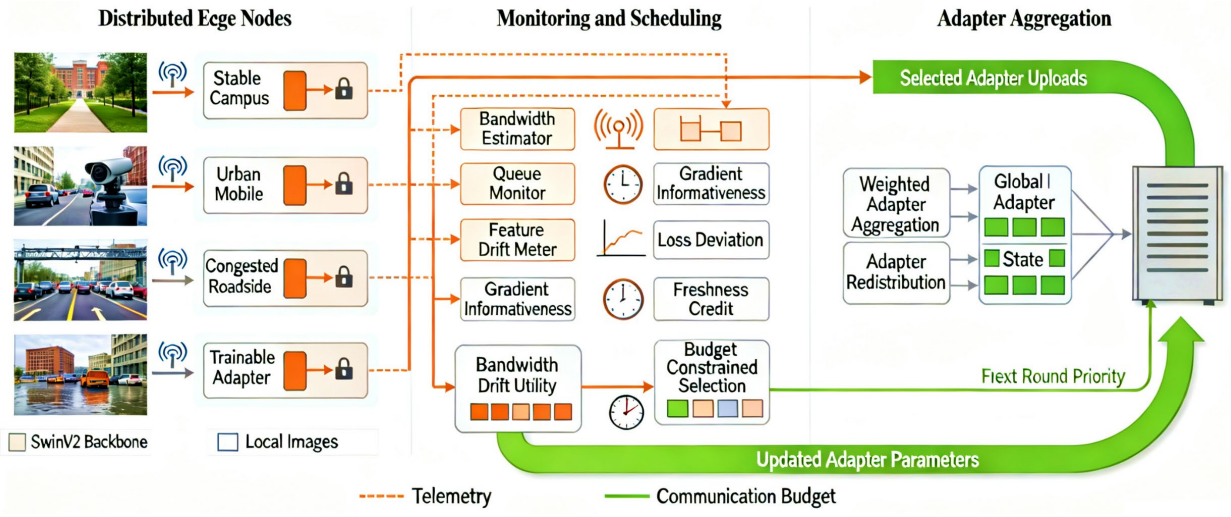


Figure 1. Distributed edge monitoring and SwinV2 Adapter update architecture

At communication round t, node i reports its current communication conditions and adaptation indicators. Bandwidth, queue delay, loss deviation, and gradient informativeness are normalized using robust running scales. Feature drift is standardized against node-specific reference statistics. These transformations allow indicators with different units to contribute to a common priority score. Their numerical ranges determine the effective influence of the

coefficients, so coefficient values must be interpreted together with the normalization scales and update windows. The utility equation combines communication feasibility, statistical urgency, and update freshness:

$$U_i(t) = \alpha B_i(t) - \beta Q_i(t) + \gamma D_i(t) + \eta L_i(t) + \rho G_i(t) + \kappa S_i(t) \tag{1}$$

Here, $B_i(t)$ denotes normalized bandwidth availability, and $Q_i(t)$ represents the queue delay penalty. The terms $D_i(t), L_i(t)$, and $G_i(t)$ describe feature drift, validation loss deviation, and gradient informativeness, respectively. The bounded credit $S_i(t)$ represents the time elapsed since the node last contributed. Nonnegative coefficients control the relative influence of these indicators and can be calibrated using representative validation traces. Bandwidth increases transmission priority, while queue pressure reduces it. The queue penalty can yield a negative utility; scheduling utility and the bounded nonnegative quantity used for aggregation therefore have different roles. The statistical terms favor updates associated with changing local observations, and the freshness term raises the priority of repeatedly deferred nodes.

Feature drift is measured using the displacement of the current feature centroid from a node-specific reference centroid. The reference statistics are maintained from a designated stable period and refreshed after an accepted model update when the local monitoring state has stabilized. For feature extractor ϕ and local batch $X_i(t)$, the following equation expresses this displacement relative to the reference feature scale:

$$D_i(t) = \frac{\left\| mean\big(\phi(X_i(t))\big) - \boldsymbol{\mu}_i^{ref} \right\|_2}{\sigma_i^{ref} + \varepsilon} \tag{2}$$

The vector $\boldsymbol{\mu}_i^{\text{ref}}$ is the reference centroid, while σ_i^{ref} is a positive scalar describing reference feature dispersion. A small constant ε ensures numerical stability. Validation loss deviation is assessed against the node's rolling loss median, which accounts for persistent differences in scene difficulty. Gradient informativeness is represented by the clipped adapter gradient norm. This quantity serves as a proxy for correction magnitude, while its relationship to generalization benefit requires separate validation. Its contribution is bounded during normalization to limit the influence of unusually large gradients. Interpreting centroid displacement also requires distinguishing changes in scene statistics from changes caused by an accepted adapter state.

Participant selection operates under a shared communication budget. Let $s_i(t)$ indicate whether node i is admitted, and let $c_i(t)$ denote its estimated adapter upload airtime, including the configured transmission overhead. For N candidate nodes, the coordinator solves:

$$\begin{aligned} &\max_{\{s_i(t)\}} \sum_{i=1}^{N} s_i(t) U_i(t) \\ &subject\ to \sum_{i=1}^{N} s_i(t) c_i(t) \le C_t, \\ &\quad s_i(t) \in \{0,1\}. \end{aligned} \tag{3}$$

In this selection equation, C_t represents the aggregate airtime available for adapter uploads during the current round. The formulation assumes a shared scheduling resource at the coordinator. A greedy approximation ranks candidates by utility per unit airtime, $U_i(t)/c_i(t)$, and admits feasible nodes while capacity remains. This ratio-based rule is a heuristic for the binary budget-constrained selection problem. Its output provides a feasible operational schedule; establishing optimality requires separate analysis. Sorting the candidate ratios requires O (N log N) comparisons, followed by a linear admission scan. Indicator estimation contributes additional computational cost. Upload costs are recalculated from the latest bandwidth and queue observations before admission, allowing participant choice to respond to changing wireless conditions. The budget is expressed in airtime; adapter payload supplies the byte count used in estimating that cost. Feasibility at admission is conditional on these estimates and on the subsequent channel state.

After admitted uploads arrive, the coordinator aggregates their adapter parameters. Each aggregation weight combines local sample count with a bounded, nonnegative transformation of scheduling utility and is normalized across the accepted participant set $\mathcal{A}_t$. The shared adapter state evolves through:

$$\boldsymbol{\theta}_A(t+1) = \boldsymbol{\theta}_A(t) + \lambda \sum_{i \subseteq \mathcal{A}_\ell} w_i(t) \big[\boldsymbol{\theta}_{A,i}(t) - \boldsymbol{\theta}_A(t) \big] \tag{4}$$

Within the aggregation equation, $\boldsymbol{\theta}_A(t)$ denotes the current shared adapter parameters, and $\boldsymbol{\theta}_{A,i}(t)$ denotes the uploaded local state. The weights $w_i(t)$ sum to unity over accepted participants. The damping coefficient λ controls how strongly the shared state moves toward the aggregated correction. This control is relevant when simultaneous scene changes produce competing local updates. Under a damping interpretation, the coefficient lies between zero and one:

zero retains the current state, and one applies the weighted accepted-state average. Successfully aggregated adapters are redistributed to the participating edge deployment.

Figure 2 presents the operational sequence from round initialization and monitoring through indicator normalization, utility ranking, admission, upload, and redistribution. Monitoring messages follow a lightweight control path, while adapter parameters use the scheduled transmission path. The coordinator retains the latest monitoring tuple and a short history for each node. Deferred nodes remain eligible in subsequent rounds, with their freshness credit updated independently of their current feature drift.

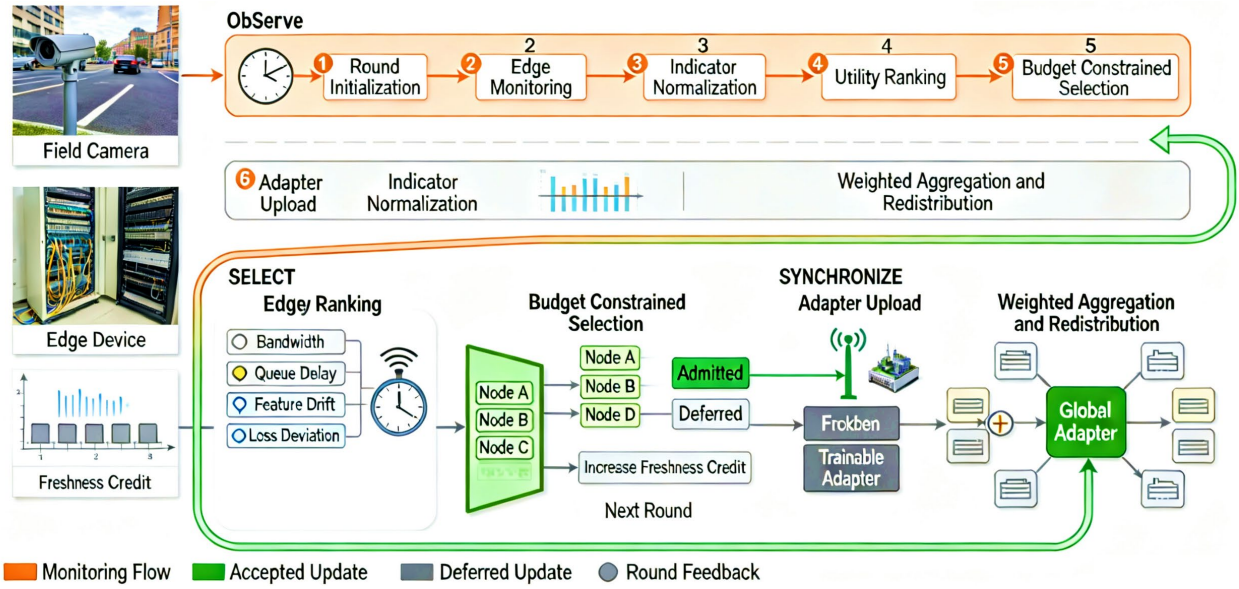


Figure 2. Bandwidth drift utility scheduling workflow for adapter synchronization

Repeated deferral is tracked through the number of rounds since a node last contributed successfully. A capped staleness equation converts this interval into a bounded priority credit:

$$S_i(t) = min\left\{\frac{t - t_i^{last}}{S_{max}}, 1\right\} \tag{5}$$

The quantity t_i^{last} records the most recent successful contribution, and $S_{\max}$ specifies the interval at which the credit reaches its maximum. Its influence remains controlled by κ, while admission continues to depend on communication feasibility. Capping the freshness credit bounds its contribution to priority; service guarantees additionally depend on the available airtime and the feasibility of each deferred upload. Monitoring and aggregation can operate at different frequencies: bandwidth and queue estimates may refresh rapidly, whereas feature statistics and validation loss are updated after local batches. At each scheduling decision, the coordinator uses the newest available observations and retains their timestamps to identify outdated monitoring information.

4. Experimental Evaluation and Analysis

4.1. Experimental Setup and Baselines

The evaluation examines the proposed scheduling mechanism under heterogeneous communication and drift conditions. The experimental configuration comprises 24 edge nodes, four wireless regimes, and five drift regimes. Images are configured at 384 by 384 pixels, with three local adapter training epochs per selected round and an 80 round scheduling horizon. The bandwidth range is 3.2 to 48.5 Mbps, covering stable campus, urban mobile, congested roadside, and emergency hotspot conditions. Drift categories comprise stable scenes, illumination change, viewpoint change, class mixture change, and abrupt scene change. Evaluation designs for heterogeneous distributed learning distinguish channel variability from differences in local data distributions [20]. Following this distinction, the experimental design varies scene urgency and communication feasibility independently, so a drifted node can have either a strong or a weak link. This configuration makes the consequences of participant choice visible under competing constraints. The same node population and drift categories are retained in the efficiency and ablation comparisons, which permits the measured metric changes to be interpreted against a common operating setting. This common configuration supports direct comparisons of participant selection policies under matched operating conditions.

Five participant selection baselines are defined against BDU-SwinV2A, the proposed bandwidth and drift utility scheduler. Periodic-All attempts to activate every node, Random-K selects a matched participant count, Bandwidth-

Only ranks throughput estimates, Drift-Only ranks feature shifts, and Loss-Trigger prioritizes validation loss excursions. Participant selection research distinguishes resource driven policies from policies based on statistical update value [21]. The two single indicator baselines isolate the respective effects of resource availability and statistical urgency, while the random policy represents selection independent of either indicator. The adapter architecture, local training duration, aggregation interval, and payload definition are held constant across the experiments. The comparisons examine model quality and communication cost under a common set of operating conditions. Periodic-All represents broad attempted participation, whereas the selective policies alter the admitted population. Their comparison therefore measures the behavior of complete scheduling policies, with participation volume and ranking both contributing to the observed differences. Completion time and energy describe the resource burden of each scheduling policy, while coverage, fairness, and staleness describe the distribution and timeliness of accepted contributions. These metrics address distinct aspects of system behavior: low upload cost can coexist with limited domain coverage, and balanced participation can coexist with lower adaptation accuracy. Attempted uploads and accepted updates have different effects on aggregation freshness. The former consume communication resources, while the latter modify the shared adapter state. Their distinction provides a basis for interpreting the relationship between transmission efficiency and effective adaptation. The analysis therefore considers accuracy alongside resource expenditure and contribution freshness throughout the baseline comparisons.

Table 1 specifies the node count, wireless range, drift categories, adapter payload, local training duration, and scheduling horizon. Each row defines a quantity that changes the interpretation of the scheduling comparison. The 24-node population determines the opportunity for participant diversity, while the bandwidth range controls the difference between a feasible transmission and a prolonged upload. The 18.6 MB adapter payload sets the per participant communication cost, and three local epochs specify the configured adaptation effort before upload. The 80 round horizon allows both initial convergence and repeated exclusion to be examined. A full model payload of 181.4 MB is used in the payload scaling comparison. The table identifies logging fields needed for implementation, including aggregation traces, feature shift, loss, energy, and time since contribution. Deadline aware distributed learning designs account for both admitted updates and their arrival times when assessing communication policies [22]. Accordingly, completion time is reported through its mean and 95th percentile, while staleness is evaluated through a separate node level measure. The coverage window spans six communication rounds and describes participant representation within that interval. Attempted admission, completed upload, and accepted aggregation are separate events, and the event used to calculate coverage must be identified when comparing this measure with successful-contribution fairness. Frequency based fairness then describes how these opportunities are distributed across the configured population.

Table 1. Experimental configuration of distributed edge update scheduling

Item	Setting	Value	Purpose	Measurement
Edge nodes	Distributed cameras	24 nodes	Evaluate heterogeneous update scheduling	Node logs and aggregation traces
Wireless regimes	Campus, mobile, roadside, hotspot	3.2-48.5 Mbps	Stress bandwidth variation	Mean and percentile completion time
Drift regimes	Normal plus four shifts	5 regimes	Measure statistical urgency	Feature drift and validation loss
Adapter payload	SwinV2 Adapter only	18.6 MB per active node	Reduce upload cost	Payload per round and energy
Training protocol	Local adapter tuning	3 local epochs	Maintain compact adaptation	F1 score and gradient norm
Scheduling horizon	Communication rounds	80 rounds	Observe convergence and staleness	F1, staleness, fairness index

4.2. Accuracy Latency and Resource Efficiency

Figure 3(a) displays five convergence trajectories over 80 communication rounds, with final F1 values of 89.3% for BDU-SwinV2A, 87.7% for Periodic-All, 86.2% for Bandwidth-Only, 87.1% for Drift-Only, and 83.8% for Random-K. The proposed curve crosses the 88% guide at approximately round 42, separating the time required to attain a target from the value at the end of the horizon. Bandwidth-Only has a relatively steep initial rise, followed by an earlier plateau, illustrating how frequent access to readily available nodes can accelerate initial updating while narrowing the diversity of later contributions. The Periodic-All trajectory illustrates a different tradeoff: broad participation increases the communication workload per round, which can delay the shared adapter refresh. The trajectories distinguish initial convergence speed from sustained adaptation across the scheduling horizon. Communication-round convergence and elapsed-time convergence are separate comparisons because the policies have different round completion times. Figure 3(b) separates the final F1 by wireless regime. The proposed values are 91.4%, 89.8%, 87.2%, and 88.6% for campus, mobile, roadside, and hotspot conditions, respectively. Corresponding Periodic-All values are 89.6%, 87.8%, 84.9%, and 86.2%, giving margins of 1.8, 2.0, 2.3, and 2.4 percentage points. The roadside comparison with Bandwidth-Only is 87.2% against 85.6%, a margin of 1.6 points. The margin against Random-K in that regime is 5.7 points. Regime level

values are reported separately from the aggregate score because pooling depends on the sample weights assigned to each regime. The larger hotspot margin motivates tracking the arrival of urgent adapter states under bursty access conditions.

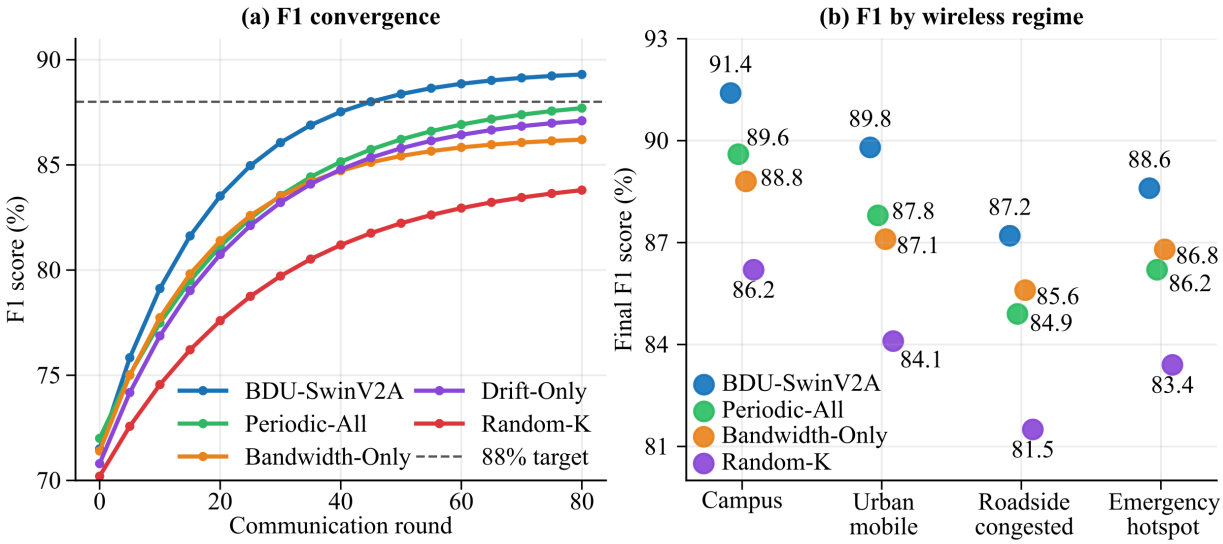


Figure 3. Accuracy behavior under joint wireless and drift heterogeneity: (a) F1 convergence over 80 rounds; (b) final F1 distribution across wireless regimes

Tail delay is relevant to aggregation because the slowest admitted uploads can extend a round beyond its intended synchronization interval [23]. Figure 4(a) compares six policies using mean completion times of 11.2, 18.3, 14.8, 10.4, 13.7, and 12.9 s for BDU-SwinV2A, Periodic-All, Random-K, Bandwidth-Only, Drift-Only, and Loss-Trigger. Their corresponding 95th percentiles are 19.6, 29.4, 25.1, 17.2, 31.8, and 23.3 s. The proposed mean is 38.8% below Periodic-All, while its upper tail remains below the plotted 20 s deadline guide. Drift-Only has a shorter mean and a longer upper tail than Periodic-All, illustrating the distinction between average upload feasibility and exposure to occasional weak links. Bandwidth-Only attains the shortest mean completion time and has a lower final F1 in the accuracy comparison. Figure 4(b) presents energy distributions centered at 42.5, 62.0, 49.4, 39.6, 55.2, and 48.7 J in the same policy order. The proposed center is 31.5% below the periodic value and 12.7% below the loss triggered value. Each violin includes its median and interquartile interval, allowing spread to be distinguished from the central value. The broader drift-oriented distribution is consistent with sensitivity to costly weak-link transmissions; separating that mechanism from local computation requires component-level energy measurements. Energy aware learning models separate transmission expenditure from local computational expenditure when comparing participant policies [24]. The reported energy metric represents the complete configured round. Its interpretation therefore accounts for both communication and local computation. Fewer admissions and shorter radio occupancy can reduce transmission expenditure, while the local training duration determines the computational cost associated with each selected node. The energy comparison is considered alongside payload and completion time to distinguish reduced resource use from changes in the amount of adaptation performed.

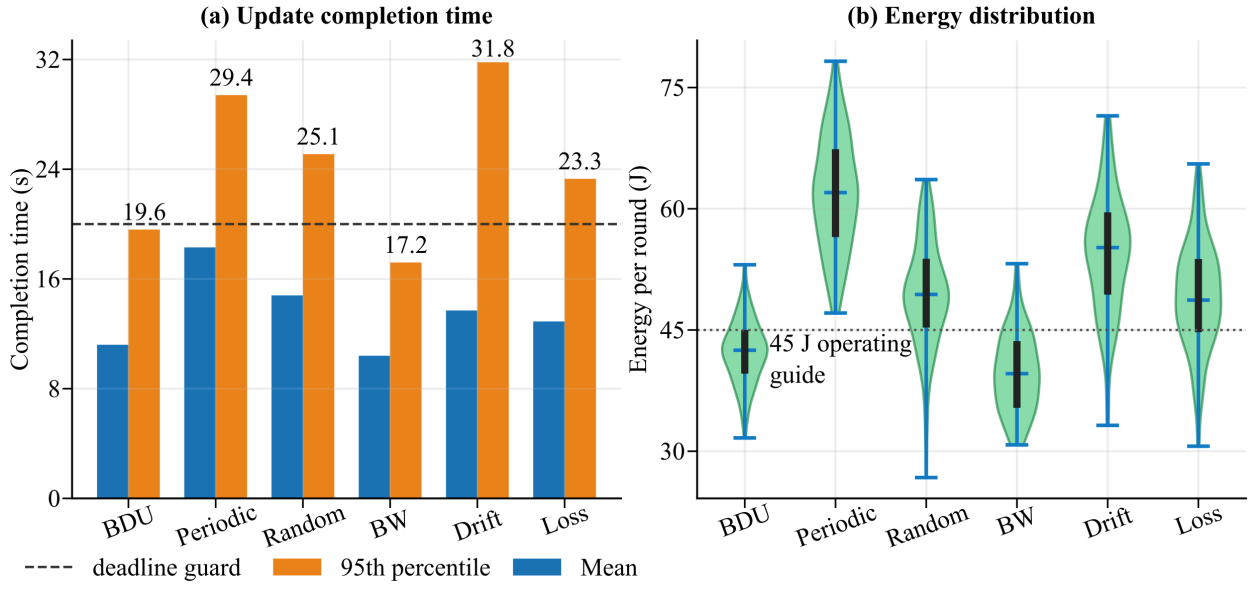


Figure 4. Communication and energy efficiency: (a) mean and 95th percentile update completion time; (b) per-round energy consumption distribution

Figure 5(a) relates the aggregate final F1 to payload normalized by Periodic-All. BDU-SwinV2A occupies the point at 42.8% payload and 89.3% F1, compared with 39.5% and 86.2% for Bandwidth-Only, 46.7% and 87.1% for Drift-Only, 43.0% and 83.8% for Random-K, and 44.1% and 86.6% for Loss-Trigger. The proposed policy requires 3.3 percentage points more normalized payload than the bandwidth policy and achieves a 3.1-point accuracy improvement. The comparison highlights an association between participant choice and adaptation quality: policies with similar payload can produce substantially different accuracy. The F1 margin is an absolute percentage-point difference, whereas the payload values are percentages of the periodic reference. Participant coverage measures whether local domains receive opportunities to influence a distributed model within a defined interval [25]. Figure 5(b) reports six round coverage for five drift regimes. BDU-SwinV2A covers 96%, 93%, 90%, 88%, and 92% of nodes in the normal, illumination, viewpoint, class mixture, and abrupt change categories. The bandwidth policy covers 95%, 85%, 80%, 77%, and 74%, respectively. The gap expands from one point in stable scenes to 18 points under abrupt change, even though the proposed payload remains below half the periodic payload. This pairing of payload and coverage illustrates the intended role of urgency: a communication budget can be redirected toward changed scenes while stable scenes require fewer repeat contributions. Coverage records the opportunity to participate, whereas accuracy records the model response after aggregation. Node identities and drift states provide the relevant distinction between repeated participation by a small subset and broader representation of a shifting domain. The coverage values therefore complement payload efficiency by describing which local distributions receive opportunities to influence the shared adapter.

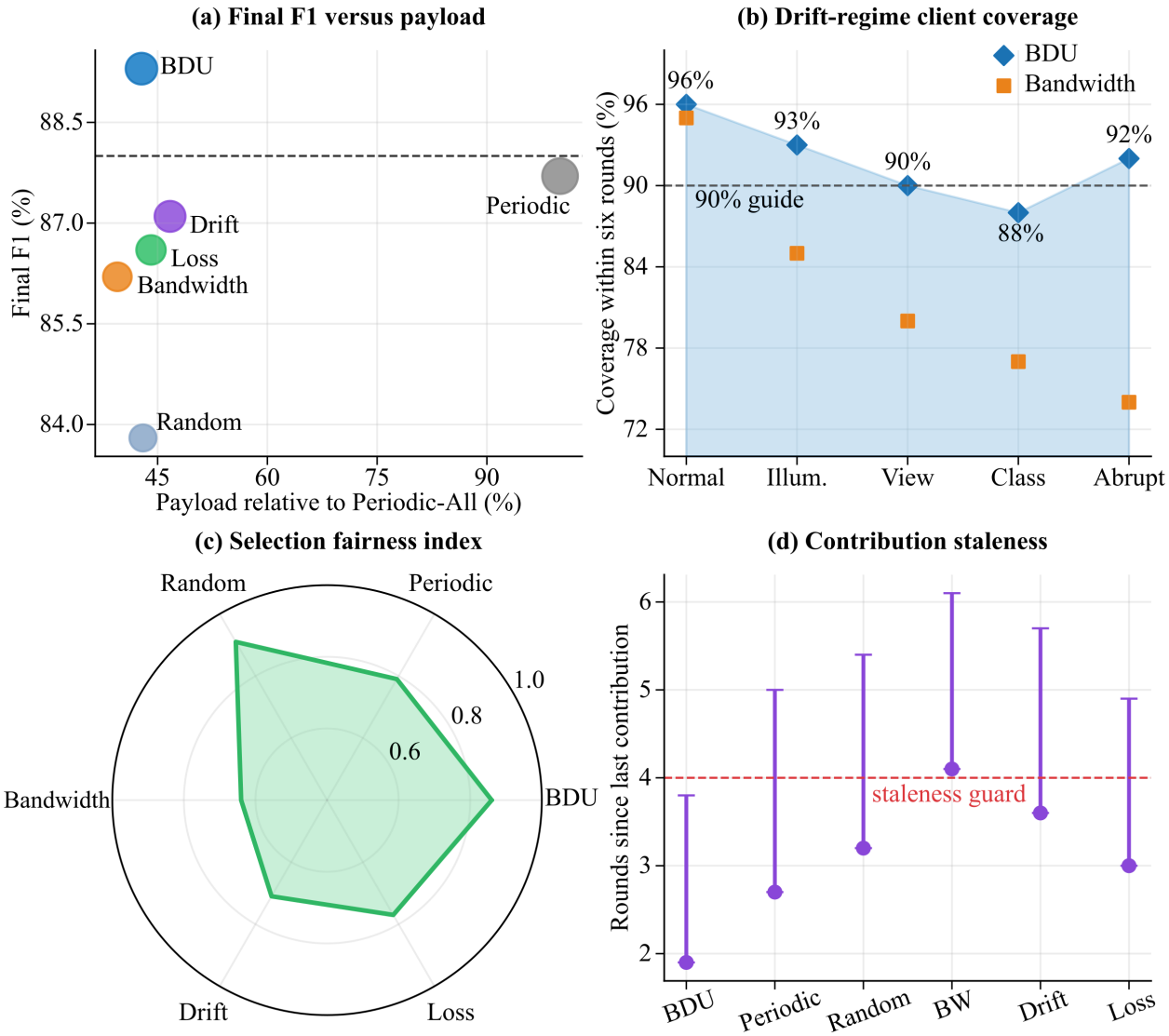


Figure 5. Payload, coverage, fairness, and staleness analysis: (a) final F1 versus transmitted payload; (b) drift-regime client coverage; (c) selected-client fairness index; (d) node-level staleness distribution

Figure 5(c) presents Jain frequency fairness values of 0.86, 0.79, 0.91, 0.64, 0.71, and 0.77 for the proposed, periodic, random, bandwidth, drift, and loss policies. Frequency based fairness describes balance in successful contributions across clients [26]. The periodic value is calculated from completed contributions, so differences in upload success affect the index. Random-K exhibits the highest fairness and the lowest aggregate F1, while Bandwidth-Only exhibits the lowest fairness and a lower payload. These pairings separate balanced access from statistical relevance. Jain fairness is computed from successful-contribution frequencies, so differences in completion success can influence the index even when attempted participation is broadly distributed. Figure 5(d) gives median staleness values of 1.9, 2.7, 3.2, 4.1, 3.6, and 3.0 rounds, with corresponding 95th percentiles of 3.8, 5.0, 5.4, 6.1, 5.7, and 4.9 rounds. Delayed contributions can interact with changing local distributions because an adapter trained earlier may reflect a superseded state [27]. The proposed 95th percentile falls below the four-round guide, whereas the bandwidth policy exceeds it by 2.1 rounds. A median above four rounds for the bandwidth policy indicates that delayed representation extends beyond a small tail of excluded nodes. The bounded freshness credit raises the priority of a node as its contribution ages, while the communication cost continues to limit admission. These paired metrics connect participant access to the age of accepted contributions. A prolonged interval at a rapidly changing node has a different adaptation consequence from the same interval at a stable node. The scheduling objective addresses this distinction through the joint influence of drift urgency and bounded freshness credit.

4.3. Ablation Sensitivity and Scenario Analysis

Sensitivity analysis of weighted participant utilities examines whether operating behavior changes sharply as resource and statistical priorities are adjusted [28]. Figure 6 displays the coefficient sensitivity surface for bandwidth weight alpha and drift weight gamma, sampled at nine values on each axis between 0.10 and 0.70. The highest cell is approximately 89.5% F1 near the designated pair of 0.36 and 0.32. Within the neighborhood bounded by alpha of 0.30 to 0.42 and gamma of 0.25 to 0.38, interpolated scores remain above 89.3%. White contours mark 86%, 87%, 88%, and 89%, so the change in accuracy can be examined beyond the highlighted operating point. At the edge represented by alpha of 0.70 and gamma of 0.10, the surface falls to approximately 86.45%; the opposite corner, alpha of 0.10 and gamma of 0.70, is approximately 85.40%. The asymmetric response reflects the differing effects of bandwidth and drift emphasis on participant choice. Higher bandwidth emphasis favors transmission feasibility, while higher drift emphasis increases the priority of changing scenes. The variation across coefficient pairs identifies how these priorities affect final accuracy under the evaluated conditions. The plot represents accuracy sensitivity; identifying a region that also satisfies delay constraints requires completion time measurements at each coefficient pair. Coefficient calibration should preserve the node population and trace ordering across grid points, with validation separated from the data used to select the operating pair. A stable region spanning several adjacent coefficients would support less frequent retuning, while an isolated maximum would require further assessment of sensitivity to trace selection. The heat map identifies the tested coefficient range and the location of the selected operating point. The neighboring high-scoring region describes local accuracy sensitivity within the evaluated coefficient grid. Interpolated contour values represent the displayed surface between tested settings; their interpretation differs from that of measurements at the sampled grid points. Accuracy sensitivity alone leaves the delay and energy sensitivity of those settings to be assessed.

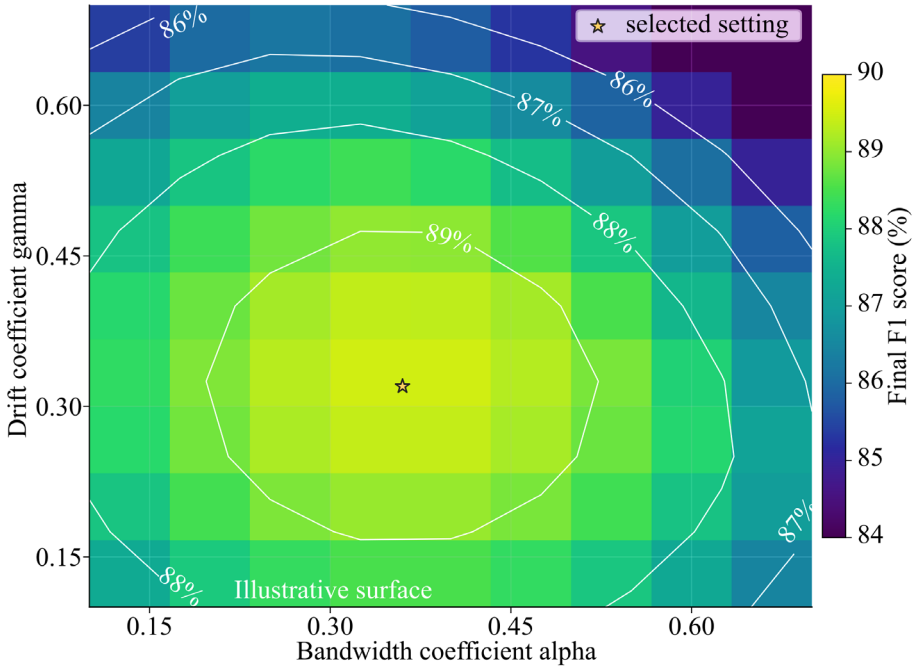


Figure 6. Coefficient sensitivity

Figure 7(a) compares the full utility with five component removals using paired completion time and F1 values. The full setting is 11.2 s and 89.3%. Removing bandwidth yields 16.9 s and 87.0%, removing drift yields 10.8 s and 86.0%, removing loss deviation yields 11.9 s and 87.5%, removing gradient informativeness yields 11.6 s and 88.1%, and removing freshness yields 12.8 s and 88.0%. Removing the drift term reduces completion time by 0.4 s and final F1 by 3.3 percentage points, whereas bandwidth removal raises completion time by 5.7 s. Component ablations in distributed learning hold other optimization choices fixed to attribute changes to the removed signal [29]. The paired accuracy and delay values distinguish the contribution of each utility component. Bandwidth awareness primarily controls transmission cost, whereas drift awareness supports adaptation quality. The loss, gradient, and freshness terms refine participant priority within these two principal objectives. Figure 7(b) contrasts adapter and full model payloads for two through fourteen active nodes in increments of two. At six active nodes, the values are 111.6 MB and 1088.4 MB; at twelve nodes, they are 223.2 MB and 2176.8 MB. The proportional reduction is 89.75% at every point because the comparison assumes fixed per-node payloads of 18.6 and 181.4 MB. This percentage describes adapter parameter bytes relative to full-model parameter bytes; the periodic-normalized payload describes a separate scheduling comparison. The linear curves describe payload accounting under a fixed size per participant. Retransmission overhead and per connection metadata would introduce departures from this proportional relationship. The shaded interval identifies the bytes avoided by adapter transmission, while the 500 MB guide marks a reference communication budget. Under that guide, every plotted adapter point remains feasible in byte terms, while the full-model curve exceeds the guide from four active nodes onward. Byte-budget compliance is distinct from airtime feasibility, which also depends on link rate, overhead, and queue conditions.

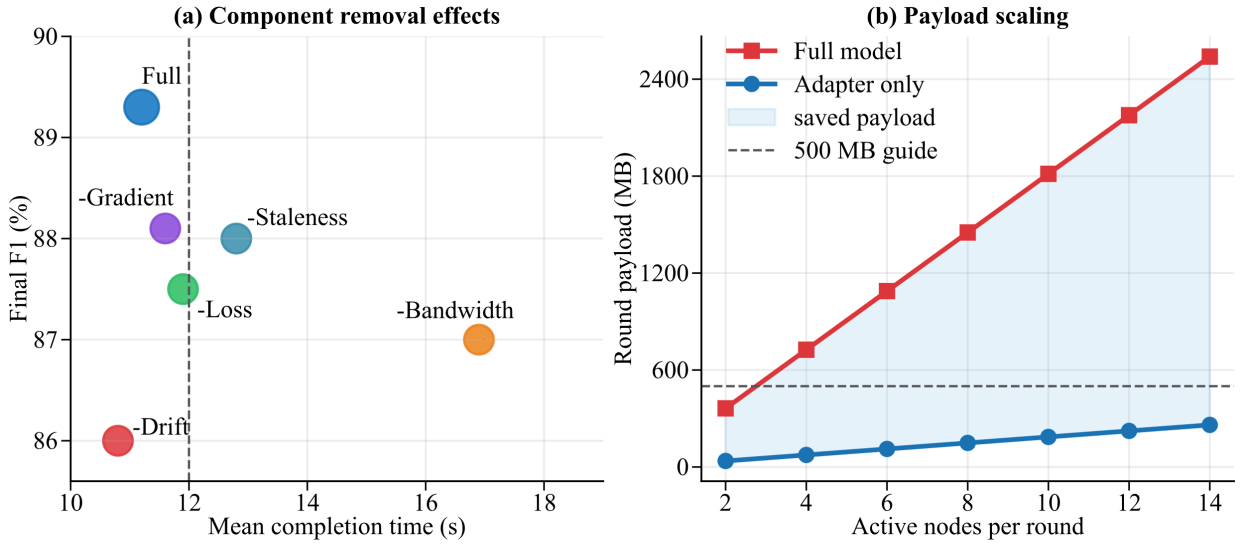


Figure 7. Utility ablation and adapter payload behavior: (a) component removal effects; (b) adapter-only versus full-model payload scaling

Figure 8(a) examines four abrupt drift amplitudes of 0.15, 0.30, 0.45, and 0.60. BDU-SwinV2A achieves F1 values of 89.7%, 88.9%, 87.9%, and 86.8%, compared with 88.8%, 87.1%, 85.0%, and 82.9% for Bandwidth-Only and 87.6%, 85.5%, 83.2%, and 80.7% for Random-K. Across the amplitude range, final F1 decreases by 2.9 percentage points for the proposed policy, 5.9 for the bandwidth policy, and 6.9 for the random policy. This comparison supports examination of the scheduling response to feature changes under a constrained airtime budget. Figure 8(b) gives missed deadline rates at low, medium, high, and extreme contention. Proposed rates are 3.8%, 6.4%, 9.7%, and 14.2%; bandwidth rates are 4.2%, 8.8%, 16.4%, and 23.9%; drift rates are 5.1%, 12.9%, 21.8%, and 34.5%. The proposed point remains below the plotted 15% guide at extreme contention, while the other two policies cross it at high contention. Grouped bars keep each policy visible at every contention level. Drift detection studies distinguish feature distribution changes from sensor faults and label composition changes when interpreting an adaptation trigger [30]. Figure 8(c) samples eighteen feature shift values between 0.05 and 0.90. The selection probability increases from approximately 0.049 to 0.947, passing approximately 0.47 near a shift of 0.40 and 0.73 near 0.55. Bubble size and color encode shift magnitude, and a vertical guide marks the 0.55 action level. The smooth probability profile describes an increasing admission tendency as feature shift grows within the reported operating conditions. This association depends on the accompanying bandwidth, queue, and competing-node states. The curve serves as an operational diagnostic under these conditions; calibration across deployment states requires separate evaluation. The probability curve must be interpreted together with the bandwidth and queue terms, which continue to constrain admission under limited transmission capacity. The plot therefore relates statistical urgency to participant access while preserving the communication feasibility conditions in the scheduling rule.

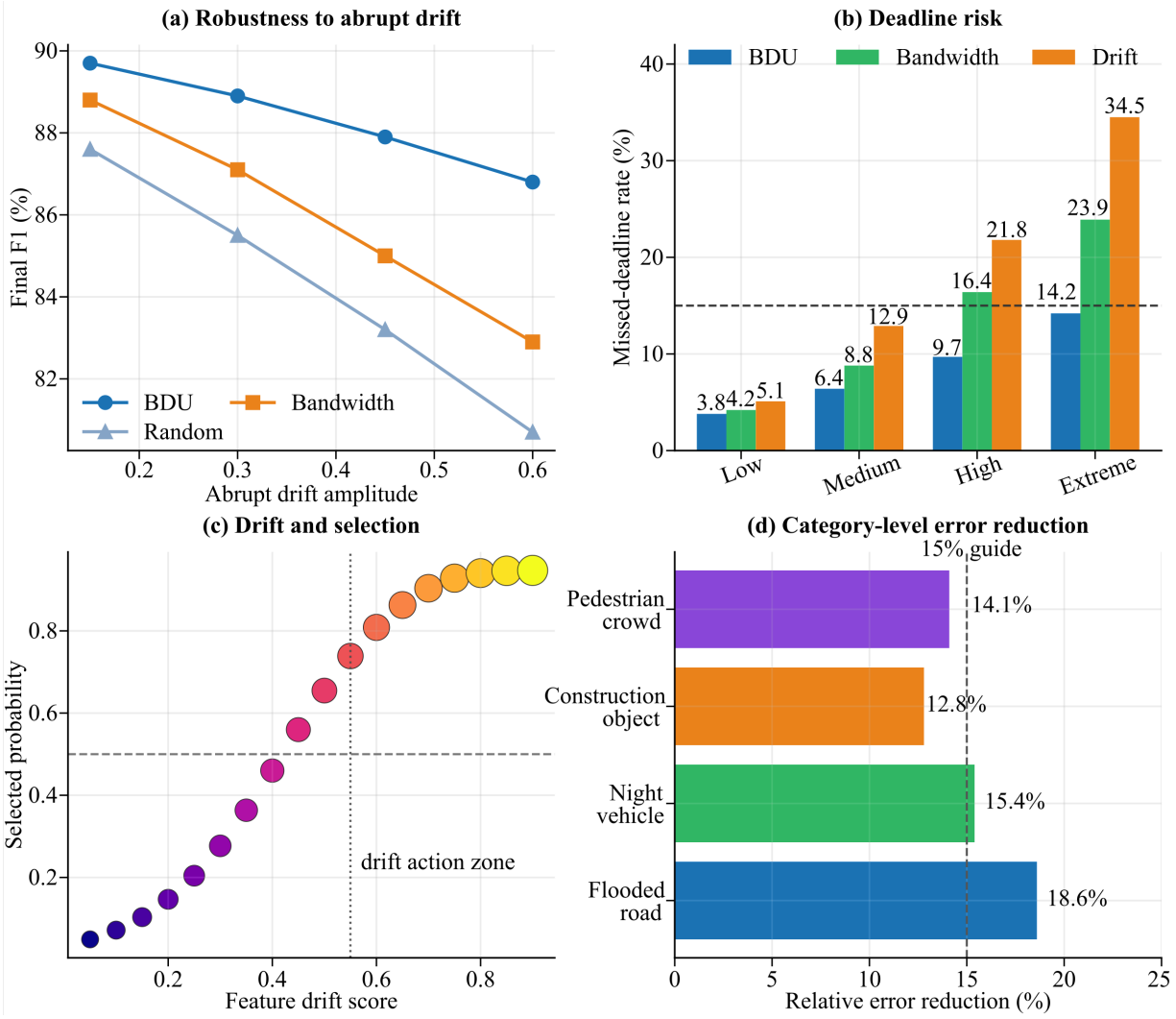


Figure 8. Scenario robustness and operational diagnostics: (a) F1 under abrupt drift amplitude; (b) missed-deadline rate under network contention; (c) selected probability versus drift score; (d) category-level error reduction

Figure 8(d) reports relative error reductions of 18.6%, 15.4%, 12.8%, and 14.1% for flooded road, night vehicle, construction object, and pedestrian crowd subsets against Periodic-All. The flooded road category exceeds the construction category by 5.8 percentage points in the experimental comparison. Relative reductions depend on the baseline error within each category, so the bars should be interpreted alongside category specific baseline errors. The 15% guide separates the flood and night categories from the construction and crowd categories. Category sample counts, baseline confusion matrices, and update timing are relevant to interpreting differences in relative error reduction. A larger reduction can reflect both the initial error burden and the responsiveness of adaptation to the affected scene category. Relative error reduction compares the decrease in category error with the periodic-policy category error; equal relative reductions can correspond to different absolute error changes. Night scenes may require several cameras to contribute because illumination changes affect both object boundaries and background contrast, while construction scenes may involve persistent changes in object mixture. The proposed feature and loss signals can assign different priorities to these events, and attribution to a particular utility term requires corresponding component ablations. Accepted adapter states link local corrections to the shared model response and provide a basis for category-specific analysis. For slow illumination variation, successive contributions should be inspected across neighboring rounds to determine whether priority changes smoothly. For abrupt scene transitions, the relevant record is the interval from a detectable feature shift to the first accepted adapter trained after that transition. Sensor corruption should be recorded as a separate event type, with gradient stability and validation behavior checked before an update is admitted. The interval between feature shift detection and the first accepted update captures the timeliness of adaptation, while gradient stability and validation behavior characterize the reliability of the admitted correction. These complementary measures support category-specific diagnosis under the same communication budget.

5. Conclusion

This paper presented a distributed edge model update scheduling method for SwinV2 Adapter models based on wireless bandwidth and drift indicators. The proposed framework used scalar monitoring signals to connect communication feasibility with statistical update value, selected participants under a per-round budget, and aggregated the selected adapter parameters through the scheduled synchronization path. The study showed that update scheduling can be treated as a model maintenance problem with explicit statistical and communication objectives.

The evaluation has several limitations. It covers a bounded range of node populations, wireless conditions, and drift categories; performance under substantially different operating conditions remains to be assessed. The drift indicators were designed to be lightweight, which reduces monitoring overhead while limiting sensitivity to complex distribution changes. The coefficient settings were examined through sensitivity analysis. Substantial changes in node count, adapter size, or aggregation interval may require coefficient recalibration. Reproducibility requires a specified dataset, train-validation-test separation, device and software configurations, normalization windows, and the complete coefficient settings. The reported point estimates support descriptive comparisons; repeated-run uncertainty and statistical significance require additional evidence. The privacy properties of transmitted summaries and the handling of empty accepted sets, aggregation-weight degeneracy, and stale monitoring inputs also require explicit implementation specifications. These dependencies delimit the current evidence for deployment readiness.

Future research will extend the scheduler to multi-cluster edge hierarchies, online coefficient adaptation, and security-aware update filtering. Further research will combine adapter scheduling with early corruption detection so that nodes affected by sensor faults can be separated from nodes affected by meaningful scene drift. Larger physical testbeds will also be needed to examine how mobility, handover, interference, and edge storage constraints interact with model freshness over long-term operation.

6. Declarations

6.1 Author Contributions

Tadas Mikalauskas contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Deividas Remeika contributes to validation, analysis, investigation, data collection, draft preparation. Orestas Grigon contributes to conceptualization, validation, analysis. All authors have read and agreed with the manuscript before its submission and publication.

6.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

6.3 Institutional Review Board Statement

Not applicable.

6.4 Informed Consent Statement

Not applicable.

6.5 Date Availability

Not applicable.

6.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

7. References

[1] Wang, X., Wang, B., Wu, Y., Ning, Z., Guo, S., & Yu, F. R. (2025). A survey on trustworthy edge intelligence: From security

and reliability to transparency and sustainability. IEEE Communications Surveys & Tutorials, 27(3), 1729–1757. https://doi.org/10.1109/comst.2024.3446585

[2] Xu, Y., Jiang, Z., Xu, H., Wang, Z., Qian, C., & Qiao, C. (2024). Federated learning with client selection and gradient compression in heterogeneous edge systems. IEEE Transactions on Mobile Computing, 23(5), 5446–5461. https://doi.org/10.1109/tmc.2023.3309497

[3] Papa, L., Russo, P., Amerini, I., & Zhou, L. (2024). A survey on efficient vision transformers: Algorithms, techniques, and performance benchmarking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 7682–7700. https://doi.org/10.1109/tpami.2024.3392941

[4] Palanisamy, B., Hassija, V., Chatterjee, A., Mandal, A., Chakraborty, D., Pandey, A., Chalapathi, G. S. S., & Kumar, D. (2025). Transformers for vision: A survey on innovative methods for computer vision. IEEE Access, 13, 95496–95523. https://doi.org/10.1109/access.2025.3571735

[5] Liu, X., Zhang, H., Ren, C., Li, H., Sun, C., & Leung, V. C. M. (2024). Multi-task learning resource allocation in federated integrated sensing and communication networks. IEEE Transactions on Wireless Communications, 23(9), 11612–11623. https://doi.org/10.1109/twc.2024.3383807

[6] He, J., Liu, G., Li, P., Su, X., Jiang, W., Zhang, D., & Zhong, S. (2024). PERS: Parameter-efficient multimodal transfer learning for remote sensing visual question answering. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 14823–14835. https://doi.org/10.1109/jstars.2024.3447086

[7] Gao, D., Song, D., Shen, G., Cai, X., Yang, L., Liu, G., Li, X., & Wang, Z. (2025). FedSTS: A stratified client selection framework for consistently fast federated learning. IEEE Transactions on Neural Networks and Learning Systems, 36(5), 9180–9194. https://doi.org/10.1109/tnnls.2024.3438843

[8] Chen, X., Zhou, X., Zhang, H., Sun, M., & Poor, H. V. (2024). Client selection for wireless federated learning with data and latency heterogeneity. IEEE Internet of Things Journal, 11(19), 32183–32196. https://doi.org/10.1109/jiot.2024.3425757

[9] Ganguly, B., & Aggarwal, V. (2024). Online federated learning via non-stationary detection and adaptation amidst concept drift. IEEE/ACM Transactions on Networking, 32(1), 643–653. https://doi.org/10.1109/tnet.2023.3294366

[10] Xie, C., Chen, Z., Yi, W., Shin, H., & Nallanathan, A. (2025). Tackling class imbalance and client heterogeneity for split federated learning in wireless networks. IEEE Transactions on Wireless Communications, 24(6), 4920–4936. https://doi.org/10.1109/twc.2025.3545236

[11] Chang, C., Lu, T., & Yao, F. (2025). MST-Adapter: Multi-scaled spatio-temporal adapter for parameter-efficient image-to-video transfer learning. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E108-A(5), 736–745. https://doi.org/10.1587/transfun.2024eap1080

[12] Singhal, P., Pandey, S. R., & Popovski, P. (2024). Greedy Shapley client selection for communication-efficient federated learning. IEEE Networking Letters, 6(2), 134–138. https://doi.org/10.1109/lnet.2024.3363620

[13] Yoon, J., Hong, J.-P., & Song, J. (2025). Communication-balancing threshold for event-triggered federated learning. IEEE Access, 13, 136038–136049. https://doi.org/10.1109/access.2025.3592860

[14] Ke, J., Zhou, J., Meng, D., Zeng, Y., Shi, Y., Qu, X., & Guo, S. (2025). JCSRC: Joint client selection and resource configuration for energy-efficient multi-task federated learning. IEEE Transactions on Computers, 74(12), 4094–4108. https://doi.org/10.1109/tc.2025.3605765

[15] Pervej, M. F., Jin, R., & Dai, H. (2024). Hierarchical federated learning in wireless networks: Pruning tackles bandwidth scarcity and system heterogeneity. IEEE Transactions on Wireless Communications, 23(9), 11417–11432. https://doi.org/10.1109/twc.2024.3382093

[16] Øren, A. E., Peuker, N. K., Adeniran-Lowe, A.-M. A., Ruepp, S., & Petersen, M. N. (2025). Concept drift under harsh constraints: A review of potential strategies for IoT systems. IEEE Access, 13, 182220–182237. https://doi.org/10.1109/access.2025.3622973

[17] Lukats, D., Zielinski, O., Hahn, A., & Stahl, F. (2025). A benchmark and survey of fully unsupervised concept drift detectors on real-world data streams. International Journal of Data Science and Analytics, 19(1), 1–31. https://doi.org/10.1007/s41060-024-00620-y

[18] Wang, Y., Liu, M., Song, X., & Nie, L. (2025). TR-Adapter: Parameter-efficient transfer learning for video question answering. IEEE Transactions on Multimedia, 27, 2232–2242. https://doi.org/10.1109/tmm.2024.3521708

[19] Trindade, S., & da Fonseca, N. L. S. (2024). Client selection in hierarchical federated learning. IEEE Internet of Things Journal, 11(17), 28480–28495. https://doi.org/10.1109/jiot.2024.3403082

[20] Park, J., & Ko, J. (2024). FedHM: Practical federated learning for heterogeneous model deployments. ICT Express, 10(2), 387–392. https://doi.org/10.1016/j.icte.2023.07.013

[21] Tan, J., Liu, Z., Guo, K., & Zhao, M. (2025). Long-term client selection for federated learning with non-IID data: A truthful auction approach. IEEE Internet of Things Journal, 12(5), 4953–4970. https://doi.org/10.1109/jiot.2024.3524389

[22] Sun, H., Chen, M., Yang, Z., Cang, Y., Pan, Y., & Huang, K. (2026). Joint task scheduling and resource allocation for multi-task federated learning over wireless network. IEEE Transactions on Wireless Communications, 25, 9062–9077. https://doi.org/10.1109/twc.2025.3644920

[23] Singh, N., & Adhikari, M. (2026). Staleness-aware semi-asynchronous federated learning with adaptive learning rates and non-IID data in edge networks. IEEE Internet of Things Journal, 13(4), 6683–6691. https://doi.org/10.1109/jiot.2025.3635674

[24] Han, X., Chen, W., Li, J., Ding, M., Wu, Q., Wei, K., Deng, X., & Mei, Z. (2025). Energy-efficient wireless federated learning via doubly adaptive quantization. IEEE Transactions on Cognitive Communications and Networking, 11(4), 2216–2231. https://doi.org/10.1109/tccn.2024.3519355

[25] Ghazi, S., Farzi, S., & Nikoofard, A. (2025). Federated learning for all: A reinforcement learning-based approach for ensuring fairness in client selection. IEEE Access, 13, 118515–118535. https://doi.org/10.1109/access.2025.3586943

[26] Shi, Y., Yu, H., & Leung, C. (2024). Towards fairness-aware federated learning. IEEE Transactions on Neural Networks and Learning Systems, 35(9), 11922–11938. https://doi.org/10.1109/tnnls.2023.3263594

[27] Liu, Z., Li, M., Zhang, B., Liu, X., & Zhu, R. (2026). Taming staleness in asynchronous federated learning with system heterogeneity. IEEE Transactions on Parallel and Distributed Systems, 37(10), 2268–2282. https://doi.org/10.1109/tpds.2026.3725721

[28] Li, Q., Lyu, S., & Wen, J. (2025). Optimal client selection of federated learning based on compressed sensing. IEEE Transactions on Information Forensics and Security, 20, 1679–1694. https://doi.org/10.1109/tifs.2025.3526050

[29] Liu, Y., Rini, S., & Chen, J. (2026). GCFed: Exploiting gradient correlation for client selection and rate allocation in federated learning. IEEE Internet of Things Journal, 13(14), 30748–30763. https://doi.org/10.1109/jiot.2026.3687127

[30] Xu, L., Han, Z., Zhao, D., Li, X., Yu, F., & Chen, C. (2024). Addressing concept drift in IoT anomaly detection: Drift detection, interpretation, and adaptation. IEEE Transactions on Sustainable Computing, 9(6), 913–924. https://doi.org/10.1109/tsusc.2024.3386667