ENERGY DATA ANALYTICS

AI-Based Analytics for Smart-Meter Electricity Data: A Review of Forecasting, Anomaly Detection, Explainability and Privacy, and the Research Gap in Session-Level Decision Intelligence

Sankitaka Lokhe, Rashmi Kulkarni

Department of Information Technology, Siddhant College of Engineering, Pune, India

Published November 2026

Abstract

Smart meters generate fine-grained electricity data that supports forecasting, anomaly detection, appliance-level analysis and consumer feedback. This paper reviews 21 sources published since 2020, including IEEE Access papers and arXiv preprints, and organises them into a taxonomy of forecasting, anomaly and theft detection, explainability, and privacy with non-intrusive load monitoring. A capability comparison shows that forecasting dominates the reviewed set, that explainability is thinly covered, and that none of the reviewed sources groups readings into consumption sessions or combines forecasting, anomaly detection and explanation in one consumer-facing application. Nine research gaps are identified and each is tied to evidence. Session-level decision intelligence is proposed as a direction, together with research questions and an evaluation plan against dashboard and ML-only baselines. The review is narrative, and its limits are stated.

Keywords: anomaly detection; decision intelligence; explainable AI; literature review; load forecasting; session detection; smart meter

1 Introduction

Smart meters record electricity consumption at fine time resolution and send it over a communication network, which makes household and building load a rich source for analytics. A recent survey places the applications of this data in the setting of a secure, reliable and robust grid and sets out future directions [1]. In practice the data is used for forecasting, anomaly and theft detection, appliance-level analysis and consumer feedback.

Research on each of these tasks has grown quickly since 2020, and several recent reviews cover single tasks: load forecasting with machine learning and deep learning [2], [3], anomaly and theft detection [4], [5], explainable AI for energy forecasting [6] and non-intrusive load monitoring [7], [8]. What is missing is a view across tasks that asks a consumer-oriented question: does any line of work turn meter readings into explained, personalised decisions? This paper reviews recent literature with that question in mind.

The contributions are: (i) a taxonomy of recent AI work on smart-meter data; (ii) a comparison of what the reviewed sources do and do not cover; (iii) a list of research gaps, each tied to evidence in the reviewed sources; and (iv) a research direction, session-level decision intelligence, that addresses the gaps.

2 Review method and scope

This is a narrative review and not a formal systematic review. Sources were located through web searches for smart-meter forecasting, anomaly detection, explainability, privacy, non-intrusive load monitoring and datasets. Only sources published in 2020 or later were kept, and a source was kept if it dealt with smart-meter or building electricity data and its abstract or text could be read. In total 21 sources are reviewed, including journal articles, IEEE Access papers and arXiv preprints.

Most of the sources were read at the level of abstract and available text, and the descriptions below are limited to what those texts state. Statements of the form "none of the reviewed sources" therefore refer to this set of 21 sources and do not claim anything about the whole literature. A formal protocol with stated databases, search strings and screening counts would be the next step.

3 Background and taxonomy

Smart-meter data is a set of time series, one per meter, usually at 15, 30 or 60 minute intervals. Figure 1 groups the reviewed work into four task families that sit on top of this data: forecasting, anomaly and theft detection, explainability, and privacy together with appliance-level analysis (NILM). A fifth box marks the gap discussed later.

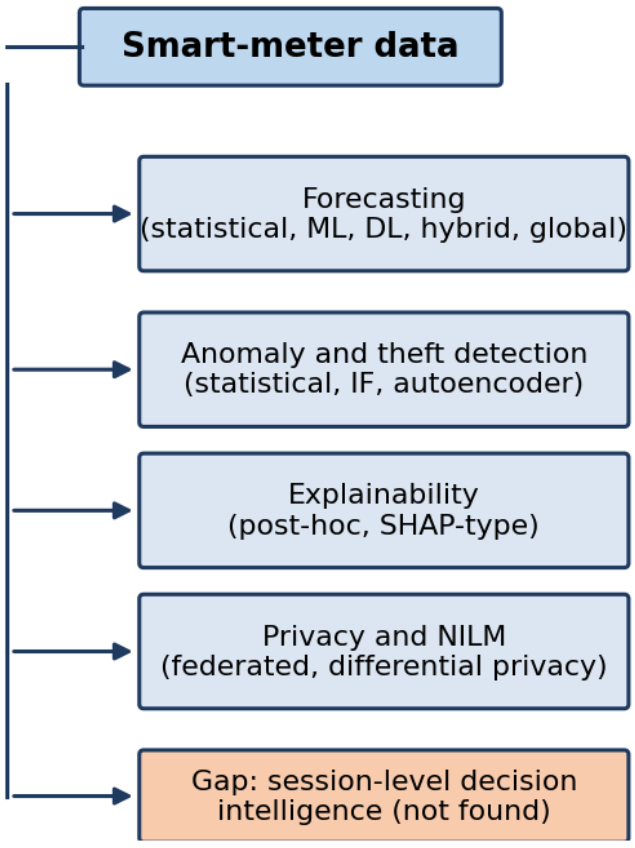


Figure 1. Taxonomy of AI work on smart-meter data and the gap identified in this review.

4 Consumption forecasting

4.1 Principles and reviews

Petropoulos et al. review forecasting theory and practice, including how forecasts are prepared, produced and evaluated [9]. Mustafa et al. survey machine learning and deep learning for electricity load and price prediction, and also compile the datasets and the performance measures used [2]. Biswal et al. review deep learning and machine learning for smart-grid load forecasting, compare them with traditional methods on measures such as MAPE and RMSE, and discuss probabilistic and ensemble forecasting for demand uncertainty [3]. Together these show that the field is mature at the level of method, and that attention has moved to data, evaluation and uncertainty.

4.2 Residential and meter-level models

At household level, Muneer et al. compared an LSTM with exponential smoothing and ARIMA on 12 houses over three months and found the LSTM best [10]. Semmelmann et al. proposed a bidirectional LSTM-XGBoost hybrid for day-ahead energy-community load that forecasts the general pattern and the peaks separately, and that improved when smart-meter data was added as input [11]. Zaboli et al. combined sequence-to-sequence learning, LSTM and stacked autoencoders for residential load profiles that include photovoltaic, battery and electric-vehicle loads [12]. Novaes et al. proposed a Transformer-based meter-level forecaster and compared it with LSTM and a plain recurrent network [13]. Syed et al. combined deep learning with clustering and consumption-pattern recognition [14].

4.3 Global models

Grabner et al. propose one deep-learning model that forecasts many loads in a distribution network, in place of one model per meter [15]. This matters for deployment, because a platform serving many households cannot tune a separate model for each.

4.4 Observations

Three observations follow. First, LSTM-type and hybrid tree-neural models are the most common choices at meter level. Second, the benchmarks differ from study to study (ARIMA and exponential smoothing, standard load profiles, LSTM, plain recurrent networks), so results are not directly comparable. Third, outputs are numbers or curves, and the studies reviewed here do not describe how a forecast is explained to the person who pays the bill.

5 Anomaly and theft detection

Himeur et al. review AI-based anomaly detection of energy consumption in buildings [4]. Nayak and Jaidhar systematically review electricity-theft and anomalous-consumption detection using machine learning, deep learning and hybrid models, and also cover privacy-preserving methods, datasets and challenges [5]. Both reviews treat anomaly detection as a mature topic with open issues in data, labels and deployment.

Individual studies illustrate the main techniques. Lim et al. used Isolation Forest to label normal and abnormal readings and then Gaussian Naive Bayes to classify new ones, reporting balanced accuracy above 89% on simulated 30-minute data, and noted that labelled ground truth is scarce [16]. Labura et al. combined Isolation Forest with Fast Fourier Transform filtering to handle point and contextual anomalies in smart-meter data and argued that data quality must be treated explicitly [17].

Czyżewski applied autoencoders, including LSTM and one-dimensional convolutional variants, to power consumption of lighting systems, and reported the autoencoder approach as best in that study [18].

Two points matter for consumer analytics. Detection is usually reported as a flag or a score, without the expected range and the reason that a household would need to judge the alert. And much of the evidence rests on simulated or injected anomalies [16], so the reported accuracy shows ability on that kind of anomaly and not on all real faults.

6 Explainability and decision support

Arabzadeh and Frank reviewed 50 studies from 2020 to 2025 that apply explainable AI to energy forecasting. Post-hoc methods, especially SHAP, dominate (62% of studies), whereas testing of explanation robustness and reporting of computational overhead remain gaps, and load forecasting shows a user-centric approach [6]. Among the 21 sources here this is the only one whose primary focus is explainability (Figure 2), which is itself a finding.

Explanation for a consumer differs from feature attribution for a modeller. A household needs a short statement: what was expected, what happened, how large the difference is and why the system calls it unusual. Whether such statements help consumers decide is not reported in the reviewed texts and is an open empirical question.

7 Privacy and appliance-level analysis

Delgado Fernandez et al. combined federated learning with differential privacy and secure aggregation for residential short-term load forecasting, and their simulations indicate that high accuracy and near-complete privacy can be achieved together [19]. NILM goes in the opposite direction, inferring appliance behaviour from the aggregate signal. Kaselimi et al. review residential NILM and argue for practical, trustworthy frameworks [7], and Irani Azad et al. review deep neural network methods, public databases and standard metrics [8]. The pairing is a reminder that fine-grained consumption data is sensitive, so any consumer analytics platform must handle it with care.

8 Datasets and evaluation practice

The UCI ElectricityLoadDiagrams20112014 dataset holds 15-minute readings for 370 clients from 2011 to 2014 [20]. Its documentation notes zero values for clients that started late and clock-change effects, such as an hour of zeros in March and a double hour in October [20]. Emami et al. include it in a benchmark, but treat the series as commercial buildings because the loads are much larger than those of homes [21]. Other studies use small residential samples, such as 12 houses over three months [10]. Dataset choice thus limits how far results transfer to households.

On evaluation, forecasting practice stresses simple baselines and held-out data [9], and the reviews use MAE, RMSE and MAPE [2], [3]. For smart-meter series with zeros, MAPE is undefined on zero days, so MAE and RMSE are safer primary measures. A chronological split, with no shuffling, is needed to prevent leakage of future readings.

9 Comparative analysis

Table 1 marks the primary focus of each reviewed source that has a clear task focus. Figure 2 counts the sources by theme. The mark is based on the abstract or text read, and a blank means the focus was not stated there.

Table 1. Capability focus of the reviewed sources.

Ref.	Type	F	A	X	P	N	S
[9]	Review	●					
[2]	Review	●					
[3]	Review	●					
[10]	Study	●					
[11]	Study	●					
[12]	Study	●					
[13]	Study	●					
[14]	Study	●					
[15]	Study	●					
[19]	Study	●			●		
[4]	Review		●				
[5]	Review		●		●		
[16]	Study		●				
[17]	Study		●				
[18]	Study		●				
[6]	Review	●		●			
[7]	Review					●	
[8]	Review					●	

F: forecasting; A: anomaly or theft detection; X: explainability; P: privacy; N: NILM; S: session-level grouping of readings. The overview and dataset sources [1], [20], [21] are not marked.

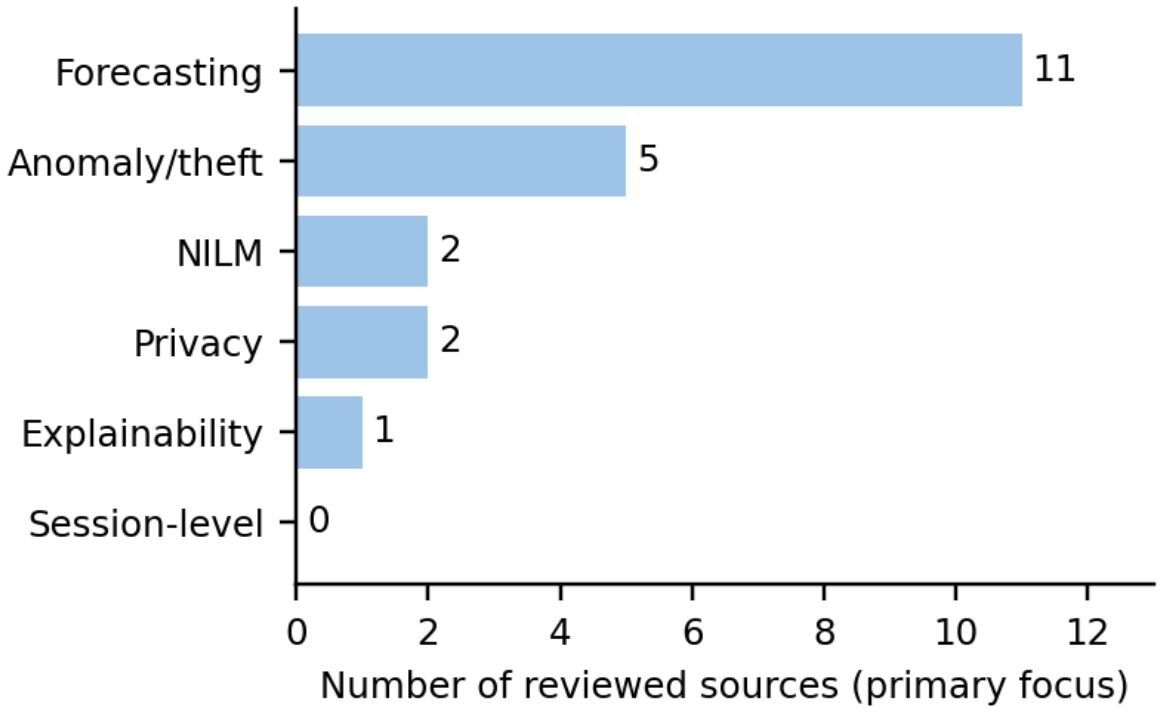


Figure 2. Reviewed sources by primary theme (a source can count in more than one theme).

Forecasting dominates, anomaly detection follows, and explainability, privacy and NILM have few sources. No reviewed source describes session-level grouping, and none combines forecasting, anomaly detection, explanation and a recommendation in one consumer-facing application.

10 Research gaps

Each gap below is tied to the evidence reviewed above. The gaps and the supporting evidence are listed in Table 2.

Table 2. Research gaps and evidence.

Gap	Evidence in reviewed sources
RG1 Isolated capabilities	Table I: each source has one or two foci; none covers all of F, A and X
RG2 No session-level unit	Table I, column S: no reviewed source groups readings into consumption sessions
RG3 Thin explainability	One XAI-focused source of 21; robustness and overhead rarely tested [6]
RG4 Scarce anomaly labels	Simulated data and scarce labels [16]; evaluation on injected anomalies
RG5 Data representativeness	UCI data treated as commercial [21]; small residential samples [10]
RG6 Inconsistent baselines	Benchmarks differ between studies; baselines stressed in practice [9]
RG7 Data quality vs anomaly	Quality faults must be separated from real anomalies [17]; UCI artefacts [20]
RG8 Privacy apart from UX	Privacy studied mostly separately [19], [5]; not tied to consumer insights
RG9 No user-facing evidence	Reviewed texts do not report whether explained advice helps consumers

RG2 and RG3 are the least explored in this set, and they are linked: a session is a unit that a household can recognise, such as the evening peak, and an explanation is easiest to give about such a unit.

11 Proposed direction: session-level decision intelligence

Session-level decision intelligence is proposed to address RG1 to RG3. Interval readings are first validated and cleaned, then grouped into sessions using baseline estimation, a dynamic threshold, activation detection and gap merging. Each session is described by its start and end, duration, energy, average and peak power, peak time and deviation from comparable past sessions. Forecasting and anomaly detection then run on sessions and daily series, and an explanation layer states the expected range, the actual value, the percentage deviation and the reason (Figure 3).

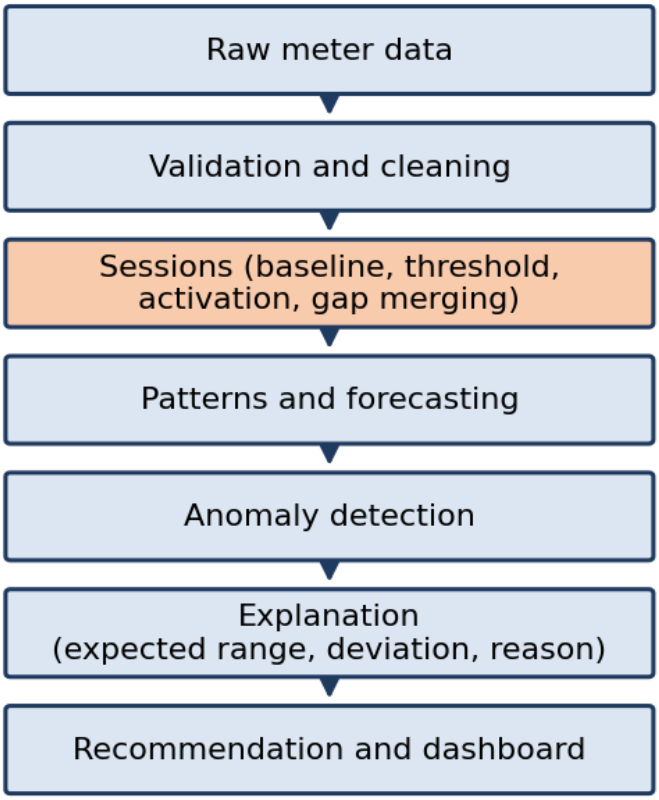


Figure 3. Proposed session-level decision-intelligence pipeline; the session stage (shaded) is the new element.

The framework is being developed as the WattWise web application using Python, FastAPI, PostgreSQL and React, with the UCI dataset for development. The proposed evaluation compares (A) a raw dashboard, (B) a dashboard with machine-learning prediction and (C) a dashboard with prediction and session-based decision intelligence. Forecasting models (ARIMA, Random Forest, XGBoost, optional LSTM) are compared with naive baselines on a 70/15/15 chronological split. Anomaly detectors (statistical, Isolation Forest, autoencoder) are compared on labelled sessions. No results are reported in this review.

Four research questions follow. RQ1: do session features improve forecasts over daily lags alone? RQ2: do session-level detectors flag unusual episodes better than reading-level detectors? RQ3: can explanations of the expected-range type be generated reliably and cheaply? RQ4: does configuration C improve decision usefulness over A and B?

12 Challenges and future work

- **Household data:** validate on residential datasets or a controlled collection, because the main development dataset is commercial-scale [21].
- **Labels:** build expert-labelled or realistically injected anomaly sets and state their limits [16].
- **Standard protocol:** fix baselines, chronological splits and zero-aware metrics so that studies can be compared [9].
- **Explanation quality:** test explanation robustness and cost [6] and study how consumers read explanations.
- **Privacy:** apply anonymisation and aggregation first, and consider federated approaches later [19].
- **Live data:** connect a real meter or API and handle communication gaps and stagnant readings.

13 Conclusion

Recent work on smart-meter data is strong on forecasting and detection but treats capabilities one at a time. In the 21 sources reviewed, none groups readings into sessions and none joins forecasting, anomaly detection and explanation in a consumer-facing application. Nine research gaps were identified and tied to evidence. Session-level decision intelligence is proposed as a direction that addresses the main gaps, and it will be evaluated against dashboard and ML-only baselines. The review itself is narrative, and a formal systematic protocol is the next step.

References

[1] Mitra S, Chakraborty B, Mitra P. Smart meter data analytics applications for secure, reliable and robust grid system: survey and future directions. *Energy*. 2024;289:129920. https://doi.org/10.1016/j.energy.2023.129920

[2] Mustafa K, et al. Toward holistic energy management by electricity load and price forecasting: a comprehensive survey. *IEEE Access*. 2023. https://doi.org/10.1109/ACCESS.2023.3334164

[3] Biswal B, et al. *Review on smart grid load forecasting for smart energy management using machine learning and deep learning techniques*. Energy Reports. 2024:3654–3670. Available from: https://eprints.whiterose.ac.uk/222121

[4] Himeur Y, et al. Artificial intelligence based anomaly detection of energy consumption in buildings: a review, current trends and new perspectives. *Applied Energy*. 2021;287:116601. https://doi.org/10.1016/j.apenergy.2021.116601

[5] Nayak R, Jaidhar CD. *Data-driven models for electricity theft and anomalous power consumption detection: a systematic review*. Applied Intelligence. 2025;55(11). Available from: https://idr.nitk.ac.in/handle/123456789/20204

[6] Arabzadeh V, Frank R. A four-dimensional analysis of explainable AI in energy forecasting: a domain-specific systematic review. *Machine Learning and Knowledge Extraction*. 2025;7(4):153. Available from: https://orbilu.uni.lu/handle/10993/66497

[7] Kaselimi M, Protopapadakis E, Voulodimos A, Doulamis N, Doulamis A. Towards trustworthy energy disaggregation: a review of challenges, methods, and perspectives for non-intrusive load monitoring. *Sensors*. 2022;22:5872. Available from: https://arxiv.org/abs/2207.02009

[8] Irani Azad M, Rajabi R, Estebsari A. Non-intrusive load monitoring (NILM) using deep neural networks: a review. *arXiv* [Preprint]. 2023. arXiv:2306.05017.

[9] Petropoulos F, et al. Forecasting: theory and practice. *International Journal of Forecasting*. 2022;38(3):705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001

[10] Muneer A, et al. Short term residential load forecasting using long short-term memory recurrent neural network. *International Journal of Electrical and Computer Engineering*. 2022;12(5):5589–5599. Available from: https://ijece.iaescore.com/index.php/IJECE/article/view/27800

[11] Semmelmann L, Henni S, Weinhardt C. Load forecasting for energy communities: a novel LSTM-XGBoost hybrid model based on smart meter data. *Energy Informatics*. 2022;5(Suppl 1). https://doi.org/10.1186/s42162-022-00212-9

[12] Zaboli A, Tuyet-Doan VN, Kim YH, Hong J, Su W. An LSTM-SAE-based behind-the-meter load forecasting method. *IEEE Access*. 2023;11:49378–49392. https://doi.org/10.1109/ACCESS.2023.3276646

[13] Novaes ALF, de Araujo RAM, Figueiredo J, Pavanelli LA. A new state-of-the-art transformers-based load forecaster on the smart grid domain. *arXiv* [Preprint]. 2021. arXiv:2108.02628.

[14] Syed D, et al. Deep learning-based short-term load forecasting approach in smart grid with clustering and consumption pattern recognition. *IEEE Access*. 2021;9:54992–55008.

[15] Grabner M, Wang Y, Wen Q, Blazic B, Struc V. A global modeling approach for load forecasting in distribution networks. *arXiv* [Preprint]. 2022. arXiv:2204.00493.

[16] Lim JY, Tan WN, Tan YF. Anomalous energy consumption detection using a Naive Bayes approach. *F1000Research*. 2022;11:64. Available from: https://f1000research.com/articles/11-64

[17] Labura P, Antic T, Capuder T. Time and frequency domain-based anomaly detection in smart meter data for distribution network studies. *arXiv* [Preprint]. 2025. arXiv:2504.18231.

[18] Czyzewski A. Autoencoder application for anomaly detection in power consumption of lighting systems. *IEEE Access*. 2023;11. https://doi.org/10.1109/ACCESS.2023.3330135

[19] Delgado Fernandez J, et al. Privacy-preserving federated learning for residential short-term load forecasting. *Applied Energy*. 2022;326:119915. Available from: https://orbilu.uni.lu/handle/10993/52125

[20] Trindade A. ElectricityLoadDiagrams20112014 [dataset]. UCI Machine Learning Repository; 2015. Available from: https://archive.ics.uci.edu/dataset/321/electricityloaddiagrams20112014

[21] Emami P, et al. BuildingsBench: a large-scale dataset of 900K buildings and benchmark for short-term load forecasting. *arXiv* [Preprint]. 2023. arXiv:2307.00142.