Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Reliable Semantic Transport with Function Preservation for Label-Efficient Federated Load Event Recognition

YiXin Zhao
Xinjiang University
Jiao Tian
jiaotian@xju.edu.cn

Xinjiang University

Research Article

Keywords:

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11548524/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Reliable Semantic Transport with Function Preservation for Label-Efficient Federated Load Event Recognition

Abstract

Federated non-intrusive load monitoring (NILM) can exploit appliance-related knowledge across households without centralizing raw measurements. Its practical use is limited by sparse local labels and household-specific operating patterns, which make cross-client semantic evidence unevenly reliable. We propose FedReST, a reliable semantic transport framework for label-efficient federated load event recognition. FedReST organizes collaboration as a two-admission process: the server assesses whether semantic evidence is sufficiently supported for sharing, and each receiving client assesses whether the resulting direction is compatible with its local function. Masked reconstruction and state-aware contrastive learning construct temporal representations that are mapped to simplex-valued semantic embeddings. Clients transmit evidence-weighted task-state prototype measures, which are aligned through reliability-gated partial optimal transport so that unsupported mass can remain unmatched. A function-preserving gradient projection suppresses shared-update components associated with locally anchor-sensitive parameter directions. On frozen REFIT 2014 rolling folds with matched two-epoch local budgets and 10% visible labels, corrected-core FedReST obtains higher mean Event F1 and Window F1 than an architecture-matched Head+BN FedAvg reference, with paired differences of +0.00524 and +0.00104, respectively. It also improves mean Event F1 by +0.01590 over equal-client FedAvg. The results support reliability-aware semantic sharing as a useful design principle for heterogeneous, label-efficient federated load recognition.

Keywords: federated learning,non-intrusive load monitoring,self-supervised learning,optimal transport,personalized learning,gradient projection

1 Introduction

Non-intrusive load monitoring (NILM) fundamentally infers appliance-level operation from aggregate household power measurements. Since its original formulation [1], the field has progressed from heuristic event-driven signal analysis to highly parameterized supervised and self-supervised neural sequence models [2–6]. Reliable recognition of appliance events strictly enables energy feedback, demand-side control, and condition-aware monitoring without the prohibitive cost of installing dedicated meters for every

device. While public datasets (e.g., UK-DALE, REFIT) and standardized software infrastructure have significantly advanced reproducibility [7–9], realistic edge deployment confronts a severe bottleneck: profound statistical heterogeneity. The exact same appliance category routinely exhibits drastically different power levels, durations, duty cycles, and co-occurring background demand across different homes. Consequently, treating cross-household generalization as an unconditional transfer problem is fundamentally flawed; an event pattern learned in one household simply cannot be assumed to possess the same semantic validity in another.

This structural limitation is critically exacerbated under extreme label scarcity. Appliance-level labels strictly require intrusive submetering or manual annotation, neither of which is routinely available at scale in edge environments. Consequently, a practical NILM client operates under a strict budget, observing vast quantities of aggregate sequences but only a minute labeled subset. Self-supervised learning (SSL) paradigms offer a principled mechanism to exploit this unlabeled majority. Masked reconstruction rigorously forces representations to retain temporal context, while state-aware contrastive learning structurally separates diverse operating patterns [10–13]. However, while these objectives mitigate label dependence, they do not render every inferred local state equally trustworthy. Rare events, overlapping concurrent loads, and weakly supervised windows inherently yield uncertain representations and highly unreliable task-state prototypes.

Federated learning (FL) provides a mathematically rigorous framework to bridge these isolated households without exporting raw, privacy-sensitive high-frequency traces [14–16]. However, while federation structurally solves the data-isolation problem, it fundamentally exacerbates semantic mismatch. A participating client may lack a specific appliance altogether, observe only a transient operating state, or possess insufficient reliable evidence to characterize a state that is dominant elsewhere. Therefore, label-efficient federated NILM must strictly quantify and gate the trust placed in each local semantic signal before it is propagated across the network.

Existing federated paradigms exhibit critical architectural gaps when navigating this heterogeneity. Parameter averaging protocols, such as FedAvg, blindly treat local updates as commensurate estimates of a shared objective, failing to establish whether the underlying appliance semantics are sufficiently supported to share [14]. While proximal methods (FedProx) or local normalization mitigate optimization drift [17, 18], they remain fundamentally blind to semantic reliability. Prototype-based federation offers a more direct semantic communication channel [19]; nevertheless, unconditionally averaging prototypes forces artificial correspondences between states that are weakly supported, entirely absent, or genuinely unique to a specific household. Although optimal transport (OT) provides a robust geometric language for aligning probability measures [20–22], standard balanced OT presumes that all local mass must participate in a correspondence. This strict mass-conservation assumption is mathematically incompatible with the reality of sparse, non-overlapping appliance states across diverse homes.

Furthermore, these paradigms systematically ignore a second, equally critical source of negative transfer: receiver-side functional interference. A semantic direction

can be statistically robust at the server yet prove catastrophically disruptive for a specific receiving client. Each household learns highly personalized decision boundaries dictated by its unique appliance mix and background noise. Directly integrating an unconstrained global semantic update inevitably alters predictions on local states that the client has already mastered. Conflating cross-client collaboration into a single, unconditional averaging or alignment operation obscures these distinct failure modes and inevitably triggers catastrophic interference.

To overcome these fundamental barriers, we introduce FedReST, a reliable semantic transport framework designed for label-efficient federated load event recognition. FedReST explicitly restructures collaboration into a rigorous two-admission process. It first learns probabilistic temporal features through masked reconstruction and state-aware contrastive learning, mapping these features to simplex-valued semantic distributions. By employing the Jeffreys divergence, the geometry used for prototype comparison accurately reflects bidirectional probability mass discrepancies rather than mere Euclidean displacement. Each client encapsulates its task-state evidence as an evidence-weighted prototype measure, where the assigned mass explicitly reflects label support and representation reliability. At the server, a reliability-gated partial optimal transport explicitly admits semantic correspondence only for mutually supported mass, leaving unsupported mass deliberately unmatched. This shareability gate decisively prevents weak local prototypes from fabricating artificial global evidence.

The second admission occurs at the receiving client. FedReST constructs a compact, localized anchor set and uses its activation-derived sensitivity structure to identify parameter directions that threaten local predictions. A function-preserving gradient projection (FPR) filters corresponding components of the admitted semantic residual before incorporation. This architecture separates local optimization, which preserves client-specific adaptation, from global semantic transport. FPR constrains shared directions that are sensitive on local anchors, providing receiver-side protection for locally supported decision functions.

Contributions.

1. **A two-admission semantic-sharing formulation.** We fundamentally restructure label-efficient federated load recognition into a dual-gated process: reliability-gated partial transport rigorously evaluates cross-client shareability, while function-preserving projection (FPR) strictly evaluates local functional acceptability.
2. **Evidence-weighted probabilistic prototype transport.** We introduce task-state prototype measures within a simplex-valued representation space. Their Jeffreys geometry and partial-OT alignment explicitly allow uncertain or household-specific semantic mass to remain unmatched, preventing the catastrophic distortion caused by forced global correspondence.

3. **Receiver-side function preservation.** We develop a low-rank, anchor-sensitive gradient projection that strictly regulates how an admitted semantic residual modifies a personalized client model. This mechanism explicitly neutralizes directions associated with local functional drift while preserving the client's localized optimization path.
4. **Event-oriented empirical evidence under strictly matched budgets.** Under the frozen REFIT 2014 rolling protocol with 10% visible labels and matched local-training budgets, the corrected FedReST core shows positive paired in-domain evidence relative to the architecture-matched Head+BN FedAvg reference. Complementary controls provide mechanism-specific evidence that reliability gating improves ranking quality and FPR reduces anchor-function drift.

2 Related Work

This work intersects four fundamental research streams: label-efficient temporal representation learning for NILM, personalized federated optimization, prototype-based semantic alignment, and function preservation under sequential updates. While these streams provide foundational tools, they individually address different aspects of heterogeneity. This section reviews their structural assumptions and positions how FedReST integrates their admission criteria into a unified semantic-sharing protocol.

2.1 Label-efficient NILM and temporal representation learning

NILM has traditionally been formulated as a mapping from an aggregate signal to appliance activity, power demand, or event states. Classical approaches exploit switching signatures and hand-designed event features [1, 2]. More recent deep models learn temporal mappings directly, including sequence-to-point prediction, denoising sequence models, and multi-task architectures [4, 5, 23]. These models possess substantial representation capacity, but their supervised objectives depend heavily on appliance-level targets that require intrusive submetering. A practical household typically generates vast continuous aggregate traces but possesses only a small number of verified device events.

Self-supervised learning provides a rigorous framework to exploit this unlabeled temporal evidence. Masked-prediction objectives compel a model to recover hidden portions of an input from its temporal context [24, 25]. Recent masked autoencoders extend this principle to continuous time-series representations [26]. Contrastive objectives independently force representations of related views to be locally consistent while separating unrelated instances [10, 11]. For time series, temporal and contextual contrasting successfully encode both local dynamics and long-range structural dependencies [12, 27]; recent federated temporal-context learning further illustrates its value

under insufficient labels [28]. Recent NILM studies confirm that these objectives significantly reduce the appliance-level supervision required during downstream learning [6, 13].

However, representation learning alone does not dictate transferability. Two distinct windows can share temporal similarities while representing entirely different physical device states, and a rare local state inherently yields a representation with high uncertainty. In a federated setting, conflating a representation's utility for local prediction with its validity for cross-client sharing can induce negative transfer. FedReST resolves this by deploying masked reconstruction and contrastive learning as candidate semantic extractors, while introducing an evidence-weighted prototype measure. This explicitly quantifies support and reliability, ensuring that representation generation is decoupled from cross-client transfer eligibility.

2.2 Heterogeneous and personalized federated learning

Federated learning (FL) enables collaborative optimization without centralizing raw data [14]. However, the foundational FedAvg formulation assumes that local updates serve as commensurate estimates of a shared global objective. When client data are statistically non-independent and local optimizations diverge, parameter averaging can trigger client drift. Frameworks such as FedProx, SCAFFOLD, and FedNova mitigate this drift through proximal constraints, control variates, or normalized updates [17, 29, 30]. These methodologies target optimization behavior at the parameter level; they do not explicitly assess the semantic validity of the appliance state information driving the update.

Personalized FL (pFL) architectures acknowledge that a single global model may be inadequate for heterogeneous networks. Existing pFL strategies retain local batch-normalization statistics, decouple shared representations from private classification heads, or employ meta-learning regularizers [18, 31, 32]. More recent approaches learn across-client common representations, explicitly model mixtures of client distributions, or seek invariant client-shared structure [33–35]. These designs successfully accommodate client-specific voltage scales and appliance ownership, governing how parameters are partitioned after a common learning objective is established.

FedReST addresses a complementary but distinct challenge prior to parameter aggregation: a client may generate a semantically weak task-state update simply because an appliance is rare, partially labelled, or absent elsewhere. Preserving a local classification head does not evaluate whether that uncertain state should influence the global representation. By introducing an explicit semantic shareability gate at the server and a functional acceptability gate at the client, FedReST extends personalization to the semantic admission level, complementing standard parameter-level pFL.

2.3 Prototype exchange and probability-measure alignment

Prototype exchange provides a compact communication channel, allowing clients to summarize task groups with representative embeddings [19]. Recent federated NILM work considers personalized, privacy-preserving, and clustered collaboration

[36–39]; nevertheless, conventional prototype-averaging assumes semantic parity across clients. This assumption is fundamentally challenged when households possess disjoint appliance inventories or highly unequal pseudo-label reliability. Furthermore, a fixed Euclidean prototype distance does not naturally capture the probabilistic uncertainty inherent to sparsely supervised edge learning.

Information geometry offers a suitable alternative framework. The Jeffreys divergence symmetrizes the discrepancy between probability distributions, while simplex geometry precisely captures shifts in physically meaningful probability masses [40, 41]. FedReST harnesses this geometry to construct a precise transport cost between evidence-weighted task-state prototype measures.

Optimal transport (OT) provides a principled geometric correspondence mechanism [20, 21]. While entropically regularized OT [22] has been applied to personalized federation [42], recent unbalanced OT-based federated cooperation also separates shared and client-specific information [43]. Standard balanced OT, however, imposes a strict mass-conservation constraint. For heterogeneous appliance states, this constraint can inadvertently force weak, missing, or uniquely localized physical modes into global correspondences.

FedReST relaxes this limitation by leveraging unbalanced and partial OT, which allows unsupported portions of a measure to remain unmatched [44, 45]. FedReST interprets unmatched mass as an active abstention from sharing rather than an alignment failure. This reliability-gated transport distinguishes the proposed mechanism from both direct prototype averaging and compulsory prototype matching.

2.4 Function preservation under sequential updates

Even when semantic correspondence is validated at the server, integrating the resulting update can trigger interference at the receiver. This relates to catastrophic interference in continual learning, where elastic weight consolidation [46] and orthogonal subspace projections [47, 48] are deployed to prevent the destruction of prior functions.

The receiver-side vulnerability in federated NILM is mechanically analogous but structurally distinct. The interfering direction is induced by a cross-client semantic residual proposed after server-side alignment, and protection must be localized to the client's established appliance states rather than a global memory. FedReST adapts the subspace perspective by extracting a low-rank sensitive subspace directly from localized anchor activations. Function-Preserving Gradient Projection (FPR) neutralizes components of the admitted global residual that intrude upon this subspace. FPR acts as a targeted safeguard, ensuring that global semantic transfer augments established local functions without inducing catastrophic drift.

2.5 Positioning of the present work

Prior literature provides foundational tools for label-efficient federation, yet a gap remains in unifying their admission criteria into a coherent sharing protocol. Self-supervision constructs temporal features, pFL mitigates parameter heterogeneity, prototype learning communicates semantic summaries, partial OT allows for unsupported abstention, and gradient projection preserves historical functions. FedReST

synthesizes these mechanisms around two fundamental imperatives: verifying that semantic evidence is sufficiently supported to be shared, and constraining whether the admitted direction is acceptable on a particular receiver's local anchors. This separation of semantic selection from local function preservation establishes a principled framework for heterogeneous edge intelligence.

3 Problem Formulation

We formalize label-efficient federated load event recognition as a personalized learning problem constrained by two distinct conditions on cross-client transfer. The first condition is semantic: a local task-state pattern should influence the global representation only when empirical evidence supports a robust cross-client correspondence. The second is functional: an admitted shared direction must strictly preserve the localized predictions that a receiving client has already established on its own supported states. This section introduces the data, model, semantic-measure, and update notation required to mathematically strictly formulate these conditions.

3.1 Federated household setting and sparse supervision

Consider a federation of C heterogeneous households indexed by $c \in \{1, \dots, C\}$. Client c maintains a localized ordered collection of aggregate-power windows,

$$\mathcal{D}_c = \{(\mathbf{x}_{c,i}, \mathbf{y}_{c,i}, \mathbf{v}_{c,i})\}_{i=1}^{n_c}, \tag{1}$$

where $\mathbf{x}_{c,i} \in \mathbb{R}^L$ is an aggregate window with L samples, $\mathbf{y}_{c,i} \in \{0,1\}^A$ contains the binary state labels for A appliance tasks, and $\mathbf{v}_{c,i} \in \{0,1\}^A$ is an explicit visibility mask. Specifically, $v_{c,i,a} = 1$ designates that the label for appliance a is admitted to the supervised objective. Conversely, $v_{c,i,a} = 0$ dictates that the corresponding window is restricted exclusively to self-supervised objectives and confidence-weighted pseudo-labeling. This formulation structurally decouples the physical availability of a signal from the availability of verified appliance supervision.

Households inherently exhibit profound statistical heterogeneity. Their aggregate sequences diverge due to distinct appliance ownership, rated power, usage frequency, and localized background demand. Consequently, neither the local sample count n_c nor the visible task coverage is balanced across the network. Let

$$\Omega_c = \{(i, a) : v_{c,i,a} = 1\} \tag{2}$$

denote the visible client–window–task index set. The global visible-label fraction is formally defined as

$$\eta = \frac{\sum_{c=1}^{C} |\Omega_c|}{\sum_{c=1}^{C} \sum_{i=1}^{n_c} \sum_{a=1}^{A} \mathbb{I}[y_{c,i,a} \text{ is available}]}, \tag{3}$$

where the denominator quantifies total available labels prior to the visibility protocol. The primary operational point is strictly constrained to $\eta = 0.1$.

Following the standard federated privacy boundary, all raw aggregate windows, appliance labels, pseudo-labels, thresholds, and anchor samples are strictly confined to client c. The server accesses exclusively model updates and compact semantic summaries. The objective is to construct a robust shared representation despite extreme label scarcity and non-overlapping local task-state distributions.

3.2 Personalized predictor and local learning objective

The architecture is decoupled into shared parameters θ_s and personalized parameters θ_c^p. The shared component encodes the temporal dynamics and semantic projections utilized across the federation, whereas the private component governs localized prediction and normalization. For an input window $\mathbf{x}_{c,i}$, the model maps to a latent representation, a semantic distribution, and an appliance-state probability vector:

$$\mathbf{h}_{c,i} = f_{\theta_s}(\mathbf{x}_{c,i}), \qquad \mathbf{z}_{c,i} = q_{\theta_s}(\mathbf{h}_{c,i}), \qquad \widehat{\mathbf{y}}_{c,i} = g_{\theta_s,\theta_c^p}(\mathbf{x}_{c,i}), \tag{4}$$

where $\widehat{\mathbf{y}}_{c,i} \in [0,1]^A$. The semantic projection is mathematically constrained to the $(d-1)$-simplex:

$$\mathbf{z}_{c,i} \in \Delta^{d-1} = \left\{ \mathbf{z} \in \mathbb{R}_+^d : \sum_{r=1}^{d} z_r = 1 \right\}. \tag{5}$$

This simplex projection enforces each embedding as a normalized probability distribution over semantic coordinates, providing a statistically sound foundation for cross-client alignment compared to unconstrained Euclidean vectors.

The visible supervised loss for client c is the mask-normalized binary cross-entropy:

$$\mathcal{L}_{\mathrm{sup},c} = -\frac{1}{|\Omega_c|} \sum_{(i,a)\in\Omega_c} \left[y_{c,i,a} \log \widehat{y}_{c,i,a} + (1 - y_{c,i,a}) \log(1 - \widehat{y}_{c,i,a}) \right]. \tag{6}$$

This objective is strictly augmented by reconstruction and contrastive terms evaluated across all available local windows. At communication round t, the unified local objective is defined as:

$$\mathcal{L}_c^t = \mathcal{L}_{\mathrm{sup},c} + \lambda_{\mathrm{rec}} \mathcal{L}_{\mathrm{rec},c} + \lambda_{\mathrm{con}} \mathcal{L}_{\mathrm{con},c} + \gamma_c^t \lambda_{\mathrm{sem}} \mathcal{L}_{\mathrm{sem},c}^t. \tag{7}$$

The semantic term is modulated by the adaptive coefficient $\gamma_c^t \in [0,1]$. When cross-client evidence is insufficient, $\gamma_c^t \to 0$, naturally reducing the local update to its self-supervised and sparsely supervised components.

3.3 Event targets and task-state semantic evidence

While the training target operates at the window level, the physical application demands continuous appliance activation events. Given a client–task threshold $\tau_{c,a}$, the predicted binary state is:

$$\widehat{s}_{c,i,a} = \mathbb{I}[\widehat{y}_{c,i,a} \geq \tau_{c,a}]. \tag{8}$$

Maximal consecutive positive windows form predicted events. Crucially, the threshold is calibrated exclusively on localized training data rather than artificially universally fixed. This structural separation guarantees that calibrated Window F1 and Event F1 evaluate fundamentally different physical endpoints: granular probability ranking versus contiguous event boundary formation.

For semantic transfer, client c aggregates simplex embeddings by appliance a and binary state $s \in \{0, 1\}$. Denoting the retained prototypes as $\{\mathbf{z}_{c,a,s,k}\}_{k=1}^{K_{c,a,s}}$, the local semantic topology is summarized by the discrete probability measure:

$$\mu_{c,a,s} = \sum_{k=1}^{K_{c,a,s}} w_{c,a,s,k} \delta_{\mathbf{z}_{c,a,s,k}}, \qquad w_{c,a,s,k} \geq 0, \qquad \sum_{k=1}^{K_{c,a,s}} w_{c,a,s,k} = 1, \tag{9}$$

where $\delta_{\mathbf{z}}$ is a Dirac measure. The explicitly assigned mass $w_{c,a,s,k}$ rigorously combines visible-sample frequency with confidence-weighted pseudo-label evidence. Consequently, $\mu_{c,a,s}$ precisely encodes both the semantic location and the statistical reliability of the state.

Semantic discrepancy between two simplex-valued prototypes $\mathbf{p}$ and $\mathbf{q}$ is strictly evaluated via the symmetric Jeffreys divergence:

$$D_{\mathrm{J}}(\mathbf{p}, \mathbf{q}) = D_{\mathrm{KL}}(\mathbf{p} \| \mathbf{q}) + D_{\mathrm{KL}}(\mathbf{q} \| \mathbf{p}). \tag{10}$$

This metric perfectly aligns with the simplex geometry, capturing relative shifts in probability mass allocation without arbitrarily designating a reference distribution.

3.4 Dual-admission semantic sharing

Let $\Delta_{\mathrm{loc},c}^{t}$ denote the shared-parameter update induced by client c's local objectives. Rather than unconditionally averaging these updates, FedReST synthesizes a global semantic reference from $\{\mu_{c,a,s}\}$ to derive a client-specific semantic residual $\Delta_{\mathrm{sem},c}^{t}$.

The framework regulates this transfer through a rigorous dual-admission protocol:

1. **Semantic Shareability (Server-Side):** The server estimates the reliability coefficient $\gamma_c^t \in [0, 1]$ via partial optimal transport. The transport plan evaluates prototype support, alignment cost, and matching entropy. Unsupported or highly uncertain semantic mass is explicitly rejected and remains unmatched.
2. **Functional Acceptability (Client-Side):** Client c extracts an anchor-sensitive projection operator $\mathcal{P}_c$ from its localized anchor set $\mathcal{A}_c$. The semantic residual is strictly filtered through this projection before integration.

The final protected shared update is formulated as:

$$\widetilde{\theta}_{s,c}^{t+1} = \theta_s^t + \Delta_{\mathrm{loc},c}^{t} + \gamma_c^t \mathcal{P}_c\big(\Delta_{\mathrm{sem},c}^{t}\big). \tag{11}$$

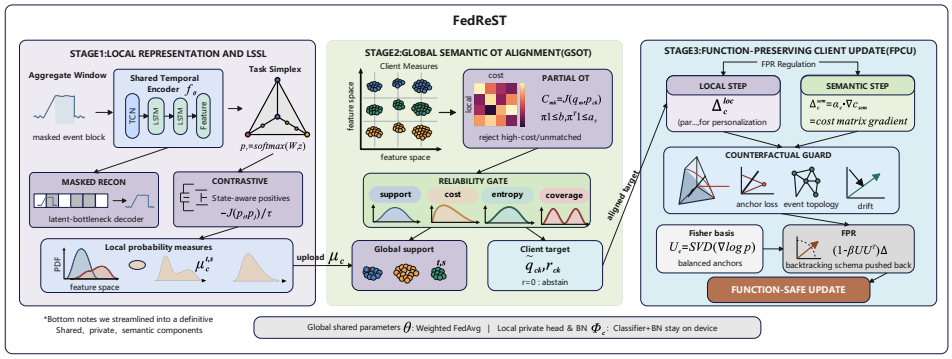


Fig. 1 Overview of FedReST. Clients learn temporal features from sparse labels and unlabeled windows, convert supported task–state embeddings into weighted probability measures, and retain local training anchors. The server constructs OT semantic references and gates unreliable mass. Each receiver then filters the resulting direction through FPR before shared-model aggregation.

The server aggregates these protected states to synchronize the shared model, while private parameters θ_c^p remain strictly localized. This architecture mathematically ensures that semantic transfer occurs exclusively when the evidence is globally supported and locally safe.

3.5 Optimization objective

The overarching goal is to minimize the weighted local objective subject to the dual-admission constraints:

$$\min_{\theta_s,\{\theta_c^p\}_{c=1}^C} \sum_{c=1}^{C} \frac{n_c}{\sum_{r=1}^{C} n_r} \mathcal{L}_c^t, \quad \text{s.t. transfer is bounded by} \quad \gamma_c^t \mathcal{P}_c\big(\Delta_{\text{sem},c}^t\big). \tag{12}$$

This formulation defines a paradigm where the semantic components of an update can be dynamically attenuated or explicitly zeroed, establishing a structurally secure mechanism for heterogeneous edge intelligence.

4 FedReST Methodology

FedReST operationalizes the dual-admission formulation established in Section 3. Moving beyond unconditional prototype alignment, the architecture constructs a compact semantic measure, tests whether that measure possesses a statistically defensible cross-client correspondence, and subsequently verifies whether the induced parameter direction is acceptable for the receiving predictor. Under this closed-loop protocol, the server explicitly leaves unsupported mass unmatched, while the client attenuates or rejects incompatible semantic directions. Figure 1 summarizes the overall architecture.

Algorithm 1 One communication round of FedReST

Require: Shared parameters θ_s^t; client-private parameters $\{\theta_c^p\}_{c=1}^C$; local training data $\{\mathcal{D}_c\}_{c=1}^C$; anchors $\{\mathcal{A}_c\}_{c=1}^C$; references $\{\mathcal{R}_c^t\}_{c=1}^C$

Ensure: Updated shared parameters θ_s^{t+1} and next-round references $\{\mathcal{R}_c^{t+1}\}_{c=1}^C$

Server broadcasts θ_s^t and $\mathcal{R}_c^t$ to each selected client c
for all selected clients c **in parallel do**
    Optimize $\mathcal{L}_{\text{sup},c} + \lambda_{\text{rec}}\mathcal{L}_{\text{rec},c} + \lambda_{\text{con}}\mathcal{L}_{\text{con},c}$ on $\mathcal{D}_c$
    Derive $\Delta_{\text{loc},c}^t$ and a semantic candidate $\mathbf{u}_c^t$ from $\mathcal{R}_c^t$
    $\mathbf{v}_c^t \leftarrow \text{FPR}(\mathbf{u}_c^t, \mathcal{A}_c)$ using Equations (27)–(28)
    Form protected shared state $\widetilde{\theta}_{s,c}^{t+1} \leftarrow \theta_s^t + \Delta_{\text{loc},c}^t + \gamma_c^t \mathbf{v}_c^t$
    Construct evidence-weighted measures $\{\mu_{c,a,s}\}$ and support metadata from local embeddings
    Upload $\widetilde{\theta}_{s,c}^{t+1}$, $\{\mu_{c,a,s}\}$, and support metadata
end for
for all active appliance–state groups (a, s) **do**
    Build a Sinkhorn barycenter $\nu_{a,s}$ from valid $\{\mu_{c,a,s}\}$
    for all clients c contributing to (a, s) **do**
        Solve partial OT $\mathbf{\Pi}_c^\star$ between $\nu_{a,s}$ and $\mu_{c,a,s}$
        Compute $\{r_{c,k}\}$ and $\gamma_{c,a,s}$; set rejected reliabilities to zero
        Store matched targets and reliability metadata in $\mathcal{R}_c^{t+1}$
    end for
end for
$\theta_s^{t+1} \leftarrow \text{Aggregate}(\{\widetilde{\theta}_{s,c}^{t+1}\}_{c=1}^C)$
return θ_s^{t+1} and $\{\mathcal{R}_c^{t+1}\}_{c=1}^C$

4.1 Round-level algorithm and local semantic learning

At communication round t, the server broadcasts global shared parameters θ_s^t and the client-specific semantic references constructed in the previous round. Private classification heads and local normalization statistics remain localized. Client c separates its local task update $\Delta_{\text{loc},c}^t$ from the incoming semantic residual $\Delta_{\text{sem},c}^t$: the former is derived from the local objective, whereas the latter is admitted exclusively through the dual-admission gates.

Algorithm 1 makes the server–client execution order explicit. A reference constructed from the current uploads is used in the subsequent round; this prevents a client from using its own just-uploaded measure as an instantaneous target.

For an aggregate window $\mathbf{x}$, the shared encoder produces $\mathbf{h} = f_{\theta_s}(\mathbf{x})$. A block mask $\mathbf{m} \in \{0,1\}^L$ hides contiguous samples, and a reconstruction head predicts $\widehat{\mathbf{x}}$ at the hidden positions:

$$\mathcal{L}_{\text{rec}} = \frac{\|\mathbf{m} \odot (\mathbf{x} - \widehat{\mathbf{x}})\|_2^2}{\|\mathbf{m}\|_1 + \epsilon}. \tag{13}$$

Two stochastic views of the same window additionally form an instance-contrastive pair. For a minibatch of B windows, its instance term is

$$\mathcal{L}_{\text{ins}} = -\frac{1}{B}\sum_{i=1}^{B} \log \frac{\exp\{s(\mathbf{h}_i^{(1)}, \mathbf{h}_i^{(2)})/\tau_{\text{con}}\}}{\sum_{j=1}^{B} \exp\{s(\mathbf{h}_i^{(1)}, \mathbf{h}_j^{(2)})/\tau_{\text{con}}\}}. \tag{14}$$

When visible labels are available, a state-aware contrastive term further aligns compatible task–state representations. Unlabelled windows participate only through masked reconstruction and the instance term, anchoring temporal representations without fabricating labels [10, 12, 24, 25, 27].

The semantic head maps a latent vector to a strictly positive probability vector,

$$\mathbf{z} = \text{softmax}(q_\phi(\mathbf{h})/\tau_{\text{sem}}) \in \Delta^{d-1}. \tag{15}$$

For two embeddings, we use the symmetric Jeffreys divergence

$$D_{\text{J}}(\mathbf{p}, \mathbf{q}) = \sum_r p_r \log \frac{p_r}{q_r} + \sum_r q_r \log \frac{q_r}{p_r}, \tag{16}$$

after flooring and renormalizing both distributions. Its symmetry prevents designating either household as the semantic reference [40, 41]. The same geometry is used for semantic contrast, clustering, and OT costs. The local objective is

$$\mathcal{L}_c^t = \mathcal{L}_{\text{sup},c} + \lambda_{\text{rec}}\mathcal{L}_{\text{rec},c} + \lambda_{\text{con}}\mathcal{L}_{\text{con},c} + \gamma_c^t \lambda_{\text{sem}}\mathcal{L}_{\text{sem},c}^t, \tag{17}$$

where the adaptive coefficient γ_c^t is evidence-derived rather than manually fixed.

4.2 Evidence-weighted prototype measures and support

Clients filter embeddings before uploading their representation. For appliance a and state s, a visible sample receives evidence weight $e_i = 1$. An unlabelled sample is retained only when its classifier confidence exceeds a local threshold, its simplex projection is closer in Jeffreys geometry to the visible centroid of the proposed state than to the competing state, and its inclusion respects a cap proportional to visible support. An accepted pseudo example receives fractional evidence

$$e_i = \omega_{\text{ps}} \max_{s' \in \{0,1\}} p_\theta(s' \mid \mathbf{x}_i), \qquad 0 < \omega_{\text{ps}} \leq 1. \tag{18}$$

Rejected pseudo labels contribute neither location nor mass. This conservative construction prevents weakly verified predictions from artificially inflating client trust and overwhelming observed supervision.

Weighted clustering yields $K_{c,a,s}$ prototypes. For the members $\mathcal{I}_{c,a,s,k}$ of cluster k,

$$\mathbf{z}_{c,a,s,k} = \frac{\sum_{i \in \mathcal{I}_{c,a,s,k}} e_i \mathbf{z}_{c,i}}{\sum_{i \in \mathcal{I}_{c,a,s,k}} e_i + \epsilon}, \quad w_{c,a,s,k} = \frac{\sum_{i \in \mathcal{I}_{c,a,s,k}} e_i}{\sum_j e_j + \epsilon}. \tag{19}$$

Locations are renormalized to the simplex and empty clusters are removed. The localized measure and its effective support are

$$\mu_{c,a,s} = \sum_{k=1}^{K_{c,a,s}} w_{c,a,s,k}\delta_{\mathbf{z}_{c,a,s,k}}, \qquad n^{\text{eff}}_{c,a,s} = \sum_{i\in\mathcal{I}_{c,a,s}} e_i. \tag{20}$$

The client reports visible and pseudo support independently. These quantities link semantic reliability to verification quality rather than raw local data volume.

4.3 Reliability-gated partial optimal transport

OT is executed independently within each appliance–state group, prohibiting alignment between unrelated tasks. For active clients $\mathcal{C}_{a,s}$, the server initializes a deterministic global support $\nu_{a,s} = \sum_{j=1}^{K_g} b_j\delta_{\mathbf{g}_j}$ and refines it through entropy-regularized Sinkhorn barycenter iterations. With $[\mathbf{C}_c]_{jk} = D_{\mathrm{J}}(\mathbf{g}_j, \mathbf{z}_{c,k})$, the balanced plan for this reference-estimation stage is

$$\mathbf{T}_c^{\text{bal}} = \arg\min_{\mathbf{T}\in U(\mathbf{b},\mathbf{w}_c)} \langle \mathbf{T}, \mathbf{C}_c\rangle - \varepsilon_{\text{ot}} H(\mathbf{T}). \tag{21}$$

Log-domain Sinkhorn updates provide stable entropic OT on compact measures [21, 22].

Balanced transport estimates the global barycenter but does not dictate receiver-specific admission. To govern sharing, the server solves a separate, rejectable partial transport problem,

$$\begin{aligned} \mathbf{\Pi}_c^\star = \arg\min_{\mathbf{\Pi}\geq 0} \quad & \sum_{j,k}(C_{c,jk} - \kappa)\Pi_{jk}, \\ \text{s.t.} \quad & \sum_k \Pi_{jk} \leq b_j, \qquad \sum_j \Pi_{jk} \leq w_{c,k}. \end{aligned} \tag{22}$$

Its upper-bound marginals make $\mathbf{\Pi}_c^\star$ a sub-coupling. Since an edge is used only when its cost is lower than the rejection threshold κ, the solution can retain unmatched mass or abandon transport entirely. This distinguishes FedReST from compulsory balanced alignment [44, 45].

For prototype k, its accepted fraction and transported target are

$$\alpha_{c,k} = \frac{\sum_j \Pi^\star_{c,jk}}{w_{c,k} + \epsilon}, \qquad \widetilde{\mathbf{z}}_{c,k} = \frac{\sum_j \Pi^\star_{c,jk}\mathbf{g}_j}{\sum_j \Pi^\star_{c,jk} + \epsilon}. \tag{23}$$

An unmatched prototype retains its local coordinates but receives zero reliability, eliminating its semantic gradient.

Continuous reliability.

FedReST computes a per-prototype value

$$r_{c,k} = \alpha_{c,k} q_{c,k}^{\mathrm{ent}} \exp(-d_{c,k}/\sigma_{\mathrm{ot}}) \times \min\left(\frac{n_{c,a,s}^{\mathrm{eff}}}{n_{\min}}, 1\right) \min\left(\frac{|\mathcal{C}_{a,s}|}{C_{\min}}, 1\right) \xi_{a,s}, \tag{24}$$

where $d_{c,k} = D_{\mathrm{J}}(\mathbf{z}_{c,k}, \widetilde{\mathbf{z}}_{c,k})$ and $q_{c,k}^{\mathrm{ent}} = (1 - \bar{H}_{c,k})^{\zeta}$ is entropy confidence from the balanced plan. $\xi_{a,s} \in \{1, 0.5\}$ penalizes singleton matches: zero entropy in a 1×1 plan signals lack of alternatives rather than an unambiguous correspondence. The factors measure accepted mass, assignment ambiguity, displacement, within-client support, and cross-client coverage. Values below a minimum threshold are set exactly to zero.

The returned group coefficient is $\gamma_{c,a,s} = \sum_k w_{c,a,s,k} r_{c,k}$, and γ_c^t is its average over active local groups. Consequently, transport can align a robust state while disabling an unsupported state within the same household; direct prototype averaging cannot express this selective operation.

4.4 Function-preserving gradient projection

Partial OT determines global shareability but not receiver-side safety. Each client maintains a compact, class-balanced anchor set $\mathcal{A}_c$ drawn from its verified training split. To avoid a full anchor Jacobian, FPR uses a low-rank task-sensitive approximation. Let $\ell_{a,s}(\theta_s; \mathcal{A}_{c,a,s})$ be mean negative log probability for state s on local anchors. Its normalized score direction is

$$\mathbf{g}_{c,a,s} = \frac{\nabla_{\theta_s} \ell_{a,s}(\theta_s; \mathcal{A}_{c,a,s})}{\|\nabla_{\theta_s} \ell_{a,s}(\theta_s; \mathcal{A}_{c,a,s})\|_2 + \epsilon}. \tag{25}$$

Stacking available score directions in $\mathbf{G}_c$ and taking a thin SVD yields $\mathbf{G}_c = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top}$. The retained basis is

$$\mathbf{B}_c = \mathbf{V}_{1:r_c}, \qquad r_c = \min\left\{ r : \frac{\sum_{j=1}^{r} \sigma_j^2}{\sum_j \sigma_j^2} \geq \eta_{\mathrm{fpr}} \right\}, \tag{26}$$

subject to a fixed maximum rank. This is a low-rank approximation to the task-sensitive Fisher/score span. Layerwise activation bases may be maintained for diagnostics, but the accepted parameter update uses this score basis so that downstream task sensitivity is represented.

For vectorized semantic update $\mathbf{u}_c$, FPR applies

$$\mathcal{P}_c(\mathbf{u}_c) = \mathbf{u}_c - \rho \mathbf{B}_c(\mathbf{B}_c^{\top} \mathbf{u}_c), \qquad 0 \leq \rho \leq 1. \tag{27}$$

The local objective update $\Delta^t_{\mathrm{loc},c}$ is exempt from projection; FPR regulates only the admitted shared semantic residual. Projection is triggered by held-out anchor evidence of increased supervised loss, and optionally by a non-improving event-structure counterfactual, rather than merely a negative parameter-space cosine.

The projected candidate is norm-clipped relative to the local update and backtracked as $\beta^j \mathcal{P}_c(\mathbf{u}_c)$ until it satisfies

$$\begin{aligned} \mathrm{Drift}_c(\mathbf{v}) &= \frac{1}{|\mathcal{A}_c|} \sum_{(\mathbf{x},\cdot)\in\mathcal{A}_c} \|p_{\theta_s+\mathbf{v}}(\mathbf{x}) - p_{\theta_s}(\mathbf{x})\|_2^2 \le \delta_c, \\ \mathcal{L}_{\mathrm{anc},c}(\theta_s+\mathbf{v}) - \mathcal{L}_{\mathrm{anc},c}(\theta_s) &\le \delta_\ell. \end{aligned} \tag{28}$$

If no candidate passes, its semantic residual is zero. This provides a deterministic, finite-anchor safeguard for locally supported decision behaviour rather than a universal preservation claim.

4.5 Implementation, complexity, and communication profile

Prototype clustering uses deterministic initialization and a fixed number of weighted iterations. The server initializes global support with a deterministic farthest-support rule, smooths barycenters between rounds with momentum, runs Sinkhorn in the log domain, and solves partial OT as a bounded-marginal linear program. Barycenter estimation and receiver admission are therefore separate operations.

For K_g global and K_c local prototypes of dimension d, a Jeffreys cost matrix costs $O(K_g K_c d)$. With I_{ot} Sinkhorn iterations and R barycenter refinements, the balanced OT cost is $O(RI_{\mathrm{ot}}K_gK_c)$. The partial program has K_gK_c variables and $K_g + K_c$ marginal inequalities. Client clustering costs $O(N_{c,a,s}K_{c,a,s}dI_{\mathrm{cl}})$. These costs depend on compact measures, not on raw-window length after prototype construction.

Let P be the shared-parameter count and r_c the retained FPR rank. After anchor-score computation, projection costs $O(r_cP)$ time and storage, plus a bounded number of anchor evaluations for backtracking. Each client uploads shared weights alongside prototype locations, masses, and scalar support metadata, giving an incremental payload of $O(\sum_{a,s} K_{c,a,s}(d+1))$ rather than $O(n_cL)$. Raw sequences, detailed pseudo-labels, private heads, and anchors never traverse the network. Measured deployment costs are reported in Section 6.

5 Experimental Protocol

This section defines the frozen evaluation protocol governing the corrected FedReST core. The foundational design principle is temporal separation: calendar months used for optimization, threshold calibration, and endpoint evaluation are strictly disjoint. Every fold–seed paired comparison also uses an identical client set, visible-label mask, optimization budget, and evaluation key across all methods. The protocol is intended to prevent threshold-tuning bias or data leakage from being mistaken for an effect of semantic transfer or function preservation.

5.1 REFIT processing, appliance states, and windows

The primary in-domain evaluation uses the 2014 partition of REFIT [8]. We retain refrigeration, washing machine, dishwasher, and tumble dryer. REFIT channel assignments are household-specific; all channels mapped to one canonical category are summed before state labeling. If a task is physically absent from a household, it is marked unavailable rather than imputed as an off state.

For every valid household and calendar month, the aggregate sequence is partitioned into maximal contiguous segments containing finite, non-negative power. Segments shorter than the input width are discarded. Aggregate windows are normalized by a fixed physical scale of 4000 W, and no statistic estimated from calibration or evaluation months enters preprocessing. An input window contains $L = 100$ consecutive samples. Its state label is extracted at the temporal centre. With summed appliance power P_a at a valid centre, the binary target is

$$y_a = \mathbb{I}[P_a \geq \vartheta_a]. \tag{29}$$

For refrigeration, washing machine, dishwasher, and tumble dryer, the fixed physical on-state thresholds are 50 W, 20 W, 10 W, and 20 W, respectively. These thresholds define ground-truth operating states and are distinct from learned probability thresholds used at inference. A task target is unavailable when its channel is absent or a required appliance value at the centre is non-finite or negative.

Windows are extracted without crossing discontinuities. Training uses stride 1200 to reduce redundant optimization samples; calibration and evaluation use the denser stride 300 to preserve temporal resolution for event scoring. This segmentation, scaling, centre-label rule, and split-specific stride are shared by every compared method.

5.2 Clients, chronology, and visible-label construction

Houses 2, 3, 5, 7, 15, and 20 constitute the federated network. They provide usable 2014 recordings for the target tasks while retaining heterogeneity in appliance ownership, activity frequency, and aggregate backgrounds. A house contributes to a task only when that appliance is monitored and has usable data in the months required by the fold. Missing calendar files are never synthesized; clients are excluded only when they fail the pre-specified minimum-month constraint.

We construct three chronological rolling folds: (i) January–May for training, June for calibration, and July for evaluation (`eval_07`); (ii) January–June, July, and August (`eval_08`); and (iii) January–July, August, and September (`eval_09`). Calibration and evaluation are future horizons relative to training and are mutually disjoint. Development decisions for the corrected core were frozen with this in-domain protocol before cross-year and cross-dataset audits.

The supervised loss accesses 10% of valid training labels. The mask is sampled once for each household, appliance, and binary state using

$$s_{c,a,y} = s + 1009c + 97a + y, \tag{30}$$

where $s \in \{42, 7, 2026\}$ is the experiment seed. Within every available client–task–state stratum, the selected count is $\max\{1, \text{round}(0.1N_{c,a,y})\}$, capped by $N_{c,a,y}$. The label mask is immutable across epochs, rounds, and baselines. Unlabelled windows participate only in masked reconstruction and contrastive objectives; they are excluded from the supervised classification loss. Pseudo-labelled examples can affect prototype construction only through the conservative support rules in Section 4.2, and never change the visible-label mask.

5.3 Probability calibration and event construction

To separate decision calibration from test performance, probability threshold $\tau_{c,a}$ is fitted only on the calibration month. For a client–task pair with at least five on and off calibration windows, the selected threshold maximizes calibration macro-F1. If local support is insufficient, the procedure backs off to a pooled task-level calibration fit; if it is also unsupported, the threshold is fixed at 0.5. The selected source and value are archived. Evaluation labels never influence threshold selection.

For calibrated Window F1, $\widehat{s}_{c,i,a} = \mathbb{I}[\widehat{y}_{c,i,a} \geq \tau_{c,a}]$. For Event F1, consecutive positive windows are merged within each client–task–month sequence; no event bridges a missing segment or calendar boundary. Predicted and reference events are matched one-to-one in descending intersection-over-union (IoU) order, using IoU ≥ 0.1. We also report balanced accuracy, AUROC, and AUPRC to distinguish thresholded classification from unthresholded ranking capacity.

5.4 Model configuration and frozen hyperparameters

The corrected FedReST core includes the shared encoder, simplex semantic head, private parameter partition, masked reconstruction, state-aware contrast, confidence-weighted prototype support, Jeffreys OT, and FPR. The optional event-veto extension is excluded from the primary corrected-core endpoint protocol. Settings governing data access, training budget, and local optimization are matched across comparative baselines. Transport and FPR settings were frozen before final endpoint aggregation and are not retuned by task or fold.

5.5 Fair baselines and ablation controls

The primary matched reference is Head+BN FedAvg-2E. It retains the temporal encoder, private classifier head, batch-normalization scope, client population, label mask, optimizer, local epochs, rounds, and temporal folds, but removes FedReST's semantic construction, OT transfer, and FPR. Equal-client FedAvg-2E is a fully shared-parameter reference under the same budget. The expanded comparison includes FedBN, FedProx, SCAFFOLD, FedProto, and a two-pass budget-matched adaptation of FUML. FedProto provides a direct semantic-sharing comparator by unconditionally aggregating Euclidean task–state prototypes without partial matching, reliability gating, or FPR [14, 17–19, 29, 49].

For every method, the architecture, initialization, participating clients, chronological folds, label mask, local optimization budget, calibration routine, and prediction

Table 1 Frozen corrected-core configuration for the REFIT 2014 rolling protocol.

Component	Hyperparameter	Value
Input	Window length; train / calibration–test stride	100; 1200 / 300 samples
Supervision	Visible-label fraction; random seeds	0.10; $\{42, 7, 2026\}$
Optimization	Rounds; local epochs; batch size	6; 2; 256
Local learning	Adam learning rate; semantic update rate	10^{-3}; 0.025
SSL	Mask ratio; reconstruction / contrastive weights; temperature	0.25; 0.1 / 0.1; 0.5
Prototypes	Maximum clusters; pseudo confidence / weight / cap	5; 0.85 / 0.5 / 4× visible support
Partial OT	Regularization; reject cost; cost scale; minimum reliability	0.05; 0.5; 0.25; 0.05
FPR	Energy; strength; norm ratio; anchors per group	0.95; 0.5; 0.5; 8
FPR acceptance	Drift budget; loss tolerance; backtracking factor / steps	10^{-4}; 0; 0.5 / 6

identity are held fixed. The FUML adaptation is explicitly not an exact reproduction of its original power-regression objective: it uses the common event-recognition architecture and one local plus one mutual update pass, matching the two-pass local-update budget. SCAFFOLD uses its control-variate update with learning rate 0.01 under this same fixed protocol. These implementation choices make the comparison a controlled study of federated sharing rules, rather than a claim that heterogeneous source methods are reproduced under their original task definitions.

Two paired corrected-core ablations deconstruct the admission mechanisms. *Core-no-FPR* disables only parameter-space projection. *Core-mean-prototype* replaces reliability-gated partial OT with support-weighted arithmetic-mean prototype aggregation. The completed reliability-only control retains the same prototype eligibility and partial-OT plan but substitutes a support-matched constant for the learned reliability coefficient. The completed isomorphic zero-projection control forms the same anchor subspace and executes the same projection bookkeeping, but applies the unprojected semantic residual. Both controls isolate the intended operation without changing data exposure or update budget. All ablations retain the remaining components, architecture, seeds, label masks, threshold calibration, and training budgets. Centralized training is reported as a non-federated reference rather than a fairness-matched deployment baseline.

5.6 Label-budget and post-freeze robustness audits

To test label efficiency without changing any architectural or optimization setting, the primary 10% mask is replaced by state-stratified visible-label fractions of 1%, 5%, and 20%. Each fraction uses the same three rolling folds and three seeds, and the masks remain paired across FedReST-core and Head+BN FedAvg. A stratum with

available data contributes at least one visible example after fractioning; this prevents an accidental all-hidden state from being conflated with an algorithmic effect, while still allowing client–task supports without visible positive examples when no valid positive window is present.

After the in-domain structure and hyperparameters were frozen, we reran the corrected core and Head+BN reference on two robustness audits without further adjustment. REFIT 2015 uses January–May for training, June for calibration, and July for evaluation; houses failing the pre-specified required-month rule are excluded rather than imputed. UK-DALE uses fixed within-house temporal 60/20/20 train/calibration/evaluation partitions over houses 1, 2, 4, and 5, with appliance availability determining task coverage. In sparse UK-DALE sequences, `centre_sample_index`, rather than a repeated month-window counter, is the unique evaluation identity used for prediction matching and event reconstruction. Both audits are labelled post-freeze robustness assessments because earlier legacy results had already been inspected; they are not used to revise the frozen configuration.

5.7 Statistical plan, audit trail, and scope

Endpoint summaries report the mean and standard deviation across nine matched fold–seed outcomes. Paired summaries report mean differences and fold- and task-level metrics. Percentile 95% intervals over paired differences are descriptive uncertainty summaries: repeated seeds within overlapping chronological folds are not independent population samples.

Every completed run archives its temporal split, missing-month handling, window counts, label-mask fingerprint, calibration threshold sources, and per-window predictions. Before a paired analysis, prediction keys and target labels are checked to ensure identical evaluation constraints. The UK-DALE audit additionally rejects non-unique prediction identities before event reconstruction. This audit prevents differences in data availability, label visibility, or temporal identity from masquerading as algorithmic gains.

REFIT 2014 rolling folds are the in-domain development protocol. REFIT 2015 and UK-DALE [7] are reserved for post-development robustness auditing and do not influence the frozen corrected-core configuration or baseline selection.

6 Results and Discussion

6.1 Main in-domain comparison under matched budgets

Table 2 reports the corrected-core comparison. Under the matched protocol, FedReST obtains higher mean Event F1 and Window F1 than the architecture-matched Head+BN FedAvg reference. The paired differences are +0.00524 for Event F1 and +0.00104 for Window F1; Event F1 is higher in seven of nine matched fold-seed outcomes. The descriptive 95% interval for the Event-F1 difference is $[-0.00780, 0.01828]$.

Relative to equal-client FedAvg-2E, FedReST improves mean Event F1 by +0.01590 and mean Window F1 by +0.00243. This is a secondary comparison because equal-client FedAvg does not match the personalized parameter scope of FedReST.

Table 2 Main REFIT 2014 in-domain comparison at 10% visible labels. Values are mean ± standard deviation over nine matched fold-seed runs.

Method	Window F1	Event F1	BA	AUROC	AUPRC
Head+BN FedAvg-2E	.663 ± .014	.612 ± .045	.722 ± .023	.833 ± .017	.455 ± .023
Equal-client FedAvg-2E	.662 ± .013	.601 ± .062	.716 ± .022	.834 ± .022	.456 ± .017
FedProto-style-2E	.651 ± .009	.619 ± .046	.695 ± .015	.794 ± .015	.442 ± .019
FedReST-core	.664 ± .012	.617 ± .051	.721 ± .027	.831 ± .017	.455 ± .011

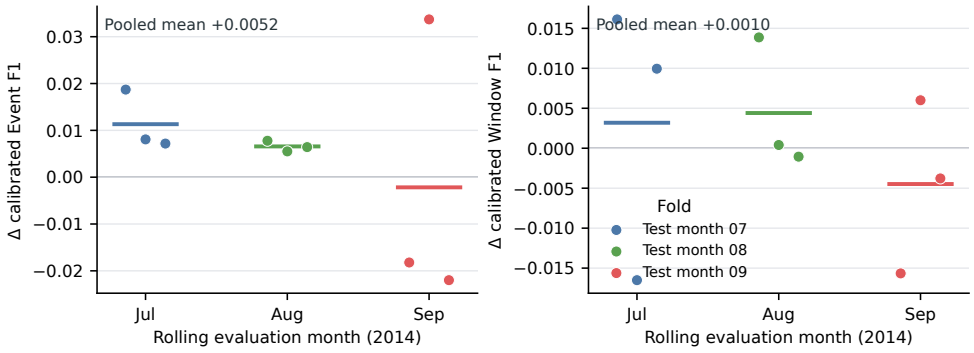


Fig. 2 Paired corrected-core comparison against the architecture-matched Head+BN FedAvg reference. Points are the nine matched fold-seed outcomes and horizontal segments are fold means. The pooled changes are descriptive because seeds and rolling folds are not independent datasets.

Figure 2 reports all nine matched outcomes, rather than hiding fold–seed variation behind aggregate means. It makes the heterogeneous September outcomes visible and is consistent with the descriptive interval reported above; the underlying paired values are retained in `j7_pairwise_deltas.csv`.

6.2 Selective versus direct prototype sharing

Compared with FedProto-style-2E, FedReST obtains higher Window F1 by +0.01271, balanced accuracy by +0.02534, AUROC by +0.03651, and AUPRC by +0.01359. Under the frozen protocol, these consistent window-level contrasts indicate that reliability-gated partial transport improves score separation and ranking quality relative to the direct prototype-sharing comparator. The aggregate Event-F1 difference is −0.00175, with three of nine pairs favoring FedReST.

This metric-dependent ordering is informative rather than contradictory: event detection additionally depends on calibrated thresholding, temporal continuity, and event grouping. Direct prototype sharing and selective transport can therefore affect window-level probability quality and thresholded event construction differently. The FedProto-style comparison is not an isolated causal test of the OT solver, because the methods differ in more than one sharing component; it does not by itself establish that forced alignment causes a particular representation distortion. We therefore additionally report the matched J8 core–mean-prototype ablation below.

6.3 Mechanism evidence

The transport logs record matched and unmatched mass, transport cost, matching entropy, reliability, and effective semantic coefficients for each client-task-state group. As illustrated in Fig. 3(A), semantic admission is not uniform: low-support or high-cost groups can retain unmatched mass and receive a small or zero semantic coefficient. This is the intended operational distinction between partial, reliability-gated transport and compulsory prototype averaging. The FPR logs record raw candidate shared-update norms, accepted projected-update norms, and removed ratios. Figure 3(B) reports the corresponding within-update diagnostic for every matched run: projection reduces pooled mean anchor-function drift by 95.94% and reduces drift in all nine displayed runs. This diagnostic is distinct from the subsequent isomorphic control, which compares two complete update executions.

Table 3 pairs the operational logs with matched endpoint controls. Relative to the identical core without FPR, the protected core has mean changes of +0.00242 Event F1 and +0.00016 Window F1, with a +0.00645 AUPRC difference; the Event-F1 interval is $[-0.01131, 0.01614]$. Relative to arithmetic-mean prototype aggregation, partial OT has higher AUROC (+0.00490) and AUPRC (+0.00787), alongside lower Event F1 (−0.00660) and Window F1 (−0.00243). The isomorphic controls then retain the same semantic pipeline while changing only the reliability coefficient or projection application, directly linking the observed transport and projection behaviour to their corresponding update operations.

Table 3 Mechanism evidence under the frozen REFIT 2014 protocol. Endpoint rows use nine matched fold–seed runs; intervals are descriptive.

Evidence	Matched observation or diagnostic
Complete design	Versus Head+BN FedAvg-2E: Δ Event F1 = +0.00524; Δ Window F1 = +0.00104.
Selective sharing	Versus FedProto-style-2E: Window F1 +0.01271; Event F1 −0.00175.
Semantic admission	Logs record matched/unmatched mass, cost, entropy, reliability, and effective coefficient; unsupported groups can receive zero weight.
FPR within-update diagnostic	Raw and projected directions are logged; pooled anchor-function drift decreases by 95.94%.
FPR endpoint ablation	Versus core-no-FPR: Event F1 +0.00242; Window F1 +0.00016; AUPRC +0.00645.
Partial-OT endpoint ablation	Versus mean prototypes: AUROC +0.00490, AUPRC +0.00787; Event F1 −0.00660.
Reliability-only control	Learned reliability versus support-matched constant: AUROC +0.00681 (7/9 positive); Event F1 +0.00369 (4/9 positive).
Isomorphic FPR control	Projected versus bookkeeping-matched unprojected residual: AUPRC +0.00512 (7/9 positive); mean anchor drift is 40.9% lower after projection.

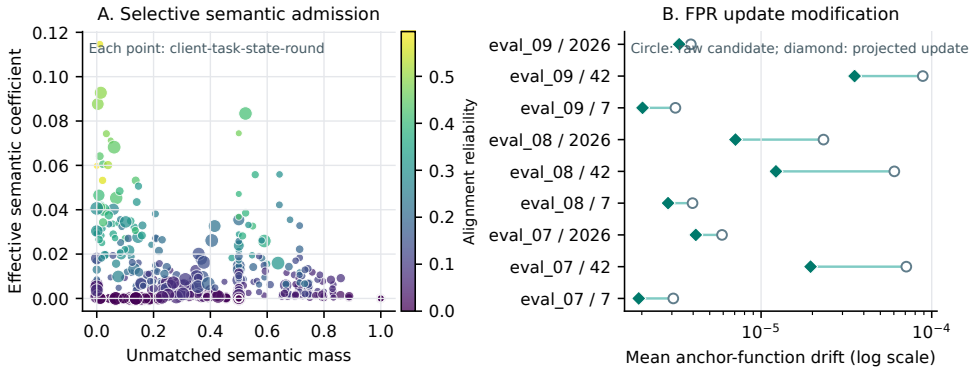


Fig. 3 Mechanism diagnostics from the frozen corrected-core logs. (A) Partial OT retains unmatched mass and scales admitted semantics by reliability; each point is a client–task–state–round group. (B) FPR changes the incoming semantic direction before incorporation and reduces mean anchor-function drift in all nine displayed runs. These operational diagnostics are not substitutes for the matched endpoint ablations in Table 3.

6.4 Measured deployment profile

Raw load signals, labels, and anchors remain local in every variant. We profiled the frozen J7/J8 corrected-core variants on a fixed hardware/software configuration over the three seeds. Timing uses rounds 2–6, treats round 1 as warm-up, synchronizes CUDA timing, and excludes evaluation. Communication is the logical numerical payload of one model upload and download per client per round, plus semantic measures where applicable; it excludes serialization headers and transport framing, and uses no compression. These definitions matter because wall-clock and network costs cannot be inferred reliably from parameter counts alone.

Table 4 reports the observed profile. The reliable semantic stack increases payload from 0.1823 to 0.1861 MiB per client-round (+2.10%) relative to Head+BN FedAvg-2E; the increase is attributable to prototype measures, while raw windows and anchors remain local. The complete core requires 6.048 ± 0.298 seconds per measured training round versus 2.015 ± 0.082 seconds for the parameter-only reference. Within the semantic stack, adding FPR changes no payload and adds 0.777 seconds per measured training round relative to the otherwise matched no-FPR core. Partial OT accounts for 0.380 ± 0.047 seconds per measured round in the complete core. The measurements characterize the tested six-client implementation.

6.5 Causal controls and label-efficiency profile

The reliability-only control preserves the prototype pipeline, group eligibility, and partial-OT plan but replaces learned reliability by a support-matched constant. Learned reliability improves AUROC by +0.00681 (seven of nine pairs; descriptive t interval $[0.00133, 0.01229]$) and changes Event F1 by +0.00369 (four of nine pairs). The isomorphic zero-projection control constructs the same anchor subspace and executes the same acceptance bookkeeping but applies the unprojected semantic residual.

Table 4 Measured deployment profile (three seeds; rounds 2–6; evaluation excluded). Payload is logical numerical communication per client-round without compression.

Variant	Payload (MiB/client-round)	Training (s/round)	OT (s/round)
Head+BN FedAvg-2E	0.1823	2.015 ± 0.082	0.000
Core without FPR	0.1861	5.271 ± 0.136	0.349 ± 0.023
Core with mean prototypes	0.1843	6.122 ± 0.202	0.001
FedReST-core	0.1861	6.048 ± 0.298	0.380 ± 0.047

Projection improves AUPRC by +0.00512 in seven of nine pairs and reduces mean anchor drift from 2.12×10^{-5} to 1.25×10^{-5} (40.9%). Together, these controls locate the clearest gains in ranking quality and protection of established local functions.

Table 5 reports the frozen label-budget study. At 10% and 20% visible labels, the core has positive mean Event-F1 changes of +0.00524 and +0.00477, respectively, against Head+BN FedAvg; the corresponding changes at 1% and 5% are −0.00107 and −0.00414. The resulting profile is budget-dependent: gains emerge at the primary and moderately expanded budgets, whereas the three scarce budgets collectively yield a near-zero pooled Event-F1 difference (+0.00001 across 27 matched outcomes). Thirty-six available client–task supports contain no visible positive label, reflecting the practical state-coverage challenge faced by sparse NILM supervision.

Table 5 Frozen label-efficiency comparison on REFIT 2014. Entries are mean paired differences (FedReST-core minus Head+BN FedAvg) over nine matched fold–seed outcomes.

Visible labels	Δ Event F1	Positive pairs	Δ Window F1
1%	−0.00107	3/9	−0.00477
5%	−0.00414	5/9	+0.00112
10%	+0.00524	7/9	+0.00104
20%	+0.00477	6/9	+0.00224

6.6 Expanded fair-baseline comparison and task heterogeneity

The expanded comparison uses the same encoder, chronological folds, visible-label masks, local-epoch budget, calibration procedure, and evaluation keys for all methods. Table 6 shows that the complete core attains a mean Event F1 of 0.61679, exceeding Head+BN FedAvg (0.61155), FedBN (0.60941), FedProx (0.59748), the budget-matched FUML adaptation (0.60020), and the evaluated SCAFFOLD configuration (0.52867). The paired improvement over FedProx is +0.01931 (8/9 positive; descriptive t interval $[0.00514, 0.03348]$). FedProto has the largest mean Event F1 (0.61854), whereas FedReST has higher Window F1, balanced accuracy, AUROC,

Table 6 Expanded fair comparison on frozen REFIT 2014 at 10% visible labels. Values are means over nine matched fold–seed outcomes; all variants use the shared evaluation and calibration protocol.

Method	Event F1	Window F1	BA	AUROC	AUPRC
FedBN	.60941	.66858	.72659	**.83785**	**.45847**
Head+BN FedAvg	.61155	.66307	.72180	.83326	.45476
FedProx	.59748	.66157	.71692	.83395	.45381
SCAFFOLD ($\eta = 0.01$)	.52867	.63063	.67748	.75779	.39804
FedProto	**.61854**	.65140	.69519	.79442	.44180
FUML-adapted (2-pass budget)	.60020	.66265	.71398	.83480	.45798
FedReST-core	.61679	**.66411**	.72054	.83093	.45539

and AUPRC. This metric-dependent ordering shows that the transport choice affects thresholded event grouping and window-level discrimination differently under the fixed evaluation protocol.

This ranking is consistent with the task-level structure of NILM: the FedReST-minus-Head+BN Event-F1 changes are +0.00459 for refrigeration, +0.00588 for washing machine, +0.01073 for tumble dryer, and −0.00226 for dishwasher. Likewise, relative to FedProto, the core improves dishwasher Event F1 by +0.03272 but decreases tumble-dryer Event F1 by −0.07300. Reporting these contrasts avoids treating heterogeneous appliance dynamics as a single interchangeable classification task.

6.7 Post-freeze robustness audits

The 2015 and UK-DALE experiments were rerun after configuration freeze with the corrected core and identical Head+BN reference. We report them as post-freeze robustness refreshes because legacy versions of these datasets had already been inspected during development. The REFIT 2015 audit contains three eligible houses under the pre-specified missing-month rule and yields positive mean changes in Event F1 (+0.01367), Window F1 (+0.00359), AUROC (+0.00908), and AUPRC (+0.02587).

For UK-DALE, sparse-window evaluation uses the unique `centre_sample_index`; the repeated month-window counter is excluded from corrected event reconstruction. Table 7 reports the resulting cross-dataset profile. The negative aggregate Event-F1 contrast identifies measurement domain and appliance coverage as consequential factors for semantic transport, motivating domain-adaptive reliability estimation.

Table 7 Post-freeze robustness audits. Values are means; Δ denotes FedReST-core minus Head+BN FedAvg. REFIT 2015 has three eligible households, and UK-DALE has three task-level paired outcomes.

Audit	Δ Event F1	Δ Window F1	Δ AUROC	Δ AUPRC
REFIT 2015	+0.01367	+0.00359	+0.00908	+0.02587
UK-DALE	−0.01523	−0.01240	−0.00520	+0.00528

6.8 Evidence synthesis

Taken together, the frozen REFIT 2014 experiments show a positive mean paired difference against the architecture-matched personalized reference, selective semantic admission that changes ranking behaviour, and FPR that suppresses local anchor-function drift. The component, label-budget, appliance, and robustness analyses further reveal where these effects concentrate: ranking and local-function preservation provide the clearest mechanism evidence, while endpoint gains vary with label budget, appliance dynamics, and domain. This evidence motivates domain-robust semantic reliability estimation as the next extension of reliable semantic transport.

7 Conclusion

Federated load event recognition is challenging because appliance states are sparsely labeled, household operating patterns are heterogeneous, and cross-client semantic correspondences cannot be assumed reliable. This work presented FedReST, a reliable semantic transport framework for label-efficient federated load monitoring. FedReST treats semantic sharing as a two-admission process: the server assesses semantic shareability, and the receiving client assesses compatibility with its local function.

The framework combines self-supervised temporal representation learning, simplex-valued semantic embeddings, evidence-weighted task-state measures, reliability-gated partial OT, and function-preserving gradient projection. Partial transport permits unsupported prototype mass to remain unmatched, while FPR attenuates components of an admitted shared update associated with local anchor-sensitive directions.

Under the frozen REFIT 2014 rolling protocol with 10% visible labels and matched two-epoch budgets, corrected-core FedReST obtains higher mean Event F1 and Window F1 than the architecture-matched Head+BN FedAvg reference. It also achieves higher mean Event F1 than equal-client FedAvg. Relative to direct prototype sharing, it improves window-level and ranking metrics, showing that semantic quality and event structure provide complementary views of federated NILM performance.

Matched causal controls and deployment measurements further show that reliability gating strengthens ranking quality and FPR suppresses anchor-function drift. Extending these mechanisms with domain-adaptive reliability estimation is a promising direction for broader heterogeneous edge deployments.

References

[1] Hart, G.W.: Nonintrusive appliance load monitoring. Proceedings of the IEEE **80**(12), 1870–1891 (1992) https://doi.org/10.1109/5.192069

[2] Zeifman, M., Roth, K.: Nonintrusive appliance load monitoring: Review and outlook. IEEE Transactions on Consumer Electronics **57**(1), 76–84 (2011) https://doi.org/10.1109/TCE.2011.5735484

[3] Kolter, J.Z., Johnson, M.J.: REDD: A public data set for energy disaggregation research. In: Proceedings of the Workshop on Data Mining Applications in Sustainability (2011)

[4] Zhang, C., Zhong, M., Wang, Z., Goddard, N.H., Sutton, C.: Sequence-to-point learning with neural networks for non-intrusive load monitoring. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, pp. 2604–2611 (2018). https://doi.org/10.1609/aaai.v32i1.11873

[5] Bonfigli, R., Felicetti, A., Principi, E., Squartini, S., Piazza, F.: Denoising autoencoders for non-intrusive load monitoring: Improvements and comparative evaluation. Energy and Buildings **158**, 1461–1474 (2018) https://doi.org/10.1016/j.enbuild.2017.11.014

[6] Chen, S., Zhao, B., Zhong, M., Luan, W., Yu, Y.: Non-intrusive load monitoring based on self-supervised learning. IEEE Transactions on Instrumentation and Measurement **72**, 1–13 (2023) https://doi.org/10.1109/TIM.2023.3246504

[7] Kelly, J., Knottenbelt, W.: The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. Scientific Data **2**, 150007 (2015) https://doi.org/10.1038/sdata.2015.7

[8] Murray, D., Stankovic, L., Stankovic, V.: An electrical load measurements dataset of UK households from a two-year longitudinal study. Scientific Data **4**, 160122 (2017) https://doi.org/10.1038/sdata.2016.122

[9] Batra, N., *et al.*: NILMTK: An open source toolkit for non-intrusive load monitoring. In: Proceedings of the 5th International Conference on Future Energy Systems, pp. 265–276 (2014). https://doi.org/10.1145/2602044.2602051

[10] Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: Proceedings of the International Conference on Machine Learning, pp. 1597–1607 (2020)

[11] He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9729–9738 (2020). https://doi.org/10.1109/CVPR42600.2020.00975

[12] Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C.K., Li, X., Guan, C.: Time-series representation learning via temporal and contextual contrasting. In: Proceedings of the International Joint Conference on Artificial Intelligence, pp. 2352–2359 (2021). https://doi.org/10.24963/ijcai.2021/324

[13] Chen, T., Gao, J., Yuan, Y., Guo, S., Yang, P.: Non-intrusive load monitoring based on MoCo_v2, time series self-supervised learning. Energy and Buildings **317**, 114374 (2024) https://doi.org/10.1016/j.enbuild.2024.114374

[14] McMahan, H.B., Moore, E., Ramage, D., Hampson, S., Arcas, B.: Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the International Conference on Artificial Intelligence and Statistics, pp. 1273–1282 (2017)

[15] Kairouz, P., *et al.*: Advances and open problems in federated learning. Foundations and Trends in Machine Learning **14**(1–2), 1–210 (2021) https://doi.org/10.1561/2200000083

[16] Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., He, B.: A survey on federated learning systems: Vision, hype and reality for data privacy and protection. IEEE Transactions on Knowledge and Data Engineering **35**(4), 3347–3366 (2023) https://doi.org/10.1109/TKDE.2021.3124597

[17] Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. In: Proceedings of Machine Learning and Systems, vol. 2, pp. 429–450 (2020)

[18] Li, X., Jiang, M., Zhang, X., Kamp, M., Dou, Q.: FedBN: Federated learning on non-IID features via local batch normalization. In: Proceedings of the International Conference on Learning Representations (2021)

[19] Tan, Y., Long, G., Liu, L., Zhou, T., Lu, Q., Jiang, J., Zhang, C.: FedProto: Federated prototype learning across heterogeneous clients. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 8432–8440 (2022). https://doi.org/10.1609/aaai.v36i8.20819

[20] Villani, C.: Optimal Transport: Old and New. Springer, Berlin, Germany (2009). https://doi.org/10.1007/978-3-540-71050-9

[21] Peyré, G., Cuturi, M.: Computational optimal transport. Foundations and Trends in Machine Learning **11**(5–6), 355–607 (2019) https://doi.org/10.1561/2200000073

[22] Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. In: Advances in Neural Information Processing Systems, vol. 26 (2013)

[23] Shin, C., Joo, S., Yim, J., Lee, H., Moon, T., Rhee, W.: Subtask gated networks for non-intrusive load monitoring. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 1150–1157 (2019). https://doi.org/10.1609/aaai.v33i01.33011150

[24] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT, pp. 4171–4186 (2019). https://doi.org/10.18653/v1/N19-1423

[25] Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., Eickhoff, C.: A

transformer-based framework for multivariate time series representation learning. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 2114–2124 (2021). https://doi.org/10.1145/3447548.3467401

[26] Liu, Q., Ye, J., Liang, H., Sun, L., Du, B.: TS-MAE: A masked autoencoder for time series representation learning. Information Sciences **690**, 121576 (2025) https://doi.org/10.1016/j.ins.2024.121576

[27] Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., Xu, B.: TS2Vec: Towards universal representation of time series. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 8980–8987 (2022). https://doi.org/10.1609/aaai.v36i8.20881

[28] Zheng, H., Liu, H., Liu, Z., Tan, J.: Federated temporal-context contrastive learning for fault diagnosis using multiple datasets with insufficient labels. Advanced Engineering Informatics **60**, 102432 (2024) https://doi.org/10.1016/j.aei.2024.102432

[29] Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S., Stich, S., Suresh, A.T.: SCAFFOLD: Stochastic controlled averaging for federated learning. In: Proceedings of the International Conference on Machine Learning, pp. 5132–5143 (2020)

[30] Wang, J., Liu, Q., Liang, H., Joshi, G., Poor, H.V.: Tackling the objective inconsistency problem in heterogeneous federated optimization. In: Advances in Neural Information Processing Systems, vol. 33, pp. 7611–7623 (2020)

[31] Collins, L., Hassani, H., Mokhtari, A., Shakkottai, S.: Exploiting shared representations for personalized federated learning. In: Proceedings of the International Conference on Machine Learning, pp. 2089–2099 (2021)

[32] Dinh, C.T., Tran, N., Nguyen, T.D.: Personalized federated learning with moreau envelopes. In: Advances in Neural Information Processing Systems, vol. 33, pp. 21394–21405 (2020)

[33] Zhang, H., Li, C., Dai, W., Zou, J., Xiong, H.: FedCR: Personalized federated learning based on across-client common representation with conditional mutual information regularization. In: Proceedings of the 40th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 202, pp. 41314–41330 (2023)

[34] Wu, Y., Zhang, S., Yu, W., Liu, Y., Gu, Q., Zhou, D., Chen, H., Cheng, W.: Personalized federated learning under mixture of distributions. In: Proceedings of the 40th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 202, pp. 37860–37879 (2023)

[35] Tang, X., Guo, S., Guo, J., Zhang, J., Yu, Y.: Causally motivated personalized federated invariant learning with shortcut-averse information-theoretic regularization. In: Proceedings of the 41st International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 235, pp. 47694–47724 (2024)

[36] Pan, Z., Wang, H., Li, C., Wang, H., Zhao, J.: PerFedNILM: A practical personalized federated learning-based non-intrusive load monitoring. Industrial Artificial Intelligence **2**(1), 1–15 (2024) https://doi.org/10.1007/s44244-024-00016-8

[37] Dai, S., Meng, F., Wang, Q., Chen, X.: DP^2-NILM: A distributed and privacy-preserving framework for non-intrusive load monitoring. Renewable and Sustainable Energy Reviews **191**, 114091 (2024) https://doi.org/10.1016/j.rser.2023.114091

[38] Wang, T., Dong, Z.: Blockchain-based clustered federated learning for non-intrusive load monitoring. IEEE Transactions on Smart Grid **15**(2), 2348–2361 (2024) https://doi.org/10.1109/TSG.2023.3326194

[39] Agarwal, V., Ardakanian, O., Pal, S.: A robust and privacy-aware federated learning framework for non-intrusive load monitoring. IEEE Transactions on Sustainable Computing **9**(5), 766–777 (2024) https://doi.org/10.1109/TSUSC.2024.3370837

[40] Jeffreys, H.: An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London. Series A **186**(1007), 453–461 (1946) https://doi.org/10.1098/rspa.1946.0056

[41] Amari, S.-i.: Information Geometry and Its Applications. Springer, Tokyo, Japan (2016). https://doi.org/10.1007/978-4-431-55978-8

[42] Ek, S., Wang, K., Portet, F., Lalanda, P., Cao, J.: FedAli: Personalized federated learning with aligned prototypes through optimal transport. arXiv preprint arXiv:2411.10595 (2024)

[43] Li, H., Huang, W., Wang, J., Shi, Y.: Global and local prompts cooperation via optimal transport for federated learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12151–12161 (2024)

[44] Figalli, A.: The optimal partial transport problem. Archive for Rational Mechanics and Analysis **195**(2), 533–560 (2010) https://doi.org/10.1007/s00205-008-0212-7

[45] Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.-X.: Scaling algorithms for unbalanced transport problems. Mathematics of Computation **87**(314), 2563–2609 (2018) https://doi.org/10.1090/mcom/3303

[46] Kirkpatrick, J., *et al.*: Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences **114**(13), 3521–3526 (2017) https://doi.org/10.1073/pnas.1611835114

[47] Farajtabar, M., Azizan, N., Mott, A., Li, A.: Orthogonal gradient descent for continual learning. In: Proceedings of the International Conference on Artificial Intelligence and Statistics, pp. 3762–3773 (2020)

[48] Zeng, F.-K., Chen, C., Liao, C.-H., Chen, T.-H., Luo, S.-C.: Continual learning with orthogonal weights modification. arXiv preprint arXiv:1812.03019 (2019)

[49] Huang, G., Zhou, Z., Lei, S., Li, Y., Yang, Q., Guo, C., Lin, Z.: Federated unilateral-mutual learning for privacy-preserving nonintrusive load monitoring. IEEE Transactions on Instrumentation and Measurement **74**, 1–12 (2025) https://doi.org/10.1109/TIM.2025.3545886