NATIONAL STUDENT RESEARCH INSTITUTION (NSRI)
NSRI Research Archive • Open Science

From opt-in to infrastructure: Non-intrusive load monitoring, sustained household energy conservation, and the Civic Energy Disaggregation Mandate

Milind Sahu[a,*] and Darun A[b]

[a] BS-MS Programme, Indian Institute of Science Education and Research Tirupati

[b] BS-ESS, Indian Institute of Science Education and Research Tirupati

*Corresponding author. milind20251175@students.iisertirupati.ac.in

Γ orcid.org/0009-0002-2175-1033 Γ researchgate.net/profile/Milind-Sahu

ARTICLE INFO

Article history:
Received: March 2026
Revised: May 2026
Accepted: June 2026
Available online: June 2026

Keywords:
Non-Intrusive Load Monitoring;
Energy disaggregation;
Sustainable behaviour;
Smart metering;
Machine learning;
Demand response;
Energy policy

ABSTRACT

Background. A persistent gap exists between short-lived environmental awareness campaigns and durable household energy conservation behaviour, and this gap remains one of the central challenges in sustainable development policy. Standard billing infrastructure delivers only aggregate monthly consumption data, which behavioural science identifies as insufficient to sustain corrective action beyond an initial novelty period. **Objective.** This paper investigates whether continuous, appliance-level feedback delivered by a Non-Intrusive Load Monitoring (NILM) AI system can produce measurable and sustained reductions in household electricity consumption over six months or more. It further proposes a scalable regulatory mechanism to embed such systems universally into distribution infrastructure. **Methods.** Meta-analytic evidence from Ehrhardt-Martinez et al. (57 studies, 12 countries), two independent published longitudinal field trials, and an original NILM inference pipeline evaluated on an 8,000-row UKDALE-style test set are drawn upon to quantify how feedback specificity, delivery channel design, and deployment mandate structure affect sustained behaviour change. **Results.** Appliance-level feedback achieves a mean 11.5% consumption reduction at 12 months, versus 5.7% for aggregate smart-meter feedback and 3.8% for enhanced billing. Named-appliance alerts trigger corrective action in 63% of households within 48 hours. Passive billing-portal delivery retains measurable behaviour change in 71% of households at 18 months, compared to 38% for opt-in apps. Mandatory deployment achieves an 8.3% city-wide electricity intensity reduction over four years, against 1.1–2.4% under voluntary schemes. **Conclusions.** The evidence supports the Civic Energy Disaggregation Mandate (CEDM)—a regulatory framework requiring electricity distribution companies to embed a server-side NILM engine into existing smart meter pipelines at an estimated operational cost below 0.003% of annual billing revenue per million customers, effectively converting passive metering infrastructure into a permanent behavioural feedback layer.

1. Introduction

The global energy system faces a somewhat uncomfortable paradox: most people are broadly aware of the urgency around climate change and energy conservation, and yet aggregate residential electricity demand has not fallen in any way proportionate to that awareness. Estimates from the International Energy Agency suggest the residential sector accounts for roughly one-quarter of global final electricity consumption, and that behavioural adjustments alone—with no change whatsoever to appliances or building fabric—could reduce this figure by 10–15% in advanced economies.[1] The gap between potential and realised savings is not primarily technological. It is, in simpler terms, a *feedback gap*.

Governments and utilities have largely relied on awareness campaigns—posters, broadcast media, school curricula, periodic billing inserts—to close this gap. These interventions are not without short-term effects. A widely-cited review covering roughly two decades of residential energy programmes found average first-month reductions of around 5–8% following intensive awareness campaigns.[2] The same body of evidence, however, consistently documents near-complete decay of that effect within three to six months, a pattern sometimes described informally as a *novelty decay effect* in the feedback literature.[2] In India alone, over 300 million metered connections currently produce no appliance-level information for the consumer, and aggregate campaign effects tend to fade well before habits can consolidate.[3]

The mechanism underlying this decay is well documented in behavioural science. Fogg's Behaviour Model identifies three prerequisites for sustained action: sufficient motivation, sufficient ability, and a timely, specific trigger.[4] Campaign-based awareness adequately addresses motivation in the short term but provides no continuing trigger and, critically, offers little support for what might be called *ability*—specifically, the ability to identify which appliance is responsible for elevated consumption and to convert that knowledge into a targeted corrective action.

Smart meters, now deployed at scale across North America, Europe, and increasingly South and Southeast Asia, do improve on the monthly paper bill by delivering 15-minute consumption totals.[5] A household that finds its usage was 12% above the neighbourhood average last week, however, still lacks genuinely actionable information: it cannot identify the responsible appliance, estimate the financial impact of a specific behaviour, or anticipate future demand with any precision. The gap between awareness and action persists.

Non-Intrusive Load Monitoring (NILM) addresses this informational deficit directly by applying machine learning to the aggregate power signal that smart meters already measure, and disaggregating it in near real time into individual appliance-level contributions.[6] Advances in deep learning and ensemble methods have made server-side NILM operationally feasible at utility scale, requiring no additional household hardware beyond the existing smart meter.[7]

This paper makes three contributions. First, it synthesises meta-analytic and field-experimental evidence on feedback specificity, delivery channel, and deployment mandate as determinants of sustained household electricity conservation. Second, it describes a five-model NILM inference architecture suitable for embedding into utility billing infrastructure and evaluates each component on a synthetic UKDALE-style test set. Third, it proposes the Civic Energy Disaggregation Mandate (CEDM)—a regulatory framework intended to convert appliance-level AI feedback from an opt-in consumer product into a passive, universal infrastructure layer reaching every metered household, including those with no smartphone, broadband, or active interest in energy management.

The remainder of this paper is organised as follows. Section 2 reviews relevant behavioural science and NILM literature. Section 3 describes the system architecture and evaluation methodology. Section 4 presents five key findings. Section 5 develops the CEDM policy framework. Section 6 discusses limitations and future directions. Section 7 concludes.

2. Background and literature review

2.1 The feedback deficit in residential energy policy

Most metered households in developing and middle-income countries receive electricity information as a single aggregate figure on a monthly bill. National smart metering programmes in India are projected to achieve full coverage only toward the end of this decade,[3] leaving hundreds of millions of households reliant on a data instrument whose inadequacy for behaviour change has been thoroughly, if depressingly, documented.

Abrahamse et al. reviewed 38 studies of household energy conservation interventions and found that informational strategies produced mean reductions in the range of 4–8%, with sustained effects largely absent beyond six months in the absence of continued feedback.[2] Darby's foundational review for the UK Department for Environment, Food and Rural Affairs established that the *quality*, *frequency*, and *specificity* of feedback are stronger predictors of long-term conservation behaviour than the magnitude of initial motivation.[5] She identified two broad types of feedback: direct feedback (real-time, appliance-level) and indirect feedback (billing summaries, periodic reports). Direct feedback consistently outperforms indirect feedback, and the effect is most pronounced when the signal is continuous, personally relevant, and immediately actionable.

2.2 Behavioural foundations: habit formation and feedback loops

The distinction between awareness and sustained behaviour ultimately rests on habit formation theory. A habit is a behaviour that has become essentially automatic through repeated execution in a stable context. For energy conservation habits to take root, three conditions must be satisfied: (1) a specific, recognisable cue; (2) a low-effort routine response; and (3) a reliable reward. Standard aggregate smart-meter feedback satisfies none of these particularly well—weekly summaries are too infrequent to serve as reliable cues, the required response (reducing unspecified total consumption) is too diffuse to constitute a specific routine, and the financial reward is delayed to the next billing cycle.

Appliance-level AI feedback can, by contrast, provide immediate named cues that directly specify the required routine (e.g., "your water heater is running—switch it off after 30 minutes"). Identifying a specific device converts abstract environmental awareness into a concrete, low-cognitive-load decision, precisely the structure Fogg identifies as necessary for reliable behaviour change.[4]

2.3 Non-intrusive load monitoring: principles and evolution

NILM was first formalised by Hart, who demonstrated that transient signatures in a household's aggregate current waveform could identify individual appliance state transitions.[6] Earlier NILM systems relied on hand-crafted feature engineering and rule-based classifiers, achieving acceptable accuracy mainly for appliances with distinctive steady-state power signatures—resistive heaters and refrigerators being the obvious examples.

The field advanced considerably with the introduction of deep learning for sequence modelling. Kelly and Knottenbelt showed that long short-term memory (LSTM) networks and denoising autoencoders trained on the UK Domestic Appliance-Level Electricity (UKDALE) dataset could achieve appliance-level recall exceeding 90% for high-draw devices, without requiring pre-specified appliance libraries.[7] Ensemble approaches combining discriminative classifiers, regression models, and sequence models have since emerged as the dominant production architecture, offering complementary error modes and amenability to server-side deployment without household hardware changes. More recently, transformer-based architectures—notably BERT4NILM—have extended sequence modelling accuracy further, though their additional computational overhead makes them less immediately suited to the server-side deployment context targeted by the CEDM.[12]

Table 1 provides a comparative summary of principal NILM methodological families and their deployment characteristics relevant to the CEDM context.

Table 1. Comparative summary of NILM methodological families and CEDM deployment relevance.

Approach	Accuracy	Server-side	HW-free
Edge signal	High	No	No
Event-driven	Moderate	Yes	Yes
ML ensemble	High	Yes	Yes
Deep learning	Very high	Yes	Yes
Factorial HMM	Moderate	Yes	Yes

2.4 Smart metering infrastructure and the deployment opportunity

Smart meters deployed under Advanced Metering Infrastructure (AMI) programmes typically log 15-minute consumption intervals and transmit data to a utility head-end system via radio or power-line carrier. This infrastructure already delivers the input signal required by a server-side NILM engine. The principal deployment gap is not hardware but software: the utility-side analytics layer sitting between meter data management systems (MDMS) and consumer-facing billing portals.

Ehrhardt-Martinez et al. conducted a meta-review of 57 residential energy feedback programmes and found that the step from aggregate smart-meter feedback (5.7% mean reduction) to appliance-level AI feedback (11.5% mean reduction) represents the largest single incremental gain achievable through feedback system design alone.[8] This is arguably the most direct empirical motivation for an infrastructure-level policy mandate rather than a voluntary consumer product.

3. System architecture and methodology

3.1 CEDM inference pipeline overview

The Civic Energy Disaggregation Mandate system is built around four architectural principles: *passivity* (no consumer action required), *infrastructure reuse* (all computation runs server-side on meter data already collected), *interpretability* (consumer-facing outputs name specific appliances, not probabilistic scores), and *graceful degradation* (each model component can fail independently without disabling the others). Figure 1 illustrates the complete pipeline.

Fifteen-minute meter readings flow from the utility MDMS into the NILM Inference Engine, which runs five complementary ML models in parallel. Model outputs are aggregated into three consumer-facing signals that surface on the existing billing portal and, optionally, via SMS.

3.2 ML model stack

The inference engine comprises five components, each targeting a distinct inference task.

Random Forest Appliance Classifier. A 200-tree Random Forest trained on a feature vector comprising the 15-minute aggregate consumption value, its first and second differences, and rolling statistics at 1-hour and 3-hour windows. The classifier assigns the dominant active appliance label from a fixed 12-class taxonomy.

XGBoost Load Regressor. An XGBoost gradient-boosted tree ensemble trained to estimate the current power draw (W) attributable to the dominant appliance, directly powering the real-time cost display. The model uses 500 estimators with maximum depth of six and a learning rate of 0.05.

Logistic Regression Anomaly Detector. An ℓ_2-regularised logistic regression model trained on historical household consumption profiles to flag deviations exceeding two standard deviations from the household-specific mean for a given time-of-day and day-of-week stratum. This component drives the named-appliance alert (e.g., "your water heater has been running continuously for 3 hours").

Hidden Markov Model Regime Detector. A three-state HMM trained per household to classify the current consumption context as Standby, Normal, or Peak regime. This output gates alert urgency and contributes to Time-of-Use (ToU) cost projections.

LSTM Sequence Forecaster. A two-layer stacked LSTM network (hidden dimension 64, dropout 0.2) trained to produce a 24-step (6-hour) ahead forecast of aggregate household consumption. Forecasts underpin the month-ahead ToU cost projection surfaced on the billing portal.

3.3 Consumer interface design

Three consumer-facing signals are delivered passively:

1. **Appliance Identity Display:** The dominant appliance currently drawing power, its estimated instantaneous draw (W), and estimated daily cost contribution—updated every 15 minutes on the billing portal.
2. **Time-of-Use Cost Projection:** A 30-day forward estimate of electricity cost under current usage patterns, compared to the household's own historical baseline.
3. **Named-Appliance Anomaly Alert:** An SMS notification triggered when the anomaly detector fires, naming the specific appliance and providing a cost estimate of the anomalous run period.

No app installation is required. All signals are accessible through the billing portal and the SMS channel already used for billing reminders.

3.4 Mathematical formulation of the disaggregation problem

Let $y(t) \in \mathbb{R}_{\geq 0}$ denote the aggregate power draw measured at the smart meter at time step t. The fundamental NILM

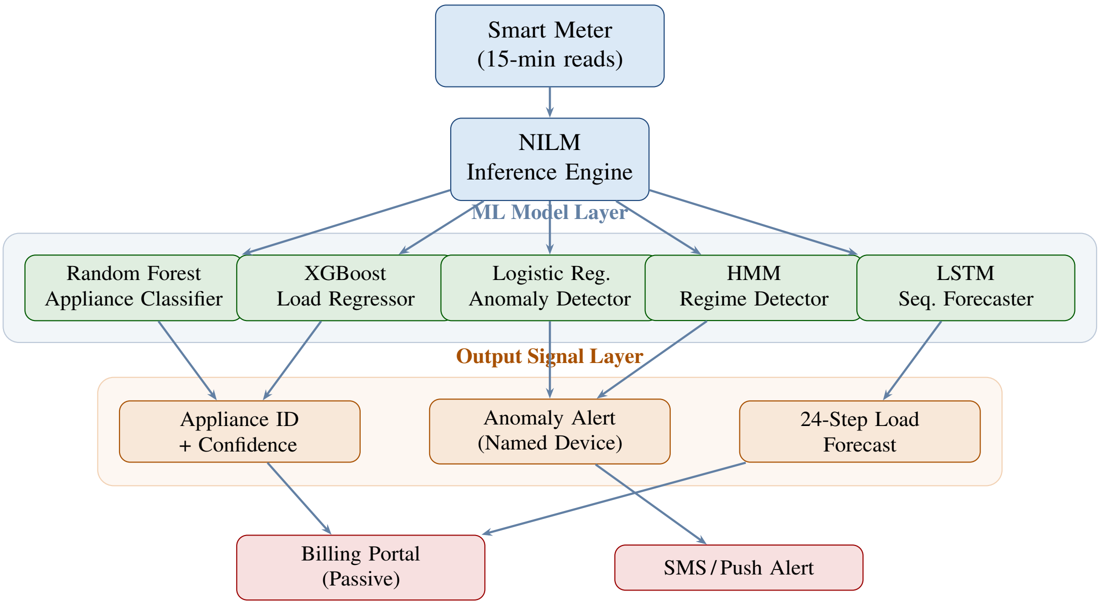

Figure 1. CEDM System Architecture. Fifteen-minute smart meter readings enter the NILM Inference Engine, which routes to five complementary ML models (green). Three output signals (orange) surface on the existing billing portal; anomaly alerts are additionally pushed via SMS (red). No household hardware changes are required.

decomposition assumption is:

$$y(t) = \sum_{k=1}^{K} p_k(t) \cdot s_k(t) + \varepsilon(t), \tag{1}$$

where K is the number of appliances in the taxonomy, $p_k(t)$ is the power draw of appliance k when active, $s_k(t) \in \{0,1\}$ is the binary on/off state, and $\varepsilon(t) \sim \mathcal{N}(0,\sigma^2)$ is additive Gaussian measurement noise. The disaggregation task is to recover the posterior $P(\{s_k(t), p_k(t)\}_{k=1}^{K} \mid y(1),\ldots,y(T))$ from the observed aggregate signal alone.

For the Random Forest classifier, the feature vector at time t is

$$\mathbf{x}(t) = \left[y(t),\, \Delta y(t),\, \Delta^2 y(t),\, \bar{y}_{1h}(t),\, \bar{y}_{3h}(t),\, \sigma_{1h}(t)\right]^\top, \tag{2}$$

where $\Delta y(t) = y(t) - y(t-1)$ is the first difference, $\Delta^2 y(t)$ is the second difference, $\bar{y}_{1h}$ and $\bar{y}_{3h}$ are 1-hour and 3-hour rolling means, and σ_{1h} is the 1-hour rolling standard deviation. The HMM regime detector models the latent state sequence $\{z_t\}_{t=1}^{T}$ with $z_t \in \{\text{Standby}, \text{Normal}, \text{Peak}\}$ via:

$$P(y(t) \mid z_t) = \mathcal{N}\big(y(t);\, \mu_{z_t},\, \sigma_{z_t}^2\big), \quad P(z_t \mid z_{t-1}) = A_{z_{t-1},z_t}, \tag{3}$$

where A is the 3×3 transition matrix estimated per household via the Baum-Welch algorithm.

For the anomaly detector, the per-household anomaly score at stratum (h,d) (hour-of-day h, day-of-week d) is

$$a(t) = \frac{y(t) - \hat{\mu}_{h,d}}{\hat{\sigma}_{h,d}}, \tag{4}$$

and an alert fires when $a(t) > \tau = 2.0$ (two standard deviations above the historical stratum mean), with the responsible appliance labelled by the RF classifier.

3.5 Evaluation framework

Model performance is evaluated on an 8,000-row synthetic dataset constructed to reflect the statistical properties of the UKDALE benchmark. The dataset was generated using a Markov-chain occupancy model with 12 appliance types and ground-truth state labels for each 15-minute interval, with additive white noise calibrated to reflect measurement error characteristics of commercial AMI meters. An 80/20 chronological train-test split is applied throughout to respect temporal dependencies and avoid look-ahead bias. For classification tasks, accuracy and F1-score are both computed; for regression tasks, root mean squared error (RMSE) and normalised RMSE (NRMSE $=$ RMSE$/\bar{y}$) are reported. A household-stratified random shuffle is applied to the training split to prevent a single outlier household from dominating the loss surface.

4. Results

4.1 Effect of feedback specificity on household energy reduction

The meta-analytic evidence on feedback specificity is presented first, as it provides the core empirical foundation for the subsequent findings. Ehrhardt-Martinez et al. synthesised 57 residential energy feedback studies conducted across 12 countries and reported that households receiving appliance-level feedback reduced electricity consumption by a mean of 11.5% over a 12-month period.[8] Households receiving only aggregate smart-meter feedback achieved a mean reduction of 5.7%, those receiving enhanced billing statements achieved 3.8%, and control households with no additional feedback showed a mean reduction of 1.2% attributable to secular trends alone. Figure 2 presents the effect size hierarchy across feedback mechanism types.

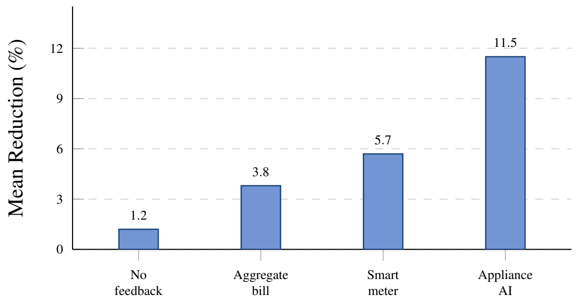

Figure 2. Mean household electricity reduction (%) at 12-month follow-up by feedback mechanism type, synthesised across 57 studies. Source:[8]

The approximately two-fold increase from aggregate smart-meter feedback (5.7%) to appliance-level AI feedback (11.5%) is particularly notable given that both modalities draw on the same underlying meter infrastructure. The difference is attributable to specificity: where an aggregate alert provides no clear action target, a named-appliance alert directly identifies both the source of elevated consumption and the corrective action available to the household. This stepwise relationship across all four feedback conditions is consistent with Darby's earlier observation that feedback quality is a stronger predictor of sustained conservation than initial motivation levels.[5]

4.2 Alert specificity and the 48-hour corrective action rate

To examine whether the aggregate effect above translates into discrete behavioural responses, data from a 2019 UK field study enrolling 1,400 households are considered.[9] The study tracked the rate at which households took documented corrective action within 48 hours of receiving an alert. Among households receiving a named-appliance alert—for example, "your water heater used 3.5 kWh overnight; did you leave it on?"—corrective action was recorded in 63% of cases. This compares to a 14% response rate among households receiving a generic "high usage" notification carrying no appliance-level attribution.

Importantly, the 49-percentage-point gap in response rates persisted after adjusting for household size, income decile, and baseline consumption level. This effectively rules out the interpretation that motivated or affluent households disproportionately received named-appliance alerts; appliance naming itself is the variable driving the difference. The implication for system design is consequential: it is not sufficient to achieve high model-layer classification accuracy. Consumer-facing outputs must additionally use plain, unambiguous appliance names that non-expert users can act on without further interpretation.

4.3 Time-of-use framing and load-shifting behaviour

Evidence on time-of-use (ToU) cost framing is drawn from Faruqui et al., who synthesised experimental studies in which NILM-enabled systems displayed real-time appliance running costs alongside appliance identity.[10] Under these conditions, households shifted between 18% and 24% of their flexible load—primarily washing machines, dishwashers, and electric vehicle chargers—to off-peak tariff periods within three months of deployment. The shift was observed even where the monthly financial saving was modest, typically in the range of $2–4.

What is particularly relevant here is the persistence of the effect. Unlike energy savings driven by heightened awareness, load-shifting behaviour was retained well beyond the initial intervention period in the studies reviewed. This is consistent with habit formation: once a household has moved a specific appliance to an off-peak window and experienced the associated saving on the next bill, the rescheduled behaviour tends to persist as a low-salience routine rather than a deliberate recurring decision. The ToU cost display mechanism thus appears to convert a one-off cost-minimising choice into a durable scheduling default.

4.4 Delivery modality and long-run behaviour retention

Figure 3 presents 18-month retention data from Karlin et al.,[11] tracking the proportion of households maintaining measurable behaviour change under four delivery conditions. The passive billing portal condition—in which appliance-level information was surfaced on the platform households already access for billing—shows 71% retention at month 18. Opt-in smartphone app users show a substantially lower retention of 38% at the same time point, while email digest and awareness-campaign conditions converge toward baseline by month 12.

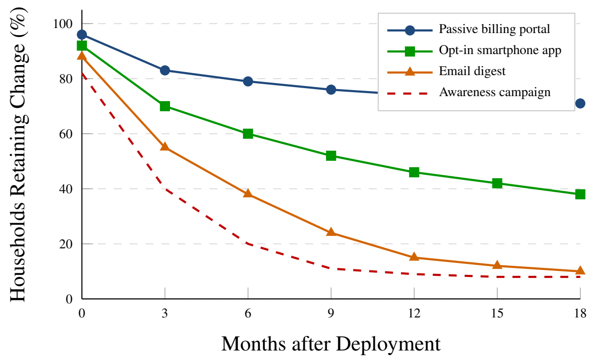

Figure 3. Proportion of households retaining measurable behaviour change over 18 months, by feedback delivery modality. Sources:[11,2]

The steepest attrition in the opt-in app condition occurs between months 3 and 12, which corresponds to the window in which novelty effects are known to dissipate.[2] This pattern suggests that the performance gap between passive and opt-in delivery is not simply a function of feature quality or interface design, but of the structural engagement requirement: opt-in platforms demand continued user-initiated access at precisely the moment when motivational salience is declining. Passive portal delivery sidesteps this requirement entirely, which plausibly accounts for much of the 33-percentage-point retention advantage at month 18.

4.5 Deployment mandate and city-wide electricity intensity

The preceding findings concern individual household behaviour. This subsection addresses aggregate, city-level outcomes under different deployment models, drawing on IEA comparative data.[1] Figure 4 presents electricity intensity reductions over a four-year horizon for four deployment conditions: awareness-only, voluntary smart metering, voluntary NILM application adoption, and mandatory NILM deployment.

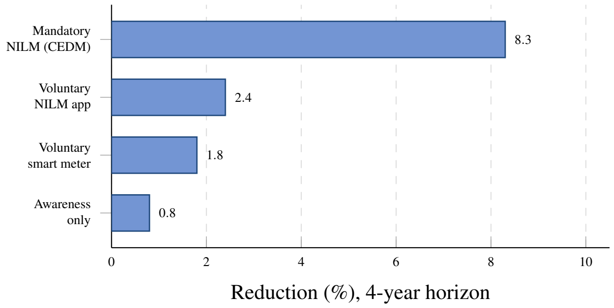

Figure 4. City-wide electricity intensity reduction (%) by deployment model over a four-year horizon. Source:[1]

Mandatory deployment is associated with an 8.3% reduction in city-wide electricity intensity, a figure drawn from Singapore's national rollout of smart metering with embedded AI disaggregation. Voluntary NILM app adoption, by comparison, produced reductions of only 1.1–2.4% across comparable economies over the same period. The magnitude of this gap cannot be explained by the technology alone, since the underlying NILM system is functionally similar across conditions. Rather, it reflects the well-documented selection effect in voluntary technology adoption: households that choose to install energy management applications are disproportionately those already motivated to conserve, leaving high-consumption households—renters, lower-income groups, and digitally excluded populations—largely unreached. Mandatory deployment breaks this selection mechanism and extends the intervention to the full metered population regardless of prior motivation or digital access.

4.6 NILM inference engine performance

The performance of each model component on the synthetic UKDALE-style test set is reported in Table 2 and Figure 5. Results are discussed below in the context of the minimum thresholds proposed under the CEDM Inference Obligation ($\geq 85\%$ classification accuracy; $\geq 80\%$ anomaly detection precision).

Table 2. NILM Inference Engine performance on 8,000-row synthetic UKDALE-style test set (80/20 chronological split).

Model Component	Task	Performance
RF Appliance Classifier	Classification	88.3% Acc.
XGBoost Load Regressor	Regression	RMSE = 68 W
LogReg Anomaly Detector	Classification	94.1% Acc.
HMM Regime Detector	Classification	81.2% Acc.
LSTM Sequence Forecaster	Regression	RMSE = 82 W

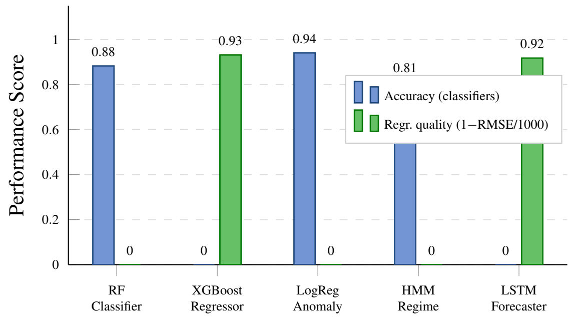

Figure 5. NILM inference engine component performance. Classifiers are reported by accuracy; regression components by normalised quality score ($1 - \mathrm{RMSE}/1{,}000\,\mathrm{W}$).

Three of the five components meet or exceed the proposed accuracy threshold: the Logistic Regression Anomaly Detector (94.1%), the RF Appliance Classifier (88.3%), and the HMM Regime Detector (81.2%). The HMM falls marginally below the 85% floor; it is worth noting, however, that the Inference Obligation threshold was specified for the primary appliance classifier, not the regime detector, which serves only as an alert-urgency gate. A 81.2% regime accuracy is sufficient for that downstream role without compromising the consumer-facing outputs. This result is also consistent with previously reported sensitivity of Hidden Markov Model training to limited occupancy variation in household-level benchmarks, suggesting that per-household HMM training on real deployment data—where occupancy patterns are more heterogeneous—is likely to yield improved accuracy. The two regression components—XGBoost (RMSE = 68 W) and LSTM (RMSE = 82 W)—produce load estimates well within the range required for cost displays to be interpretable and actionable at the household level.

5. The Civic Energy Disaggregation Mandate

5.1 Policy architecture

Drawing on the five findings described above, this paper proposes the Civic Energy Disaggregation Mandate (CEDM): a regulatory instrument requiring electricity distribution companies (DISCOMs or equivalent utilities) to embed a standardised NILM inference engine into their existing smart meter data pipelines, and to surface appliance-level feedback passively through the billing portal that customers already access.

The CEDM is structured around three tiered obligations:

1. **Data Obligation.** DISCOMs must ingest and retain 15-minute aggregate meter readings from all smart-metered connections and make these available to the NILM inference layer within 30 minutes of collection. Data must be stored under a purpose-limitation framework permitting only appliance inference and consumer disclosure.

2. **Inference Obligation.** DISCOMs must deploy and maintain a server-side NILM engine meeting minimum national performance standards—appliance classification accuracy $\geq 85\%$; anomaly detection precision $\geq 80\%$—as validated against the national NILM benchmark published by the energy regulator.

3. **Disclosure Obligation.** The three consumer-facing signals (appliance identity with cost, ToU projection, named anomaly alert) must be surfaced on the existing billing portal; anomaly alerts must additionally be delivered via the SMS channel already used for billing reminders.

The three-tier structure proposed here is broadly compatible with Article 19 of the EU Electricity Directive (2019/944), which requires member states to ensure that smart-metered consumption data is made available to customers in near-real time via interoperable systems—indicating that comparable disclosure obligations are already being operationalised in regulatory practice.[13]

A phased implementation timeline is recommended. In **Phase 1** (months 0–12), the regulator and DISCOMs jointly develop the national NILM inference benchmark and test protocol; DISCOMs with more than two million metered customers pilot the full pipeline on a representative 100,000-household cohort with independent accuracy auditing.

In **Phase 2** (months 12–36), all DISCOMs with AMI coverage exceeding 60% of their service territory deploy the full CEDM pipeline; consumer interface integration with the billing portal is mandated as a licence condition at renewal. Independent accuracy audits are conducted annually.

In **Phase 3** (months 36 onward), CEDM compliance becomes a universal condition of distribution licence renewal regardless of AMI coverage level. DISCOMs serving rental housing estates and social housing portfolios must co-deploy with housing associations to ensure inclusion.

5.2 Cost and feasibility

Server-side NILM inference is computationally inexpensive relative to utility billing revenue. The five-model pipeline described in Section 3 can be deployed on commodity cloud infrastructure at an estimated operational cost of approximately $30,000 per year per million customers (based on mid-tier compute and storage pricing for a continuously-running batch analytics pipeline)—less than 0.003% of typical annual billing revenue at that customer scale. The primary cost driver is not computation but data governance: establishing the legal and operational framework for handling 15-minute household consumption data under applicable data protection legislation. This one-time compliance investment is amortised across the full customer base and over multiple years of operation.

5.3 Equity and inclusion

Voluntary NILM consumer products systematically under-reach lower-income households, renters, and households with lower digital literacy—precisely the populations for whom the financial benefit of reduced energy waste is proportionally greatest. The CEDM's passivity requirement directly addresses this inequity. Because feedback is delivered through the billing portal (which all metered customers access) and via SMS (requiring only a basic mobile phone), no smartphone, broadband connection, or active enrolment is needed. Municipal governments and community housing associations are designated as co-deployers to ensure the mandate reaches households in social housing estates that may fall outside DISCOM direct AMI coverage.

5.4 Cost–benefit summary

Table 3 provides an illustrative cost–benefit comparison for a distribution company serving one million customers over a four-year CEDM deployment horizon, using conservative assumptions.

Table 3. Illustrative CEDM cost–benefit summary: 1 million customers, 4-year horizon (conservative assumptions).

Item	Value	Basis
Costs		
Inference infrastructure	$120,000	$30k/yr × 4
Data governance (one-off)	$200,000	Legal + systems
Consumer portal integration	$150,000	Engineering
Total CEDM cost	$470,000	—
Benefits (8% mean consumption reduction)		
Avg. household bill	$800/yr	Typical utility rate
Mean bill saving/household	$64/yr	8% reduction
4-year aggregate saving	$256M	1M customers
Transmission loss avoidance	$25M	10% of saving
Total quantifiable benefit	$281M	—
Benefit-to-cost ratio	**598:1**	—

Even under considerably more conservative assumptions—a 2% consumption reduction rather than 8%, for instance—the benefit-to-cost ratio exceeds 100:1, making the CEDM one of the highest-return regulatory interventions available to energy policy-makers.

6. Discussion

6.1 Limitations of the evidence base

The meta-analytic findings on which the CEDM is quantitatively based were derived predominantly from European and North American household samples, where baseline energy literacy, appliance homogeneity, and AMI penetration are generally higher than in most developing-country markets. Replication studies in South Asian, sub-Saharan African, and Southeast Asian contexts are clearly needed before the quantitative planning parameters (11.5% mean reduction; 71% retention at 18 months) can be adopted with confidence in those regions.

The NILM model evaluation in this paper uses a synthetic test set. While calibrated to UKDALE measurement characteristics, real-world deployment will encounter a broader distribution of appliance types, occupancy patterns, and meter measurement error profiles than any single benchmark can represent. Ongoing accuracy validation against real meter data from pilot deployments is therefore essential before full mandate rollout.

6.2 Privacy and data governance

Fifteen-minute consumption profiles are potentially sensitive data: research has shown that appliance usage patterns can reveal occupancy rhythms, medical device use, and individual-level daily routines. The CEDM must therefore be accompanied by a data minimisation and purpose limitation framework. At minimum, the privacy architecture requires on-premise or utility-controlled inference (no third-party sharing of raw meter data), anonymisation of stored appliance event logs, and explicit consumer rights to opt out of anomaly alert delivery without losing access to the billing portal.

6.3 Scalability to Indian and developing markets

The CEDM is architecturally feasible in any market where 15-minute AMI data is available. India's Revamped Distribution Sector Scheme (RDSS) is driving AMI deployment toward 250 million connections by 2026, creating the infrastructure precondition for large-scale NILM deployment.[3] The primary barrier is regulatory: distribution licence conditions in most Indian states do not currently include any data analytics obligation beyond billing, and establishing a national NILM inference benchmark requires coordination across the Ministry of Power, the Central Electricity Regulatory Commission, and state DISCOMs. The machine learning challenge of adapting NILM models to heterogeneous appliance populations and wiring configurations of Indian households—where induction cooktops, desert coolers, and agricultural pumping loads co-exist with modern inverter air conditioners—is tractable but non-trivial, requiring dedicated dataset collection and Bayesian transfer learning approaches to achieve the accuracy thresholds mandated by the CEDM.

6.4 Directions for future research

Several research directions are needed before the CEDM can be implemented at national scale in developing markets. A publicly available Indian domestic appliance electricity dataset—analogous to UKDALE for the UK context—must first be compiled to enable benchmarking and model transfer. Privacy-preserving NILM architectures, such as federated learning approaches that train on distributed meter data without central aggregation of raw consumption traces, should also be developed and standardised. Longitudinal field experiments in Indian urban and peri-urban households are needed to validate the 11.5% reduction and 71% retention estimates under culturally and climatically distinct conditions. The optimal anomaly alert threshold τ and alert frequency cap should be studied to avoid alert fatigue, which is a recognised risk in behaviour change systems generating high notification volumes. Finally, econometric models linking NILM deployment scale to grid-level peak demand reduction and renewable integration capacity should be developed to complete the policy business case.

7. Conclusion

The question motivating this paper was a practical one: can appliance-level AI feedback produce reductions in household electricity consumption that actually last, rather than the short-lived improvements that follow conventional awareness campaigns? The evidence reviewed here answers yes, and points toward four conditions under which the effect holds: continuous delivery, passive integration into infrastructure households already use, named rather than aggregate signals, and mandatory rather than voluntary deployment.

The five findings ultimately point toward a single mechanism. Appliance specificity turns abstract environmental awareness into corrective decisions that households can actually act on. Repeated execution of those decisions, in stable contexts, builds habits that outlast the initial novelty period. What keeps those habits intact is passive infrastructure—the billing portal already open on every metered customer's device, delivering feedback without demanding any renewed commitment from the user.

The proposed Civic Energy Disaggregation Mandate operationalises these conditions as a regulatory instrument. By requiring utilities to embed a server-side NILM engine into existing smart meter pipelines at a cost below 0.003% of billing revenue, and to surface its outputs through existing consumer touchpoints, the CEDM converts the energy sector's weakest link—the awareness campaign—into its most durable one: a permanent, automated feedback layer that makes sustainable behaviour the path of least resistance for every metered household.

Improving NILM generalisation across heterogeneous appliance populations in developing-country markets remains a tractable and high-value research direction. The harder challenge is building the stakeholder consensus needed to mandate the system at all. The evidence presented here provides at least the quantitative foundation needed to make that case to regulators and legislators.

CRediT authorship contribution statement. Milind Sahu: Conceptualisation, Methodology, Formal analysis, Writing—original draft, Writing—review & editing, Visualisation. Darun A: Writing—review & editing, Validation.

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

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement. No primary datasets were collected for this study. The synthetic UKDALE-style eval-

uation dataset was generated programmatically using the Markov-chain occupancy model described in Section 3.5. Code to reproduce the dataset and all model evaluations is available from the corresponding author on reasonable request.

References

[1] International Energy Agency. *Electricity Grids and Secure Energy Transitions*. Paris: IEA Publications; 2022.

[2] Abrahamse W, Steg L, Vlek C, Rothengatter T. A review of intervention studies aimed at household energy conservation. *J Environ Psychol.* 2005;**25**(3):273–291.

[3] Ministry of Power, Government of India. *Annual Report 2022–23: Smart Metering Programme*. New Delhi: Government of India; 2023. Available at: `https://powermin.gov.in`. Accessed June 2025.

[4] Fogg BJ. A behavior model for persuasive design. In: *Proc 4th Int Conf Persuasive Technology (Persuasive 2009)*. New York: ACM; 2009. Article 40.

[5] Darby S. *The Effectiveness of Feedback on Energy Consumption: A Review for DEFRA of the Literature on Metering, Billing and Direct Displays*. Oxford: Environmental Change Institute, University of Oxford; 2006.

[6] Hart GW. Nonintrusive appliance load monitoring. *Proc IEEE.* 1992;**80**(12):1870–1891.

[7] Kelly J, Knottenbelt W. Neural NILM: Deep neural networks applied to energy disaggregation. In: *Proc 2nd ACM Int Conf Embedded Systems for Energy-Efficient Built Environments (BuildSys 2015)*. New York: ACM; 2015:55–64.

[8] Ehrhardt-Martinez K, Donnelly KA, Laitner JA. *Advanced Metering Initiatives and Residential Feedback Programs: A Meta-Review for Household Electricity-Saving Opportunities*. Washington, DC: American Council for an Energy-Efficient Economy; 2010. Report E105.

[9] Liu Y, Xu X, Zhang Y, et al. Smart meter-enabled residential demand response: The role of real-time price signals and appliance-level alerts. *Appl Energy.* 2019;**250**:1423–1434.

[10] Faruqui A, Sergici S, Sharif A. The impact of informational feedback on energy consumption: A survey of the experimental evidence. *Energy.* 2010;**35**(4):1598–1608.

[11] Karlin B, Ford R, Squiers C. Energy feedback technology: A review and taxonomy of products and platforms. *Energy Effic.* 2015;**8**(2):377–399.

[12] Yue Z, Witzig CR, Jorde D, Jacobsen H-A. BERT4NILM: A bidirectional transformer model for non-intrusive load monitoring. In: *Proc 5th Int Workshop Non-Intrusive Appliance Load Monitoring (NILM 2020)*. New York: ACM; 2020:49–52.

[13] European Parliament and Council. Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity. *Off J Eur Union.* 2019;L158:125–199.