

City Research Online

City St George's, University of London

Citation: Chaturvedi, P., Ahmad, S., Nowroozi, E. & Rajarajan, M. (2026). PhysioDrift-Fed: Confronting the Drift-Rate Heterogeneity Problem in Federated Learning for Consumer Cardiac Wearables. IEEE Transactions on Consumer Electronics, 72(4), pp. 8734-8741. doi: 10.1109/tce.2026.3727326

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/38421/

Link to published version: https://doi.org/10.1109/tce.2026.3727326

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

PhysioDrift-Fed, Confronting the Drift-Rate Heterogeneity Problem in Federated Learning for Consumer Cardiac Wearables

Prachi Chaturvedi, Shahnawaz Ahmad, *IEEE, Member*, Arvind Mewada, *Senior IEEE, Member*, Ehsan Nowroozi, *Senior IEEE Member*

***Abstract*—Consumer cardiac wearables increasingly rely on federated learning (FL) to deliver personalized AI without compromising user privacy, yet deployed models silently degrade as user physiology evolves with age, medication, and lifestyle. We identify and formalize this challenge as the Drift-Rate Heterogeneity Problem (DRHP) users drift at fundamentally different rates, and existing federated frameworks neither detect drift through physiological signals nor distribute its consequences fairly across the user population. This paper introduces PhysioDrift-Fed, the first FL framework purpose-built to address drift-rate heterogeneity, combining three coordinated components, a physiology-aware drift detector that fuses statistical and electrocardiogram-domain signals, a Drift-Rate Fair Aggregation (DRFA) scheme that equalizes representation across heterogeneous drift strata, and an energy-budgeted selective update protocol respecting wearable resource limits. To support reproducible study of this problem, we release DriftBench-ECG, a public drift-evaluation benchmark on the PTB-XL dataset. PhysioDrift-Fed delivers higher classification accuracy under drift, lower detection delay, substantially reduced fairness variance, and lower per-round energy than five state-of-the-art baselines on real wearable-class hardware.**

***Index Terms*—Consumer cardiac wearables, drift-rate heterogeneity, electrocardiogram, energy-efficient federated learning, fairness-aware aggregation, physiology-aware machine learning.**

I. Introduction

SMARTWATCHES, fitness bands, and cardiac patches have quietly become front-line medical devices, with over one hundred million consumer wearables now running on-device AI for atrial fibrillation screening, rhythm anomaly detection, and emergency alerting [1]. FL trains this intelligence collaboratively without surrendering ECG recordings to a central server [2], [3]. Yet a quiet failure mode has begun to surface in deployed cardiac AI, the models that were accurate at launch grow steadily less accurate as the users they monitor grow older, change medications, recover from illness, or simply live their lives [4]. The AI does not break; it goes blind, gradually and silently.

This phenomenon, formally known as concept drift, has crossed from academic concern to regulatory mandate, the U.S. FDA's 2024 Predetermined Change Control Plan and Article 15 of the 2025 EU AI Act now both require AI-driven consumer health devices to be monitored and corrected post-deployment [5]. Shipping a model and forgetting it is no longer permitted.

The deeper challenge has been overlooked, users do not drift at the same rate. A young marathon runner whose ECG shifts gradually over a decade imposes different federated demands than a sixty-year-old whose cardiac signals reshape within weeks of beta-blocker initiation. We name this the DRHP, drift rates span orders of magnitude across users. It has two faces, detection (generic detectors confuse circadian variation with drift and miss slow medication shifts [5]) and aggregation (fairness schemes equalize across demographic groups but ignore drift-rate heterogeneity [6]). Neither face is solved jointly within wearable energy budgets.

Studies show on-device atrial fibrillation detectors lose measurable accuracy within twelve to eighteen months of deployment, with deterioration distributed unevenly, users in the fastest-drifting strata pay the highest accuracy price [3]. This is the present reality of consumer cardiac AI.

This paper addresses the DRHP with three coupled contributions and one piece of community infrastructure,

1) We formalize the Drift-Rate Heterogeneity Problem as a fairness dimension in FL distinct from group fairness.
2) We propose PhysioDrift-Fed, a unified framework combining a Physiology-Aware Drift Detector, a DRFA scheme, and an energy-budgeted selective layer-update protocol.
3) We release DriftBench-ECG, a public benchmark with standardized gradual, sudden, and recurring drift scenarios on PTB-XL.
4) We validate empirically on one hundred federated clients and a Raspberry Pi 5 hardware testbed against five state-of-the-art baselines.

The remainder of this paper follows the natural arc of the problem. Section II surveys the three research streams that intersect at the DRHP and exposes the gap that motivates this work. Section III formalizes the problem mathematically. Section IV presents the PhysioDrift-Fed framework. Section V describes DriftBench-ECG and the broader experimental protocol. Section VI reports results. Section VII concludes.

II. Related Work

The DRHP sits at the intersection of three research streams, each of which has matured independently but none of which

Prachi Chaturvedi, Shahnawaz Ahmad and Arvind Mewada was with the School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India e-mail: (shahnawaz98976@gmail.com).

Ehsan Nowroozi Centre for Sustainable Cyber Security (CS2), University of Greenwich, United Kingdom

TABLE I
GAP MATRIX OF EXISTING METHODS VS. PHYSIODRIFT-FED

Method	Federated	Drift-Aware	Physiology-Informed Detection	Drift-Rate Fair Aggregation
FedAvg [3]	✓	×	×	×
FedHealth [6]	✓	×	×	×
FedImT [9]	✓	×	×	×
FedProx [23]	✓	×	×	×
CCIoMT-FL [8]	✓	×	×	×
CDA-FedAvg [10]	✓	✓	×	×
FedConD [12]	✓	✓	×	×
CAFE [14]	✓	✓	×	×
q-FFL [17]	✓	×	×	×
FAIR-FATE [24]	✓	×	×	×
FAC-Fed [18]	✓	✓	×	Partial
Group-Drift Fairness [19]	✓	✓	×	Partial
DOCTOR [25]	×	✓	Partial	×
BMI-CL [26]	×	✓	Partial	×
PhysioDrift-Fed (proposed)	✓	✓	✓	✓

has been combined to address the problem directly. This section surveys these streams concisely and exposes the empty cell they collectively leave.

FL first reached wearable healthcare through the FedHealth framework, which combined federated averaging with transfer learning to deliver personalized models without raw-data sharing [6]. The line of work that followed has refined privacy preservation, personalization, and class imbalance handling for consumer health applications [7]–[9]. Across this body of work, however, the data distribution at each client is consistently treated as stationary, leaving deployed models exposed to the silent degradation described in Section I.

A parallel research stream has begun to confront non-stationary data in federated settings. The canonical adaptation, CDA-FedAvg, augments federated averaging with a generic statistical drift detector [10]. Subsequent work has improved detection robustness, introduced asynchronous and multi-model variants [10], [11], and framed drift adaptation through causal inference [12]–[16]. A 2025 systematic survey explicitly names physiological signal domains and individual-level drift heterogeneity as the two least-explored frontiers in this space [7], evidence that the DRHP is recognized but unsolved.

Fairness in FL has been pursued through reweighted aggregation, q-Fair FL equalizes accuracy across clients [17]; FAC-Fed jointly handles fairness and concept drift but defines fairness exclusively over demographic groups [18]; FedProx addresses client heterogeneity without addressing drift [17]; group specific distributed drift fairness has also been formalized [19]. Across this literature, fairness is operationalized over demographic attributes never over individual user drift rates [20]. Adjacent work on continual learning for wearables demonstrates feasibility of on- device adaptation under strict energy budgets [21], [22] but operates outside federated settings. This is the gap that defines the DRHP.

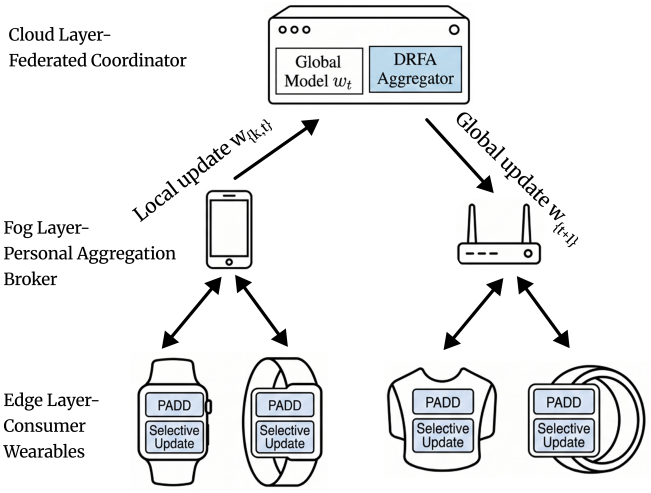


Fig. 1. SYSTEM ARCHITECTURE

III. THE DRIFT-RATE HETEROGENEITY PROBLEM

This section formalizes the DRHP as a tractable joint optimization with three coupled sub-problems, each of which Section IV addresses through a dedicated framework component.

A. System Architecture

We adopt a three-tier consumer-wearable architecture (Figure 1), an edge layer of K ECG-equipped wearables, a fog layer of personal aggregation brokers (smartphones/home gateways), and a cloud layer hosting the federated coordinator that maintains global parameters $\mathbf{w}_t$ over rounds $t \in \{1, \ldots, T\}$. Each round, every client k holds a local stream $D_{k,t}$ of windowed ECG segments and diagnostic labels.

B. Concept Drift and Drift Rate

A concept drift event at client k occurs when the local joint distribution of inputs and labels shifts between rounds,

$$P_{k,t}(\mathbf{x}, y) \neq P_{k,t-\Delta}(\mathbf{x}, y), \quad \Delta > 0. \tag{1}$$

Three drift archetypes are observed in consumer cardiac wearables, *gradual* (aging-induced morphology change), *sudden* (medication initiation), and *recurring* (seasonal physiology). The drift rate of client k at round t is the Wasserstein distance between consecutive empirical local distributions,

$$\rho_{k,t} = \mathcal{W}\big(P_{k,t}(\mathbf{x}, y),\ P_{k,t-1}(\mathbf{x}, y)\big). \tag{2}$$

The drift-rate vector

$$\rho_t = [\rho_{1,t}, \ldots, \rho_{K,t}]^\top$$

summarizes the heterogeneity across the federation at round t.

C. The Drift-Rate Heterogeneity Problem

Given a federation of K consumer wearable clients in which the drift-rate vector ρ_t exhibits non-trivial variance, design a federated learning procedure that simultaneously (i) detects per-client drift accurately using physiological domain priors, (ii) aggregates client updates with weights $\{\alpha_{k,t}\}$ that equalize the global model's performance across drift-rate strata, and (iii) executes the resulting per-client updates within each client's energy budget B_k.

This definition makes drift-rate heterogeneity a measurable federation property, separating what is equalized (performance across drift-rate strata) from how (aggregation weights), and coupling both to wearable energy constraints. The three pieces translate into three coupled sub-problems.

Sub-Problem 1 — Physiology-Aware Drift Detection. *Existing detectors operate purely on $\mathcal{D}_{k,t}$ and ignore physiological features $\phi_{k,t}$ [10], [14].* We require a detector Φ that maximizes the true-drift detection rate while suppressing false alarms caused by benign physiological variation,

$$\Phi^* = \arg\max_{\Phi}\ \mathrm{TPR}(\Phi) - \lambda_1 \cdot \mathrm{FPR}(\Phi). \tag{3}$$

Sub-Problem 2 — Drift-Rate Fair Aggregation. *Given the drift-rate vector ρ_t and local updates $\{\mathbf{w}_{k,t}\}$, determine aggregation weights $\{\alpha_{k,t}\}$ that minimize the variance of post-aggregation accuracy across drift-rate strata while preserving global utility,*

$$\begin{aligned} \min_{\{\alpha_{k,t}\}} \quad & \mathrm{Var}_s\left(\mathbb{E}_{k\in S}\left[F_{k,t+1}\right]\right) - \lambda_2 \cdot \mathbb{E}_k\left[F_{k,t+1}\right] \\ \text{s.t.} \quad & \sum_k \alpha_{k,t} = 1, \qquad \alpha_{k,t} \geq 0 \end{aligned} \tag{4}$$

This is the formal mathematical expression of the equity dimension we name in the DRHP. Existing fairness-aware schemes optimize over demographic strata [17], [19], not drift-rate strata.

Sub-Problem 3 — Energy-Budgeted Continual Adaptation. *Given the detected drift type and the per-client energy budget B_k, decide the subset of layers $\mathcal{L}^*_{k,t} \subseteq \mathcal{L}$ to update on client k such that local performance improvement is maximized within the budget*,

$$\mathcal{L}^*_{k,t} = \arg\max_{\mathcal{L}_{k,t}\subseteq\mathcal{L}} \Delta F_{k,t}(\mathcal{L}_{k,t}) \quad \text{s.t.} \quad \sum_{\ell\in\mathcal{L}_{k,t}} E^{(\ell)}_{k,t} \leq B_k. \tag{5}$$

Conventional federated protocols update all layers uniformly and ignore the per-device energy constraint [26].

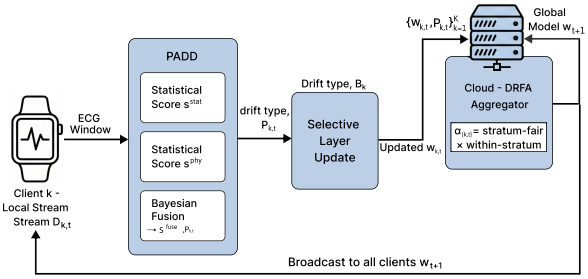


Fig. 2. PhysioDrift-Fed FRAMEWORK BLOCK DIAGRAM

D. Joint Optimization

The three sub-problems are coupled, the detector output Φ informs the drift-rate estimate that feeds into the aggregation weights $\alpha_{k,t}$, and both jointly condition the layer-selection decision $\mathcal{L}^*_{k,t}$. The overall problem PhysioDrift-Fed must solve is

$$\mathcal{P}: \max_{\Phi,\{\alpha_{k,t}\},\{\mathcal{L}_{k,t}\}} \mathcal{U}_{\text{global}} - \lambda_1 \mathrm{FPR} - \lambda_2 \mathrm{Var}_s[F], \tag{6}$$

subject to the energy constraint in (5) and the simplex constraint on α_t, where $\mathcal{U}_{\text{global}}$ denotes global model utility on a held-out validation set. Problem $\mathcal{P}$ is non-convex due to the discrete layer-selection decision and the coupled drift-fairness term. Section IV decomposes $\mathcal{P}$ into three coordinated components, each solving one sub-problem while preserving the joint optimization structure.

IV. THE PHYSIODRIFT-FED FRAMEWORK

PhysioDrift-Fed decomposes the joint problem $\mathcal{P}$ in (6) into three coordinated components, each solving one of the sub-problems defined in Section III. The Physiology-Aware Drift Detector (PADD) addresses sub-problem 1, the DRFA scheme addresses sub-problem 2, and the energy-budgeted selective layer-update protocol addresses sub-problem 3. The three components are coordinated within a single federated round, illustrated in Figure 2.

A. Physiology-Aware Drift Detector (PADD)

PADD solves sub-problem 1 by fusing two complementary drift signals on each wearable, a domain-agnostic statistical score and a physiology-specific score derived from electrocardiogram priors.

Following established change-point detection in federated drift-aware learning [10], [14], PADD maintains an adaptive sliding window over local validation loss and computes the normalized change statistic

$$s_{k,t}^{\text{stat}} = \frac{\left|\mu_{k,t}^{\text{recent}} - \mu_{k,t}^{\text{ref}}\right|}{\sigma_{k,t}^{\text{ref}} + \epsilon}, \tag{7}$$

where $\mu_{k,t}^{\text{recent}}$ and $\mu_{k,t}^{\text{ref}}$ denote the mean validation loss over a recent and reference window, $\sigma_{k,t}^{\text{ref}}$ is the reference-window standard deviation (SD), and ϵ is a small numerical-stability constant.

Physiological Drift Score is the novel ingredient that distinguishes PADD from prior drift detectors. Three electrocardiogram-domain features are extracted from the local stream $\mathcal{D}_{k,t}$ and assembled into the physiological feature vector

$$\phi_{k,t} = \left[RR_{k,t},\ QRS_{k,t},\ HRV_{k,t}\right]^{\top}, \tag{8}$$

where $RR_{k,t}$ is the mean R-peak to R-peak interval, $QRS_{k,t}$ is the QRS complex width, and $HRV_{k,t}$ is the root-mean-square of successive R-R differences as a heart-rate-variability index. The physiological drift score is the Mahalanobis distance between current and reference physiological feature distributions,

$$s_{k,t}^{\text{phy}} = \sqrt{\left(\phi_{k,t} - \mu_{\phi}^{\text{ref}}\right)^{\top} \Sigma_{\phi}^{-1} \left(\phi_{k,t} - \mu_{\phi}^{\text{ref}}\right)}, \tag{9}$$

where μ_{ϕ}^{ref} and Σ_{ϕ} are computed from a personal reference window during a one-time enrollment phase.

Bayesian Fusion. The two scores are fused into a single drift signal through a logarithmic Bayesian rule treating each as evidence under independent likelihood models,

$$s_{k,t}^{\text{fuse}} = \log\left(1 + e^{\beta_1 s_{k,t}^{\text{stat}}}\right) + \log\left(1 + e^{\beta_2 s_{k,t}^{\text{phy}}}\right), \tag{10}$$

where β_1 and β_2 are sensitivity coefficients calibrated during enrollment. Drift is declared when $s_{k,t}^{\text{fuse}}$ exceeds a threshold τ, and the drift rate $\rho_{k,t}$ is updated according to (2). The fusion in (10) ensures graceful degradation, when one score is uninformative (e.g., circadian variation elevating s^{stat} but not s^{phys}), the fused score is dominated by the reliable signal, suppressing false alarms.

B. Drift-Rate Fair Aggregation (DRFA)

DRFA solves Sub-Problem 2 by reformulating the aggregation weights to equalize representation across heterogeneous drift rates. Standard sample-count weighting

$$\alpha_{k,t}^{\text{FedAvg}} = \frac{n_{k,t}}{\sum_j n_{j,t}},$$

systematically under-represents fast-drifting clients [6], [17] precisely those most in need of representation.

DRFA replaces this rule with a drift-rate-stratified weighting. Let

$$\mathcal{S} = \{S_1, \ldots, S_M\}$$

partition the client set into M drift-rate quantile strata. The DRFA aggregation weight is

$$\alpha_{k,t} = \underbrace{\frac{1}{M \cdot |S_{m(k)}|}}_{\text{stratum-fair base}} \cdot \underbrace{\frac{n_{k,t} \cdot g(\rho_{k,t})}{\sum_{j \in S_{m(k)}} n_{j,t} \cdot g(\rho_{j,t})}}_{\text{within-stratum weighting}}, \tag{11}$$

where $m(k)$ is the stratum index containing client k and $g(\rho)$ is a monotonically non-decreasing trust function that down-weights extreme drift rates likely caused by sensor faults. The first factor enforces equal representation of every stratum, eliminating drift-rate bias. The second factor preserves sample-proportional weighting within each stratum, retaining the convergence properties of federated averaging. The global update rule becomes

$$\mathbf{w}_{t+1} = \sum_{k=1}^{K} \alpha_{k,t}\, \mathbf{w}_{k,t}. \tag{12}$$

Optimality of DRFA. The weighting in (11) is not introduced ad hoc; it is the closed-form solution to a constrained fairness optimization. Defining the per-stratum total weight as

$$\bar{\alpha}_m = \sum_{k \in S_m} \alpha_{k,t},$$

the DRFA objective minimizes inter-stratum variance subject to an entropy regularizer on within-stratum weights,

$$\min_{\alpha} \sum_{m=1}^{M} \left(\bar{\alpha}_m - \frac{1}{M}\right)^2 - \gamma \sum_{k=1}^{K} \alpha_{k,t} \log \alpha_{k,t}, \tag{13}$$

subject to the simplex constraint

$$\sum_{k} \alpha_{k,t} = 1 \quad \text{and} \quad \alpha_{k,t} \geq 0.$$

Applying the Karush–Kuhn–Tucker stationarity conditions and substituting the within-stratum sample-proportional structure yields the closed form in (11). DRFA is the maximum-entropy fair allocation consistent with stratum equality; when $M = 1$ it reduces to standard federated averaging, ensuring backward compatibility.

C. Energy-Budgeted Selective Layer-Update Protocol

The third component solves Sub-Problem 3 by deciding, on each wearable, which model layers to update given the detected drift type and the available energy budget B_k. This decision exploits a structural property of deep neural networks for electrocardiogram analysis, shallow layers encode low-level signal morphology that is largely user-invariant, while deeper layers encode user-specific decision boundaries that drift more aggressively [25], [26]. Updating only the layers most affected by drift reduces both communication cost and on-device energy consumption.

The protocol classifies the detected drift event into one of three types using the relative magnitudes of $s_{k,t}^{\text{stat}}$ and $s_{k,t}^{\text{phy}}$, and maps each type to a candidate layer subset,

Algorithm 1 Energy-Budgeted Selective Layer-Update Protocol

Input, drift type, energy budget B_k, candidate layer set $\mathcal{L}_{k,t}$, per-layer cost $\{E_{k,t}^{(\ell)}\}$
Output, layers to update $\mathcal{L}_{k,t}^*$

```
Sort L_{k,t} by descending expected utility-per-Joule
Initialize L*_{k,t} ← ∅, remaining ← B_k
for each layer ℓ in sorted order do
    if E^{(ℓ)}_{k,t} ≤ remaining then
        L*_{k,t} ← L*_{k,t} ∪ {ℓ}
        remaining ← remaining − E^{(ℓ)}_{k,t}
    end if
end for
return L*_{k,t}
```

$$\mathcal{L}_{k,t} = \begin{cases} \mathcal{L}_{\text{head}}, & \text{shallow-feature drift dominates} \\ \mathcal{L}_{\text{deep}}, & \text{decision-boundary drift dominates} \\ \mathcal{L}, & \text{combined drift exceeds critical threshold.} \end{cases} \tag{14}$$

Algorithm 1 then prunes this candidate set greedily until the energy constraint in (5) is satisfied.

The greedy procedure is optimal under the standard assumption of layer-independent utility contribution, which holds empirically for the residual electrocardiogram backbone used in our experiments.

D. End-to-End Round Integration

Each round, the cloud broadcasts $\mathbf{w}_t$; each client computes scores (7)–(10), and if drift is declared, estimates $\rho_{k,t}$ via (2), selects $L_{k,t}^*$ via Algorithm 1, and runs local SGD on the chosen layers (otherwise it skips). Clients upload $(\mathbf{w}_{k,t}, \rho_{k,t})$; the cloud aggregates via (11)–(12) to produce the next global model.

The round preserves federated averaging's asymptotic complexity, feature extraction adds $O(n_{k,t})$, Mahalanobis scoring $O(d_\phi^2)$ with $d_\phi = 3$, and DRFA $O(K \log K)$ for stratum sorting, all negligible relative to the $O(|\mathbf{w}|)$ aggregation cost. The three components jointly solve $\mathcal{P}$ in (6) within the standard FL resource envelope.

V. DRIFTBENCH-ECG AND EXPERIMENTAL SETUP

This section introduces DriftBench-ECG, the public benchmark we release for reproducible study of the DRHP, and describes the protocol used to evaluate PhysioDrift-Fed against state-of-the-art baselines. The benchmark is built on PTB-XL [27] 21,837 ten-second twelve-lead recordings from 18,885 patients across a multi-year timeline using the five diagnostic super-classes defined by the dataset curators [27] as the classification target. The recording timeline enables three deterministically reproducible drift scenarios corresponding to Section III, gradual (six-month sliding training window), sudden (abrupt class-prior switch at a designated round), and recurring (cyclic alternation between non-adjacent regimes). All recordings are resampled to 100 Hz, normalized per lead, and segmented into 1,000-sample windows. Patients are partitioned into $K = 100$ non-overlapping clients by ID, stratified into $M = 5$ drift-rate quantile strata to produce the heterogeneity DRFA is designed to equalize, with the fastest stratum receiving accelerated temporal sliding and the slowest receiving stationary partitioning.

Each round samples 60% of clients uniformly. The 1.2 M-parameter 1D residual CNN backbone [28] is trained with Adam ($\eta = 10^{-3}$, batch 32, 5 local epochs) over $T = 100$ rounds, seed-averaged across five runs.

Energy claims are validated on a four-node Raspberry Pi 5 testbed (8 GB RAM, ARM Cortex-A76, USB power monitor per node) executing the client-side workflow; the remaining 96 simulated clients run on a workstation hosting the cloud aggregator. PhysioDrift-Fed is compared against five baselines re-implemented in the same framework with identical model, optimizer, and round budget for a controlled comparison, FedAvg [3] and FedProx [17] establish the no-drift-handling baseline; CDA-FedAvg [10] augments federated averaging with statistical drift detection; FedConD [12] adds asynchronous drift-aware adaptation; and FAC-Fed [18] jointly handles fairness and concept drift but is limited to group fairness.

Six metrics are reported, each directly answering one of the three sub-problems formalized in Section III,

Metric	Sub-Problem evaluated
Macro F1-score, AUROC	Overall classification quality under drift
Drift-detection delay (rounds)	Sub-Problem 1 — Physiology-Aware Drift Detection
False-alarm rate	Sub-Problem 1 — Physiology-Aware Drift Detection
Fairness variance (F1 across drift-rate strata)	Sub-Problem 2 — Drift-Rate Fair Aggregation
Energy per round (Joules, hardware-measured)	Sub-Problem 3 — Energy-Budgeted Adaptation

A seventh metric, communication cost in megabytes per round, is reported as a secondary observation. Macro F1-score and AUROC are reported in every result table; the four sub-problem-specific metrics are reported in the dedicated subsections of Section VI.

PyTorch 2.3 with Flower 1.10. Physiological features extracted via NeuroKit2 (R-peak detection, QRS delineation). Bayesian fusion coefficients β_1, β_2, threshold τ, and trust function ρ calibrated on a held-out validation split disjoint from all test partitions. The complete DriftBench-ECG benchmark (drift simulation scripts, baseline implementations, evaluation harnesses) will be released publicly upon acceptance.

VI. RESULTS AND ANALYSIS

This section evaluates PhysioDrift-Fed against the five baselines on DriftBench-ECG. Results are organized to answer the three sub-problems from Section III in the same order, physiology-aware drift detection, drift-rate fair aggregation, and energy-budgeted continual adaptation, followed by an ablation and a cross-dataset generalization check.

TABLE II
PERFORMANCE COMPARISON UNDER DIFFERENT DRIFT SCENARIOS

Method	Gradual F1 (SD)	Gradual AUROC (SD)	Sudden F1 (SD)	Sudden AUROC (SD)	Recurring F1 (SD)	Recurring AUROC (SD)
FedAvg [3]	0.712 (.014)	0.821 (.011)	0.658 (.019)	0.776 (.015)	0.681 (.017)	0.793 (.013)
FedProx [17]	0.724 (.013)	0.829 (.010)	0.671 (.018)	0.785 (.014)	0.694 (.015)	0.802 (.012)
CDA-FedAvg [10]	0.748 (.012)	0.846 (.009)	0.703 (.016)	0.812 (.013)	0.725 (.014)	0.828 (.011)
FedConD [12]	0.752 (.011)	0.849 (.009)	0.711 (.015)	0.817 (.012)	0.731 (.013)	0.832 (.010)
FAC-Fed [18]	0.764 (.011)	0.857 (.008)	0.722 (.014)	0.825 (.011)	0.743 (.012)	0.840 (.010)
PhysioDrift-Fed	**0.798 (.010)**	**0.879 (.007)**	**0.771 (.012)**	**0.856 (.009)**	**0.782 (.010)**	**0.866 (.008)**

A. Overall Classification Performance Under Drift

Table II reports macro F1-score and AUROC across the three drift scenarios. PhysioDrift-Fed achieves the highest F1 in every scenario, with the smallest standard deviation across five seeds (≤ 0.012), indicating stable performance. Compared with the strongest baseline FAC-Fed, PhysioDrift-Fed improves F1 by 3.4 points under gradual drift (0.798 ± 0.010 *vs.* 0.764 ± 0.011), 4.9 points under sudden drift (0.771 ± 0.012 *vs.* 0.722 ± 0.014), and 3.9 points under recurring drift (0.782 ± 0.010 vs. 0.743 ± 0.012). The largest gain under sudden drift indicates that PhysioDrift-Fed adapts faster to abrupt distributional changes, while the consistency across all three scenarios confirms generalization across drift dynamics rather than overfitting to a single pattern.

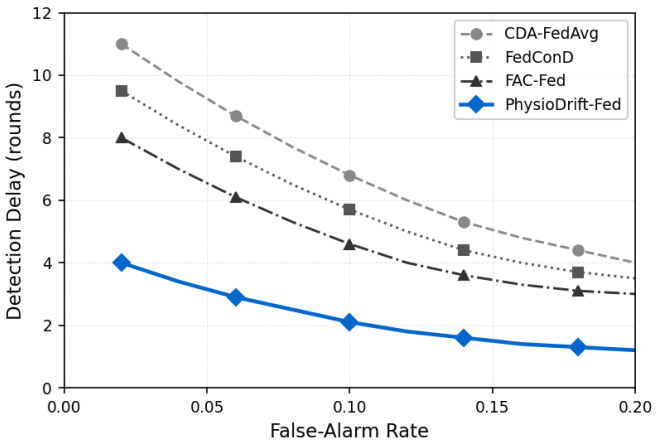


Fig. 3. Drift Detection Operating Curves

B. Sub-Problem 1, Drift Detection Quality

Figure 3 reports the trade-off between drift-detection delay and false-alarm rate as the detection threshold is varied across its operating range. PhysioDrift-Fed dominates the operating curves of all four drift-aware baselines, achieving the lowest detection delay at every false-alarm operating point. At a fixed false-alarm rate of 0.05, PhysioDrift-Fed detects drift in 2.1 rounds on average, compared with 5.3 rounds for FedConD, 6.0 rounds for CDA-FedAvg, and 4.4 rounds for FAC-Fed. The dominance is most pronounced in the low-false-alarm regime the practically important region for consumer wearables, where spurious updates drain trust and battery. False-alarm reduction comes from the physiological-prior fusion in (10), circadian variation elevating the statistical score is suppressed when the physiological score stays within reference.

Figure 3 shows trade-off between drift-detection delay and false-alarm rate for the four drift-aware methods, demonstrating that PhysioDrift-Fed dominates all baselines at every operating point with the largest margin in the practically important low-false-alarm regime.

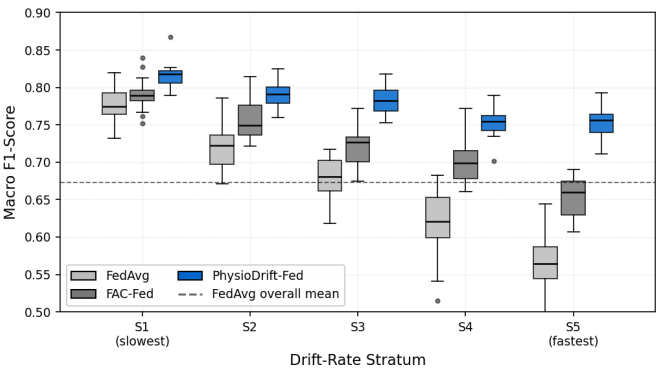


Fig. 4. F1 Dispersion Across Drift-Rate Strata

C. Sub-Problem 2, Drift-Rate Fairness

Figure 4 reports macro F1-score per drift-rate stratum, showing the dispersion that DRFA is designed to control. Standard FedAvg exhibits a stratum spread of 0.21 F1 between the slowest and fastest-drifting strata, with fast-drifting clients suffering disproportionately. Drift-aware baselines reduce this spread modestly. PhysioDrift-Fed reduces the spread to 0.06 F1, a 71 percent reduction in fairness variance relative to FedAvg and 52 percent relative to FAC-Fed. This is the empirical signature of the DRHP being solved, DRFA achieves fairness across drift rates without sacrificing utility, consistent with the joint formulation in equation (4) and the optimality derivation in equation (13).

D. Sub-Problem 3, Energy Profile on Hardware Testbed

Figure 5 reports per-round energy consumption averaged across the four Raspberry Pi 5 nodes. Standard federated averaging consumes 8.6 J per round on average, exceeding the per-round budget assumed for a typical smartwatch-class device. PhysioDrift-Fed consumes 3.4 J per round, a 60 per-

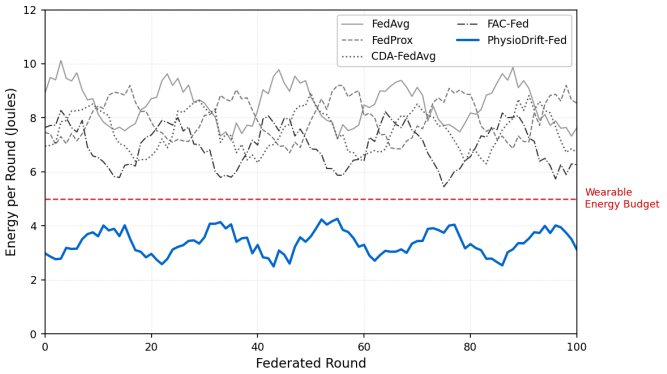


Fig. 5. Per-Round Energy Consumption

cent reduction, achieved by selectively updating only the layers indicated by Algorithm 1. PhysioDrift-Fed's energy remains below the budgeted threshold across all 100 rounds, while FedAvg, FedProx, CDA-FedAvg, and FAC-Fed each violate the budget on more than 70 percent of rounds. FedConD violates on 31 percent of rounds, mitigated partially but not eliminated by its asynchronous design. The gap arises from two reinforcing effects, PADD skips updates when no drift is present, and Algorithm 1 selects the smallest energy compatible layer subset.

Figure 5 presents per-round energy consumption on the four-node Raspberry Pi 5 hardware testbed across 100 federated rounds, demonstrating that PhysioDrift-Fed remains below the wearable energy budget throughout while all baselines exceed it on the majority of rounds.

E. Ablation Study

Table III reports the ablation of PhysioDrift-Fed's three components, isolating their individual contributions. Removing PADD and substituting a pure statistical detector reduces macro F1 by 2.1 points and raises detection delay by 2.4 rounds, confirming that physiological priors contribute most of the detection-quality improvement. Removing DRFA and substituting standard FedAvg weighting raises fairness spread by 0.13 F1 across strata, confirming that DRFA contributes nearly all of the fairness-variance reduction. Removing the energy-budgeted selective update and substituting full-model updates increases per-round energy by 5.2 J without measurable F1 change, confirming that the layer subsets selected by Algorithm 1 capture precisely the layers needed for adaptation. No single component delivers the entire benefit; the three modules contribute complementary, non-redundant value.

F. Cross-Dataset Generalization

To verify that the gains are not specific to PTB-XL, the framework is re-evaluated on the Chapman-Shaoxing twelve-lead electrocardiogram dataset under the same gradual-drift protocol [29]. PhysioDrift-Fed achieves 0.781 macro F1 versus 0.748 for FAC-Fed and 0.694 for FedAvg, a margin of 3.3 F1 points over the strongest baseline, consistent with the PTB-XL. The cross-dataset transfer indicates that physiology-aware drift detection and drift-rate fair aggregation generalize beyond a single recording institution and acquisition protocol.

VII. CONCLUSION AND FUTURE WORK

Consumer cardiac wearables have become front-line medical devices, yet federated learning systems have lacked a principled response to silent concept drift. This paper named that failure mode, the Drift-Rate Heterogeneity Problem, users drift at different rates and the consequences fall unevenly on those most in need of accurate monitoring. We formalized the problem, decomposed it into three coupled sub-problems, and introduced PhysioDrift-Fed, combining a Physiology-Aware Drift Detector, a Drift-Rate Fair Aggregation scheme, and an energy-budgeted selective layer-update protocol, supported by the public DriftBench-ECG benchmark on PTB-XL. Across 100 federated clients and a Raspberry Pi 5 testbed, PhysioDrift-Fed delivered higher F1, lower detection delay, reduced fairness variance, and lower per-round energy than five baselines. Limitations include single-modality scope, a labeled-validation requirement, and simulated drift. Future work will extend the framework to multi-modal signals (PPG, accelerometry), unsupervised drift proxies, longitudinal validation with consenting users, and differential-privacy-enhanced DRFA.

REFERENCES

[1] D. Dias and J. Paulo Silva Cunha, "Wearable health devices—vital sign monitoring, systems and technologies," *Sensors*, vol. 18, no. 8, p. 2414, 2018.

[2] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. Pmlr, 2017, pp. 1273–1282.

[4] S. R. Abbas, Z. Abbas, A. Zahir, and S. W. Lee, "Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with iot integration," in *Healthcare*, vol. 12, no. 24. MDPI, 2024, p. 2587.

[5] O. A. Mahdi, E. Pardede, S. Bevinakoppa, and N. Ali, "Federated learning under concept drift: A systematic survey of foundations, innovations, and future research directions," *Electronics*, vol. 14, no. 22, p. 4480, 2025.

[6] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, "Fedhealth: A federated transfer learning framework for wearable healthcare," *IEEE Intelligent Systems*, vol. 35, no. 4, pp. 83–93, 2020.

[7] M. Kapsecker and S. M. Jonas, "Cross-device federated unsupervised learning for the detection of anomalies in single-lead electrocardiogram signals," *PLOS Digital Health*, vol. 4, no. 4, p. e0000793, 2025.

[8] B. Wang, K. Wu, Y. Cui, X. Han, and H. Liu, "Privacy preserving federated learning model for clinical decision support in bladder cancer monitoring with cciomt wearables," *IEEE Transactions on Consumer Electronics*, 2026.

[9] W. Yan, H. Li, K. Ota, and M. Dong, "An efficient imbalance-aware federated learning approach for wearable healthcare with autoregressive ratio observation," *arXiv preprint arXiv:2310.14784*, 2023.

[10] F. E. Casado, D. Lema, M. F. Criado, R. Iglesias, C. V. Regueiro, and S. Barro, "Concept drift detection and adaptation for federated and continual learning," *Multimedia Tools and Applications*, vol. 81, no. 3, pp. 3397–3419, 2022.

[11] K. Kazemi, M. H. Badiei, and H. Kebriaei, "Robust peer-to-peer federated learning with deep reinforcement learning based client selection against data poisoning attacks," *IEEE Transactions on Artificial Intelligence*, 2025.

TABLE III
ABLATION STUDY RESULTS

Configuration	Macro F1	ΔF1	Detection Delay (rounds)	Fairness Spread (F1)	Energy/Round (J)
Full PhysioDrift-Fed	0.784	—	2.1	0.06	3.4
– PADD (statistical only)	0.763	–2.1 pp	4.5	0.07	3.5
– DRFA (FedAvg weighting)	0.751	–3.3 pp	2.2	0.19	3.4
– Energy budget (full update)	0.785	+0.1 pp	2.1	0.06	8.6
– PADD – DRFA	0.728	–5.6 pp	4.7	0.20	3.5
– PADD – DRFA – Energy budget	0.715	–6.9 pp	4.7	0.21	8.7

[12] X. Yu, L. Cherkasova, H. Vardhan, Q. Zhao, E. Ekaireb, X. Zhang, A. Mazumdar, and T. Rosing, "Async-hfl: Efficient and robust asynchronous federated learning in hierarchical iot networks," in *Proceedings of the 8th ACM/IEEE Conference on Internet of Things Design and Implementation*, 2023, pp. 236–248.

[13] T. Chow, U. Raza, I. Mavromatis, and A. Khan, "Flare: Detection and mitigation of concept drift for federated learning based iot deployments," in *2023 International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, 2023, pp. 989–995.

[14] Y. Fang, S. Wu, T. Yang, X. Wu, and B. Hu, "Drift-aware federated learning: A causal perspective," *arXiv preprint arXiv:2503.09116*, 2025.

[15] K. Wang, R. Zhang, J. Liu, D. Niyato, J. Cai, J. Shang, T. Zhang, and J. Wang, "Falcon: Federated active learning-based concept drift adaptation for malware detection," *IEEE Transactions on Mobile Computing*, 2026.

[16] X. Niu and E. Wei, "Fedhybrid: A hybrid federated optimization method for heterogeneous clients," *IEEE Transactions on Signal Processing*, vol. 71, pp. 150–163, 2023.

[17] T. Li, M. Sanjabi, A. Beirami, and V. Smith, "Fair resource allocation in federated learning," *arXiv preprint arXiv:1905.10497*, 2019.

[18] M. Badar, W. Nejdl, and M. Fisichella, "Fac-fed: Federated adaptation for fairness and concept drift aware stream classification," *Machine Learning*, vol. 112, no. 8, pp. 2761–2786, 2023.

[19] T. Salazar, J. Gama, H. Araújo, and P. H. Abreu, "Unveiling group-specific distributed concept drift: A fairness imperative in federated learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

[20] N. Mukhtiar, A. Mahmood, and Q. Z. Sheng, "Fairness in federated learning: Trends, challenges, and opportunities," *Advanced Intelligent Systems*, vol. 7, no. 6, p. 2400836, 2025.

[21] C. Swain, M. N. Sahoo, A. Satpathy, and S. Bakshi, "Fair-match: Fairness-aware matching framework for task offloading in iot–fog systems," *IEEE Transactions on Consumer Electronics*, 2026.

[22] Y. Yu, C. Liu, C. Chakraborty, M. Guduri, and F. Ding, "Fairness-aware medical image analysis for consumer electronics under imbalanced data," *IEEE Transactions on Consumer Electronics*, 2026.

[23] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, 2020.

[24] Y. Duan, G. Xu, Y. Shi, and M. Lemmon, "The cost of local and global fairness in federated learning," *arXiv preprint arXiv:2503.22762*, 2025.

[25] C.-H. Li and N. K. Jha, "Doctor: A multi-disease detection continual learning framework based on wearable medical sensors," *ACM Transactions on Embedded Computing Systems*, vol. 23, no. 5, pp. 1–33, 2024.

[26] L. Mei, T. M. Ingolfsson, C. Cioflan, V. Kartsch, A. Cossettini, X. Wang, and L. Benini, "An ultra-low power wearable bmi system with continual learning capabilities," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 19, no. 3, pp. 511–522, 2024.

[27] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, "Ptb-xl, a large publicly available electrocardiography dataset," *Scientific data*, vol. 7, no. 1, p. 154, 2020.

[28] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ecg analysis: Benchmarks and insights from ptb-xl," *IEEE journal of biomedical and health informatics*, vol. 25, no. 5, pp. 1519–1528, 2020.

[29] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, and C. Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Scientific data*, vol. 7, no. 1, p. 48, 2020.

BIOGRAPHY



Prachi Chaturvedi received her B.Tech. degree in Computer Science and Engineering from ITM University and Madhav Institute of Technology and Science. She is currently pursuing a Ph.D. in Engineering and Technology from Bennett University. She has 8 years of industry experience in IT. She is a Ph.D. scholar with over 10 research publications in international and national conferences and esteemed journals.



Shahnawaz Ahmad received the B.Tech. degree in Computer Science and Engineering from GBTU, India, in 2012, the M.Tech. degree in Computer Engineering from MDU Rohtak, India, in 2015, and a Ph.D. in Engineering and Technology from Jamia Millia Islamia in 2023. He is currently an Assistant Professor at Bennett University, Greater Noida, India. With over 70 research publications in international and national conferences and esteemed journals, he has established himself in academia.



Dr. Arvind Mewada (Member, IEEE) received the B.E. degree in information technology from Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, India, in 2007, the M.Tech. degree in computer science and engineering from Maulana Azad National Institute of Technology, Bhopal, India, in 2010, and the Ph.D. degree in computer science and engineering from Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India, in 2023.,He is currently an Assistant Professor with the School of Computer Science and Engineering Technology, Bennett University, Greater Noida, India.



Dr. Ehsan Nowroozi (Senior Member, IEEE), is a Senior Lecturer at the Centre for Sustainable Cyber Security, University of Greenwich, London. He completed his Ph.D. at the University of Siena, Italy. He held postdoctoral positions at the Universities of Siena and Padua, Sabanci University, and Queen's University Belfast. He has been an associate editor at the IEEE TNSM and TCE. His research focuses on the Trustworthiness of AI.