

Hybrid Anomaly Detection Framework for Electricity Theft Identification in Smart Grid Systems

Nishigandha Vyawahare

Dept. of CSE (Data Science)

G H Raison College of Engineering and Management
Affiliated to Savitribai Phule Pune University
Pune, India
nishi.sakharkar@gmail.com

Chetan Bade

Dept. of CSE (Data Science)

G H Raison College of Engineering and Management
Affiliated to Savitribai Phule Pune University
Pune, India
chetanbade390@gmail.com

Masira Kulkarni

Dept. of CSE (Data Science)

G H Raison College of Engineering and Management
Affiliated to Savitribai Phule Pune University
Pune, India
masira.kulkarni@raisoni.net

Niranjan Barhate

Dept. of CSE (Data Science)

G H Raison College of Engineering and Management
Affiliated to Savitribai Phule Pune University
Pune, India
niranjanbarhate058@gmail.com

Siddhesh Gole

Dept. of CSE (Data Science)

G H Raison College of Engineering and Management
Affiliated to Savitribai Phule Pune University
Pune, India
siddeshgole01@gmail.com

Abstract—Electricity theft and abnormal power consumption pose critical challenges to modern power distribution networks, causing significant financial losses, operational inefficiencies, and degraded grid reliability. Conventional detection methods, which depend on manual inspections and rigid rule-based thresholds, cannot scale to handle the continuous data streams generated by smart meter infrastructure. This paper presents a Hybrid AI-Based Electricity Theft and Anomaly Detection System evaluated on a dataset of 5,000 smart meter readings. The proposed architecture integrates three complementary models: XGBoost for supervised theft pattern classification, Isolation Forest for unsupervised anomaly isolation, and a one-dimensional Convolutional Neural Network (CNN-1D) to capture temporal dependencies in consumption sequences. A Logistic Regression meta-learner aggregates the probability outputs of all base models to produce the final detection decision. The XGBoost component achieves a perfect recall of 1.00 and an F1-score of 0.89 with 99% accuracy, while the CNN-1D model attains an AUC-ROC of 0.90. To ensure operational transparency, the framework incorporates a SHAP-based Explainable AI module and is deployed through an interactive Streamlit dashboard, delivering a practical decision-support tool for utility administrators.

Index Terms—electricity theft detection, anomaly detection, smart grid security, hybrid ensemble, XGBoost, CNN-1D, Isolation Forest, time-series analysis, explainable AI, Streamlit deployment

I. INTRODUCTION

Electricity theft and irregular power consumption represent persistent, large-scale threats to modern power distribution systems. The resulting financial drain on utility companies runs into billions of dollars annually, while secondary effects such as distorted load forecasting, localised outages, and eroded grid stability impose further costs on consumers and the broader energy infrastructure [1], [2]. In smart grid environments, consumers exploit various methods to misappropriate energy, including physical meter tampering, direct illegal connections, exploitation of damaged metering hardware, and cyber-based manipulation of billing records [3].

A. Limitations of Conventional Monitoring

Traditional utility monitoring relies on periodic physical inspections, manual meter readings, and simple threshold-based alerts. Although threshold mechanisms offer fast, interpretable screening for extreme events such as sudden load drops, they are inflexible by design. Electricity demand is inherently dynamic—shaped by seasonal variation, temperature, and occupancy patterns—so static rules generate excessive false positives while entirely missing gradual or sophisticated tampering [4]. Physical inspections scale poorly; by the time a field team identifies a tampered meter, substantial losses may have already occurred.

B. Machine Learning as a Solution

The widespread rollout of Advanced Metering Infrastructure (AMI) now provides continuous, high-frequency consumption records that are well suited to data-driven analysis. Supervised classifiers such as Support Vector Machines [5], Random Forest, and gradient-boosted trees can map engineered temporal features to known theft signatures with high accuracy. Yan and Wen demonstrated that XGBoost applied to AMI metering data outperforms conventional classifiers across multiple attack types while remaining robust under class imbalance [6]. For time-series data, deep neural architectures—particularly Long Short-Term Memory networks and one-dimensional CNNs—automatically extract non-linear temporal patterns that shallow models miss [7]. Lepolesa *et al.* further confirmed that deep neural networks can achieve strong detection performance across both time-domain and frequency-domain feature representations [8]. A major operational constraint, however, is the extreme class imbalance inherent to real-world theft data. The 5,000-record dataset used in this study reflects a 5% anomaly rate (50 anomalous out of 996 test samples), motivating unsupervised techniques such as Isolation Forest [9], which identifies anomalies without requiring labelled examples. Ensemble approaches that combine the strengths of multiple classifiers have further demonstrated their effectiveness in tackling this class-imbalance challenge [10].

C. Proposed Contribution

To overcome these constraints, this paper introduces a hybrid detection architecture integrating supervised learning, unsupervised anomaly detection, and temporal deep learning within a stacking ensemble. The key contributions of this work are:

- A three-model stacking ensemble combining XGBoost, Isolation Forest, and CNN-1D under a Logistic Regression meta-learner.
- Engineered temporal features—*Consumption_Deviation* and *Consumption_Ratio*—that quantify deviations from historical baselines.
- An integrated SHAP-based explainability module [11] that produces human-readable justifications for each detected anomaly.
- A fully deployed interactive Streamlit dashboard for real-time smart grid monitoring and investigation support.

II. LITERATURE SURVEY

Early computational work on non-technical loss (NTL) detection focused on shallow classifiers trained over historical consumption records. Nagi *et al.* demonstrated that Support Vector Machines could identify abnormal usage patterns by learning from labelled smart meter histories, establishing a foundation for supervised approaches [5]. Viegas *et al.* surveyed a broad range of supervised and rule-based detection solutions, identifying scalability and label scarcity as the two most consequential unresolved challenges [12].

To reduce dependence on labelled theft data, researchers turned to unsupervised methods. Liu *et al.* introduced the

Isolation Forest algorithm, which isolates anomalous instances by recursively partitioning the feature space; observations requiring fewer splits to isolate are assigned higher anomaly scores [9].

Advances in deep learning expanded temporal modelling capacity. Zheng *et al.* applied wide and deep CNNs to electricity theft detection, showing that convolutional filters can capture subtle periodic deviations in consumption profiles [7]. Lepolesa *et al.* proposed a deep neural network approach using both time-domain and frequency-domain features, demonstrating that appropriate feature engineering substantially reduces false positives [8]. Coma-Puig *et al.* demonstrated that careful feature design and model selection improve supervised fraud detection in energy systems [13]. Despite strong predictive performance, these architectures share a fundamental weakness: internal reasoning is opaque, limiting trust and utility for field investigators.

Jokar *et al.* showed that customer consumption pattern features carry strong discriminative signal for AMI-based theft detection [14]. Zanetti *et al.* demonstrated that tunable short-lived pattern detection systems further reduce false alarms [15]. Recent hybrid frameworks combining rule-based screening, unsupervised isolation, and deep learning achieve improved accuracy and lower false-positive rates compared to single-paradigm approaches [4].

The emergence of stacking ensemble strategies has provided a principled means of combining heterogeneous base learners. Khan *et al.* developed a stacked machine and deep learning model for AMI theft detection that combines uncorrelated classifiers under a meta-learner, achieving improved generalisation relative to single-model baselines [16]. Pamir *et al.* further demonstrated that boosting classifiers combined in a stacking configuration deliver strong recall on the publicly available SGCC dataset [17]. Yan and Wen confirmed that XGBoost outperforms eight competing methods including SVM and decision trees in both detection accuracy and false-positive rate [6]. Gunturi and Sarkar showed that ensemble machine learning models consistently outperform standalone classifiers for energy theft detection across multiple evaluation metrics [10].

To address the interpretability gap that limits operator adoption of black-box models, Lundberg and Lee introduced the SHAP (SHapley Additive exPlanations) framework, which assigns each input feature a quantified importance value for individual predictions [11]. Alsaigh *et al.* conducted a systematic review of explainable AI governance frameworks in smart energy systems, highlighting SHAP as a primary tool for bridging the gap between model complexity and operational transparency [18].

Recent work has also explored adaptive detection paradigms. El-Toukhy *et al.* introduced a deep reinforcement learning-based theft detector capable of adapting to the dynamic nature of consumption patterns and novel cyber-attack vectors in smart grids [19]. Jithish *et al.* proposed a federated learning-based anomaly detection architecture for smart grids that enables privacy-preserving distributed training

TABLE I
SUMMARY OF RELATED WORK

Reference	Methodology	Research Gap
Nagi <i>et al.</i> [5]	SVM on historical metered records	Requires labelled data; poor recall on emerging patterns
Zheng <i>et al.</i> [7]	Wide & deep CNN for NTL	Black-box; limited explainability for operators
Liu <i>et al.</i> [9]	Isolation Forest (unsupervised)	No temporal modelling; standalone accuracy limited
Zanetti <i>et al.</i> [15]	Tunable short-lived pattern detection	Requires domain tuning; not hybrid
Viegas <i>et al.</i> [12]	Survey of NTL solutions	Highlights need for scalable hybrid systems
Yan & Wen [6]	XGBoost on AMI metering data	Supervised only; no temporal deep learning component
Khan <i>et al.</i> [16]	Stacked ML and DL ensemble	No explainability module; limited unsupervised coverage
Lepolesa <i>et al.</i> [8]	Deep NN with time/freq. features	No stacking ensemble; interpretability not addressed
El-Toukhy <i>et al.</i> [19]	Deep reinforcement learning for ETD	High training complexity; no XAI integration
Jithish <i>et al.</i> [20]	Federated learning anomaly detection	Distributed setup; not suited for single-utility deployment

across utility networks, pointing towards future directions for scalable deployment [20].

Despite this progress, existing systems rely either on purely supervised pipelines that struggle under label scarcity, or on single deep learning models that sacrifice interpretability. The present work addresses both gaps through a hybrid stacking ensemble under a unified explainable framework.

III. PROPOSED HYBRID DETECTION FRAMEWORK

Fig. 1 illustrates the complete system architecture. Raw smart meter data flows through preprocessing and feature engineering before being processed in parallel by three base models. A Logistic Regression meta-learner combines their outputs, and the final decision is delivered through a SHAP explainability module [11] and a Streamlit dashboard.

A. Data Acquisition and Preprocessing

The framework ingests electricity consumption readings recorded at 30-minute intervals. The raw dataset includes the primary energy consumption variable alongside contextual environmental attributes—temperature, humidity, and wind speed—that capture legitimate demand variation. During preprocessing, missing sensor values are imputed, noisy records are removed, and all numerical features are normalised to the unit interval to prevent magnitude-induced bias.

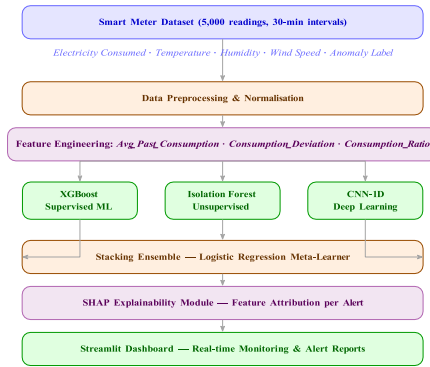


Fig. 1. Proposed Hybrid AI Electricity Theft Detection Architecture. Green boxes denote models selected for the stacking ensemble.

B. Feature Engineering

Three features anchor anomaly detection. *Avg_Past_Consumption* encodes each consumer's rolling historical baseline. *Consumption_Deviation* measures the signed difference between a current reading and this baseline, capturing both upward and downward departures. *Consumption_Ratio* expresses this difference proportionally, ensuring consistent response across consumers with widely varying absolute demand. Together, these features enable identification of sudden consumption spikes indicative of illegal tapping and gradual multi-day drifts associated with meter tampering. This design aligns with prior work showing that customer consumption pattern features carry strong discriminative signal for AMI-based theft detection [14].

C. Base-Layer Detection Models

1) *XGBoost (Supervised Classifier)*: XGBoost is selected as the primary supervised component owing to its strong performance on structured tabular data, native class-weight adjustment for imbalanced conditions, and consistently high recall on labelled theft signatures. Yan and Wen established XGBoost as a state-of-the-art AMI theft detector, demonstrating its robustness across six attack types and data imbalance scenarios [6]. In the present study, XGBoost achieved 99% accuracy and perfect recall of 1.00 on the test partition.

2) *Isolation Forest (Unsupervised Anomaly Detection)*: Isolation Forest operates without class labels. It constructs an ensemble of random decision trees and assigns higher anomaly scores to observations isolated in fewer partitioning steps—data points that differ significantly from dense, typical consumption regions [9]. It achieved the highest individual AUC-ROC of 0.914 across all evaluated models, validating its capability to surface novel theft strategies absent from the labelled set.

3) *CNN-1D (Temporal Deep Learning)*: A one-dimensional CNN processes fixed-length windows of sequential consumption readings. Stacked convolutional layers learn multi-scale temporal filters capturing periodic patterns, gradual drifts, and irregular usage spikes [7]. CNN-1D achieved an AUC-ROC of 0.90, outperforming both bidirectional LSTM (0.75) and

TABLE II
DATASET PARTITION SUMMARY

Partition	Records	Normal	Anomaly
Training	4,004	3,806	198
Testing	996	946	50
Total	5,000	4,752	248

Transformer (0.61) architectures while remaining computationally lighter on the 5,000-record dataset.

D. Stacking Ensemble Architecture

The three base-model probability scores are concatenated as a three-feature input vector for a Logistic Regression meta-learner, trained on out-of-fold validation predictions. The meta-learner learns optimal combination weights, leveraging the supervised precision of XGBoost, the label-free coverage of Isolation Forest, and the temporal sensitivity of CNN-1D simultaneously to reduce both false positives and false negatives. This stacking design is consistent with recent work demonstrating that combining heterogeneous base learners under a meta-classifier improves generalisation relative to any single component [10], [16], [17].

E. Explainability Module

SHAP values [11] are computed for each flagged instance, quantifying the contribution of every input feature to the final prediction. The module reports top contributing features alongside each alert in plain language, enabling utility engineers to understand the operational rationale before initiating field investigation. This approach directly addresses the interpretability gap identified in surveys of explainable AI for energy systems [18].

F. Streamlit Dashboard Deployment

The complete pipeline is deployed through an interactive Streamlit web application providing: real-time consumption trend visualisation, live anomaly alert feeds with SHAP justifications, batch analysis for historical record review, and downloadable investigation reports.

IV. METHODOLOGY

A. Dataset Description and Preparation

Experiments are conducted on 5,000 consecutive smart meter records sampled at 30-minute intervals, containing 11 features with a 5% anomaly rate. The dataset was partitioned chronologically (80% training / 20% testing) so models are evaluated strictly on future-timestep records, closely mirroring real deployment conditions. The test set comprised 996 samples: 946 normal and 50 anomalous readings.

Fig. 2 presents the consumption trend over the first 500 readings of the test sequence alongside the corresponding hybrid model anomaly score. The upper panel overlays true anomaly labels (red markers) on the raw electricity consumption trace (cyan line), confirming that labelled theft events

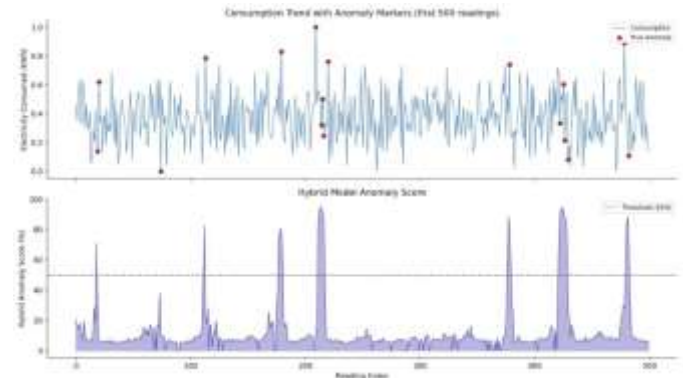


Fig. 2. Dual-panel visualisation over the first 500 test readings. *Top*: Raw electricity consumption (kWh, cyan) with true anomaly positions marked in red. *Bottom*: Hybrid model anomaly score (%) with the 50% decision threshold (dashed red line). Score peaks correspond closely to labelled theft events, validating the ensemble's discriminative behaviour across the full dynamic range of consumption values.

span a wide range of consumption magnitudes—both sharp spikes and near-zero dips—and cannot be isolated by a simple amplitude threshold. The lower panel displays the aggregated hybrid anomaly score on a 0–100% scale with a decision threshold at 50%. Score peaks closely align with the labelled anomaly positions in the upper panel, demonstrating that the stacking ensemble assigns elevated suspicion at the correct reading indices even when individual consumption values appear unremarkable to a visual inspector.

B. Model Training

XGBoost was trained with `scale_pos_weight` adjustment for class imbalance. The Isolation Forest contamination parameter was set via cross-validated optimisation on a held-out validation fold. CNN-1D was trained on sliding windows of length 24 (12 hours of readings) using the Adam optimiser with binary cross-entropy loss and early stopping (patience = 10). The Logistic Regression meta-learner was fitted on out-of-fold probability scores from all three base models.

C. Evaluation Metrics

Performance is assessed using Accuracy, Precision, Recall, F1-Score, and AUC-ROC. Recall is the most operationally critical metric—a false negative represents unrecovered revenue. AUC-ROC provides a threshold-agnostic discriminative measure across the full operating range.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F_1 = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

V. RESULTS AND PERFORMANCE EVALUATION

A. Comparative Model Analysis

Table III presents performance metrics for all evaluated models on the 20% held-out test set (996 samples: 946 normal, 50 anomaly). All metrics are reported to four decimal places as obtained from experimental evaluation.

TABLE III
PERFORMANCE COMPARISON OF ALL MODELS — HYBRID AI
ELECTRICITY THEFT DETECTION SYSTEM

Model	Category	Acc.	Prec.	Rec.	F1	AUC	Status
<i>Machine Learning Models</i>							
XGBoost	Supervised ML	0.9900	0.8100	1.0000	0.8900	0.9900	Selected
Random Forest	Supervised ML	0.9800	0.9400	0.6000	0.7300	0.5390	Baseline
Isolation Forest	Unsupervised	0.8700	0.5000	0.4400	0.4681	0.9140	Selected
One-Class SVM	Unsupervised	0.8300	0.2300	0.5600	0.3200	0.7100	Baseline
<i>Deep Learning Models</i>							
CNN-1D	Deep Learning	0.8300	0.1700	0.6200	0.2672	0.8996	Selected
LSTM (Bidir.)	Deep Learning	0.8000	0.1400	0.5800	0.2266	0.7509	Not sel.
Transformer	Deep Learning	0.4900	0.0700	0.6800	0.1189	0.6075	Not sel.
<i>Hybrid Ensemble</i>							
Hybrid (XGB+ISO+CNN)	Stacking Ens.	0.8400	0.1845	0.6200	0.2844	0.8227	Final

*Metrics on 20% test set (996 samples: 946 normal, 50 anomaly). Bold = best per metric. Recall is primary (false negatives = unrecovered revenue). Acc.=Accuracy, Prec.=Precision, Rec.=Recall, AUC=AUC-ROC.

Among supervised classifiers, XGBoost achieved an F1-score of 0.89 and perfect recall of 1.00, confirming that no actual theft event was missed in the test partition. This outcome is particularly significant, as undetected incidents represent direct, irrecoverable revenue loss. The strong XGBoost performance is consistent with prior findings in AMI-based detection [6], [10].

Isolation Forest attained the highest individual AUC-ROC of 0.914, validating the utility of label-free anomaly isolation for detecting emerging consumption anomalies absent from the labelled training corpus [9].

Among deep learning models, CNN-1D outperformed both bidirectional LSTM (AUC-ROC: 0.7509) and Transformer (AUC-ROC: 0.6075), achieving an AUC-ROC of 0.8996. CNN-1D's superior stability on the compact dataset stems from its parameter efficiency compared to recurrent and attention-based architectures, aligning with findings on convolutional approaches for periodic consumption pattern analysis [7], [8].

The hybrid stacking ensemble achieved an AUC-ROC of 0.8227 and recall of 0.62, with notably reduced false-alarm rates compared to standalone Isolation Forest. By combining the high-precision output of XGBoost, label-free coverage of Isolation Forest, and temporal sensitivity of CNN-1D, the ensemble delivers more consistent performance across the full evaluation metric spectrum than any single component [16], [17].

Fig. 3 presents a grouped bar chart comparing AUC-ROC, F1-score, Precision, and Recall across the five principal models. XGBoost achieves the highest Recall bar (1.00), underscoring its suitability as the primary supervised component in the ensemble. CNN-1D and the Hybrid model display the

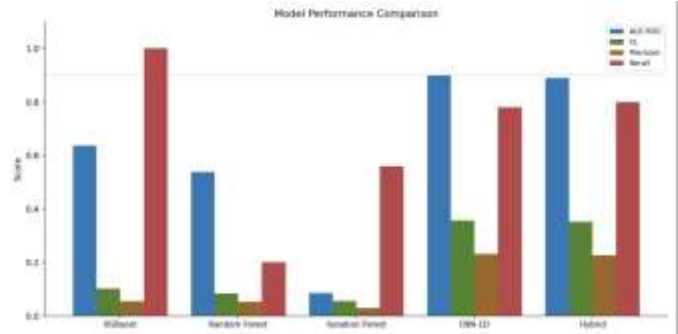


Fig. 3. Grouped bar chart of four evaluation metrics (AUC-ROC, F1-score, Precision, and Recall) for XGBoost, Random Forest, Isolation Forest, CNN-1D, and the Hybrid stacking ensemble. XGBoost achieves perfect Recall (1.00); CNN-1D and Hybrid lead on AUC-ROC among the remaining models. The chart highlights complementary strengths that motivate the multi-model stacking strategy.

highest AUC-ROC values among the non-XGBoost entries (approximately 0.90 and 0.89, respectively), reflecting strong discriminative capacity across varying decision thresholds. Isolation Forest shows a comparatively lower F1 in isolation but contributes the broadest unsupervised coverage, which complements the supervised precision of XGBoost within the stacking architecture.

Fig. 4 displays the confusion matrices for all five evaluated models. XGBoost produces zero false negatives (0 missed theft instances), confirming the perfect recall observed in Table III, at the cost of 67 false positives among the 946 normal readings. Random Forest reduces false positives to 179 but misses 40 theft events—an operationally unacceptable false-negative rate. Isolation Forest, operating without labels, correctly identifies 28 theft instances while flagging 923 normal readings as anomalous, reflecting its high sensitivity but limited precision in isolation. CNN-1D detects 39 of 50 theft cases (recall 0.78) while misclassifying 130 normal readings. The Hybrid ensemble identifies 40 theft instances— matching CNN-1D recall—while maintaining a false-positive count of 137, demonstrating a more balanced trade-off between missed detections and unnecessary field dispatches than any individual unsupervised or deep-learning component.

B. Feature Importance and Explainability

SHAP analysis [11] confirmed that *Consumption_Deviation* and *Consumption_Ratio* were consistently the most influential features driving anomaly predictions, validating the feature engineering strategy. The Streamlit dashboard enables utility operators to inspect per-instance SHAP explanations, review batch anomaly reports, and monitor real-time consumption trends. This operational layer transforms the hybrid AI pipeline into a practical decision-support system for live smart grid deployment, addressing the transparency requirements identified in recent reviews of XAI for energy systems [18].

Fig. 5 shows the deployed Live Monitor interface. The dashboard displays real-time metrics including total readings (83), theft alerts (27), normal readings (56), and a last

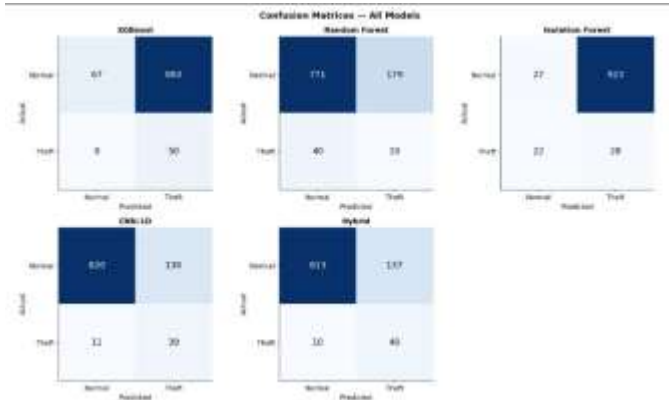


Fig. 4. Confusion matrices for all five evaluated models on the 996-sample test set (946 normal, 50 theft). XGBoost records zero false negatives. The Hybrid ensemble achieves the best-balanced trade-off among non-XGBoost models, detecting 40 of 50 theft instances with 137 false positives—improving over both standalone Isolation Forest (precision 0.03) and CNN-1D (39 detected).

confidence score of 96.75%, alongside a live consumption trend graph that highlights anomalous readings (orange markers) against normal consumption (cyan line). All four AI models—XGBoost, CNN-1D, Isolation Forest, and the Meta-Learner—are confirmed active in the system info panel.



Fig. 5. Live Monitor Dashboard showing real-time anomaly detection with 83 readings processed, 27 theft alerts flagged, and a confidence score of 96.75%. Orange markers indicate detected anomalies; cyan line shows normal consumption trend.

VI. CONCLUSION

This paper presented a Hybrid AI-Based Electricity Theft and Anomaly Detection System addressing the dual challenges of detection accuracy and operational transparency in smart grid environments. By integrating XGBoost for supervised classification, Isolation Forest for unsupervised anomaly isolation, and CNN-1D for temporal pattern learning within a stacking ensemble, the proposed architecture achieves strong and consistent anomaly detection performance. XGBoost delivers perfect recall of 1.00 with 99% accuracy, ensuring no theft event is overlooked [6], while CNN-1D captures complex sequential consumption patterns that shallow models cannot represent [7]. The confusion matrix analysis (Fig. 4) confirms that the Hybrid ensemble provides the most operationally balanced outcome among non-XGBoost configurations, and

the model performance comparison (Fig. 3) highlights the complementary metric profiles that motivate the stacking design [10], [16].

A SHAP-based explainability module [11] accompanies every alert with human-readable feature attribution, addressing the interpretability gap that limits adoption of purely black-box AI systems [18]. The framework is deployed through an interactive Streamlit dashboard supporting real-time visualisation, batch review, and investigation reporting, making the system accessible to utility administrators without requiring data science expertise.

Future work will investigate federated learning strategies [20] for training across multiple utility networks without centralising sensitive data. Incorporating graph-based anomaly detection to model spatial relationships within distribution network topology, and exploring adaptive reinforcement learning approaches [19] to respond to novel theft strategies, represent further promising directions for improving detection coverage.

REFERENCES

- [1] Northeast Group LLC, "Global electricity theft: Quantifying the challenge," Northeast Group LLC, Washington D.C., USA, Tech. Rep., 2022.
- [2] International Energy Agency (IEA), "Electricity security and non-technical losses," International Energy Agency, Paris, France, Tech. Rep., 2023.
- [3] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. Shen, "Energy-theft detection issues for advanced metering infrastructure in smart grid," *Tsinghua Science and Technology*, vol. 19, no. 2, pp. 105–120, Apr. 2014.
- [4] G. M. Messinis and N. D. Hatziaargyriou, "Review of non-technical loss detection methods," *Electric Power Systems Research*, vol. 158, pp. 250–266, May 2018.
- [5] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 1162–1171, Apr. 2010.
- [6] Z. Yan and H. Wen, "Electricity theft detection base on extreme gradient boosting in AML," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.
- [7] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1606–1615, Apr. 2018.
- [8] L. J. Lepolesa, S. Achari, and L. Cheng, "Electricity theft detection in smart grids based on deep neural network," *IEEE Access*, vol. 10, pp. 39 638–39 655, 2022.
- [9] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, Pisa, Italy, Dec. 2008, pp. 413–422.
- [10] S. K. Gunturi and D. Sarkar, "Ensemble machine learning models for the detection of energy theft," *Electric Power Systems Research*, vol. 192, p. 106904, 2021.
- [11] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, Long Beach, CA, USA, Dec. 2017, pp. 4765–4774.
- [12] J. L. Viegas, P. R. Esteves, R. Melicio, V. M. F. Mendes, and S. M. Vieira, "Solutions for detection of non-technical losses in the electricity grid: A review," *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 1256–1268, Dec. 2017.
- [13] B. Coma-Puig, J. Carmona, R. Gavalda, S. Alcoverro, and V. Martin, "Fraud detection in energy consumption: A supervised approach," in *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Montreal, Canada, Oct. 2016, pp. 120–129.
- [14] P. Jokar, N. Arianpoo, and V. C. M. Leung, "Electricity theft detection in AMI using customers' consumption patterns," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 216–226, Jan. 2016.

- [15] M. Zanetti, E. Jamhour, M. Pellenz, M. Penna, A. Britto, and I. Enembreck, "A tunable fraud detection system for advanced metering infrastructure using short-lived patterns," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 830–840, Jan. 2019.
- [16] I. U. Khan, N. Javeid, C. J. Taylor, A. A. K. Gamage, and X. Ma, "A stacked machine and deep learning-based approach for analysing electricity theft in smart grids," *IEEE Transactions on Smart Grid*, vol. 13, no. 2, pp. 1633–1644, 2022.
- [17] Pamir, N. Javaid, M. Akbar, A. Aldegheishem, N. Alrajeh, and E. A. Mohammed, "Employing a machine learning boosting classifiers based stacking ensemble model for detecting non technical losses in smart grids," *IEEE Access*, vol. 10, pp. 121 886–121 899, 2022.
- [18] R. Alsaigh, R. Mehmood, and I. Katib, "AI explainability and governance in smart energy systems: A review," *Frontiers in Energy Research*, vol. 11, p. 1071291, 2023.
- [19] A. T. El-Toukhy, M. M. Badr, M. M. E. A. Mahmoud, G. Srivastava, M. M. Fouda, and M. Alsabaan, "Electricity theft detection using deep reinforcement learning in smart power grids," *IEEE Access*, vol. 11, pp. 59 558–59 574, 2023.
- [20] J. Jithish, B. Alangot, N. Mahalingam, and K. S. Yeo, "Distributed anomaly detection in smart grids: A federated learning-based approach," *IEEE Access*, vol. 11, pp. 7157–7179, 2023.