BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

IoT and Machine Learning for Real-Time Energy Theft Detection in Smart Grids : A Review of IoT-Based Monitoring and Machine Learning Approaches

Sarthak Deore

Department of Artificial Intelligence and Data Science Engineering
Ajeenkya DY Patil School of Engineering, Lohegaon, Pune, Maharashtra, India

Dr. Divya Sharma

Guide

Abstract

Electricity theft continues to create practical and financial difficulties for power utilities. It contributes to non-technical losses, reduces revenue, makes demand estimates less dependable, and can affect the operation of distribution networks. Inspections and simple threshold-based checks remain useful in some situations, but they can be slow when a utility must monitor a large number of consumers. This review examines how the Internet of Things (IoT), smart meters, edge computing, and machine learning (ML) could be combined to screen electricity-use data and identify patterns that deserve closer attention. Smart meters record consumption over time, and the readings can be sent to an edge gateway or a utility platform. Before analysis, the data need to be checked and prepared so that a model can work with meaningful features. The approaches considered include support vector machines, random forests, neural networks, CNN/LSTM models, hybrid methods, active learning, and techniques for dealing with imbalanced datasets. The paper also describes a conceptual system design and explains how a model score and a selected threshold can be used to raise an alert. Precision, recall, F1-score, false-positive rate, and detection latency are discussed as measures of performance. Local processing may help reduce response time and the amount of detailed data sent to a central system; central platforms, meanwhile, can support wider analysis and model maintenance. In practice, the results will depend on data quality, dependable labels, secure communication, and responsible handling of customer information. An alert should prompt a review, not be treated as proof of theft. For this reason, IoT–ML detection is considered here as a support tool that needs careful testing and human oversight. Keywords: Internet of Things, smart grid, machine learning, electricity theft detection, smart meters, edge computing, anomaly detection, data privacy.

Keywords : Internet of Things, smart grid, machine learning, electricity theft detection, smart meters, edge computing, anomaly detection, data privacy.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Introduction

Electricity theft remains a difficult issue for power providers. It can cause financial losses, make it harder to estimate how much electricity is actually being used, and create concerns for the reliable operation of distribution networks. In many cases, utilities still depend on site visits, simple consumption rules, or checks carried out after data have been collected. Such methods may be difficult to apply quickly when readings arrive from a large number of meters. Smart meters provide a more regular picture of consumption, and machine-learning techniques can help find patterns that differ from a customer's usual usage. Some of this analysis may be performed near the meter using edge computing. Still, collecting more data does not automatically produce reliable results. The system needs suitable examples for training, consistent data preparation, dependable communication, and safeguards for customer information. Its predictions should be tested before they are used to guide inspections or other decisions.

Problem Statement

The main question addressed in this review is how a utility can identify unusual electricity-use patterns across many consumers in a timely and dependable way. Consumption is not the same for every household or business, and it can change from one period to another. As a result, a single fixed limit may miss suspicious activity or flag an ordinary change in demand. An IoT–ML solution must be able to work with frequent readings while accounting for missing or noisy data, the small number of confirmed theft cases, communication delays, and false alarms. The design must also protect customer information and consider the security of connected devices and the data they transmit.

Objectives

The review is guided by five objectives:

• Explain how IoT devices and smart meters gather electricity-use readings.

• Review machine-learning methods that have been considered for possible electricity-theft detection.

• Describe a conceptual workflow from data collection and preparation to model output and alert review.

• Discuss the measures used to assess performance and the difficulties that can arise in practice.

• Identify requirements for a scalable approach that takes edge processing and privacy into account.

Relevance To Current Technology Trends

Modern smart grids increasingly use connected meters and data analysis to understand electricity consumption. IoT devices can collect readings at regular intervals, while edge computing allows some of the analysis to take place close to where the data are produced. This can help a system respond sooner and may reduce the need to send detailed readings to a central platform. Utility platforms still have an important role: they can compare patterns across many meters, maintain models, and support coordinated monitoring. A design that combines edge and cloud processing can divide the work between them. The choice depends on how quickly an alert is needed, what computing resources are available, how much communication costs, and what privacy protections are required.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Energy-Theft Detection Techniques

A. Data Collection

Smart meters form the main source of time-stamped electricity-consumption data. If the equipment supports it, the system may also use event logs from meters, voltage or current measurements, and other relevant context. Every record needs a valid meter identifier and a timestamp so readings can be placed in the right order. Before the data are analysed, the collection stage should look for missing intervals, repeated records, delayed readings, and values that fall outside reasonable limits. Without these checks, an error in the data could be mistaken for an unusual pattern in a customer's use.

B. Preprocessing and Feature Extraction

Raw readings are not always ready to be used by a model. Depending on the dataset, preparation may involve checking values, dealing with gaps, scaling selected features, and arranging readings into consistent time intervals. The system can then calculate indicators such as a customer's daily usage profile, peak-to-average ratio, abrupt changes, or differences from their earlier consumption. It is important to use the same preparation process during training and after deployment. If the input data are handled differently in the two stages, the model's predictions may become less reliable.

C. Conventional and Deep Learning Methods

The methods considered in this area do not all work in the same way. A support vector machine uses a learned boundary to separate examples, whereas a random forest combines the results of multiple decision trees and can offer clues about which features influence its output. Neural networks can learn more complicated relationships, and CNN/LSTM approaches may be useful when both short-range patterns and changes over time matter. Hybrid models bring together two or more techniques. To compare these options fairly, researchers should use consistent data splits and comparable testing conditions. A model with more layers or greater complexity is not necessarily more useful if it is expensive to run, hard to interpret, or performs poorly outside the training data.

D. Active Learning and Imbalanced Data

Records of confirmed electricity theft are usually much fewer than records of normal consumption. This imbalance can make accuracy misleading: a model might classify most records correctly but still miss many of the cases the utility wants to find. Resampling or giving additional weight to rare cases during training may help, though the effect needs to be checked on separate evaluation data. Active learning can also be considered. It selects uncertain or informative examples for review by a human expert, who can provide or verify a label. These methods do not remove the need for good-quality labels and repeated validation.

E. Suspicion Score and Alert Threshold

Let x stand for features calculated from recent meter readings, and let f(x) represent the model's estimated probability that the pattern is linked to theft. An alert can be generated when $f(x) \geq \tau$, with τ representing a threshold selected using validation results and the utility's priorities. The choice involves a trade-off. A lower threshold may catch more suspicious patterns, but it can also send more innocent customers for inspection. A higher threshold may reduce unnecessary visits while allowing some cases

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

to pass unnoticed. The score is therefore a signal for further checking; it should not be treated as a final finding of wrongdoing.

F. Evaluation Measures

A useful evaluation should consider more than accuracy alone. Precision tells us what share of the cases flagged by the model are positive, while recall shows how many of the known theft cases were found. The F1-score combines precision and recall, and the false-positive rate and time taken to detect a pattern are also important for deployment. Where theft is rare, a high accuracy figure can hide poor detection of the cases that matter most. Testing should use data that were not used for training. Researchers should also guard against data leakage—for example, when readings from the same customer or a later time period unintentionally appear in both training and test sets. Results are most useful when the test conditions resemble the setting in which the system will operate.

Proposed IoT-ML System Architecture

The proposed architecture begins with smart meters and sensors that send readings through a communication network to an edge gateway or a utility platform. The incoming data are checked and prepared before being passed to the machine-learning component. When the model's output meets the selected alert rule, the system sends the case for review by utility staff. This architecture is a proposed framework for implementation and evaluation. It should not be read as a claim that a new system has already been built or deployed.

A. Model Inference and Alerts

Once trained, a classifier or anomaly detector can assign a score to a meter or a chosen period of its consumption. If the score crosses the selected threshold, an alert is created along with information that may help utility staff understand why the case was flagged. A gateway could run a smaller model close to the data source, while a central platform handles more detailed analysis, monitors groups of meters, and coordinates model updates. Whichever arrangement is used, communication should be protected, access should be controlled, and model updates should be checked for authenticity.

B. Edge and Cloud Processing

Running part of the analysis at the edge can make alerts available sooner and reduce the volume of raw readings sent to a central system. A central platform has other strengths: it can support larger models, compare data across groups of meters, and manage updates in a consistent way. A combined design may therefore use edge devices for initial screening and central systems for broader analysis. There is no single arrangement that suits every utility. Connectivity, device capability, operating cost, and privacy requirements all influence the decision.

C. Operational Applications and Advantages

If tested successfully, this type of system could help a utility decide which sites to inspect first, identify unusual consumption, examine losses along feeders, plan audits, and look for patterns that recur over time or across locations. Automated screening may help staff direct their time towards cases that need attention and could be extended to many meters. Models may also need to be reviewed as customer usage changes. These advantages should be treated as possibilities, not guaranteed results. Field testing,

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

suitable comparison data, and a clear process for checking mistaken alerts are needed to determine whether the system provides real benefits.

Challenges, Limitations And Ethical Concerns

A. Data and Model Limitations

Several issues can reduce a model's reliability after deployment. The training data may be incomplete, outdated, noisy, or unrepresentative of the customers who will later be monitored. Because confirmed theft cases are uncommon, the model may overlook them; changes in consumption patterns can also cause model drift. Some advanced models need more computing power and may be harder to explain to utility staff or customers. False alarms are a practical and ethical concern as well. They can waste inspection time and may lead to unfair treatment if a risk score is accepted as proof.

B. Privacy and Cybersecurity

Electricity readings collected at fine time intervals can sometimes reveal routines or activities in homes and businesses. Utilities should therefore limit collection to information that is needed for the intended task and protect records both in transit and in storage. Access permissions, retention periods, and monitoring of data access should be clearly defined. Since smart meters and model-update systems are connected devices, they also need protection against unauthorized access, interference, and tampering.

C. Human Review and Accountability

A model's score should never be enough, by itself, to accuse a customer of theft or impose a penalty. Utility staff should examine the available evidence, check whether there are reasonable explanations for the pattern, and follow the relevant procedures before acting. The system should also be monitored to see whether its performance changes across customer groups or operating conditions. When an alert is wrong, there needs to be a documented way to challenge it and correct the record.

Research Gap

A number of questions still need further investigation. Reliable detection depends on historical data with trustworthy labels, but real meter datasets may contain gaps, errors, or noise. Edge devices also face limits in computing power, and communication delays can affect timely analysis. In some cases, a system may identify an unusual pattern without showing exactly where the problem began. Privacy and cybersecurity remain important throughout deployment. The material reviewed for this paper does not provide one shared benchmark with comparable numerical results for every method, so a firm numerical ranking cannot be justified here. Future studies should test the approaches on consistent datasets, report their evaluation methods clearly, and consider realistic operating conditions.

Proposed Working Approach

The proposed workflow is conceptual and is intended to explain how the technologies may work together. It is not a report of a completed implementation.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Stage	Description
Step1—Data collection	Smart meters and IoT sensors record electricity consumption readings at regular intervals.
Step2—Communication	Readings are sent through a suitable communication network to a local gateway or processing service
Step3—Data preparation	The system checks missing, noisy, invalid, or duplicated readings and organizes the data in a consistent format.
Step4—Feature extraction	Features such as interval consumption, changes over time, and deviations from a customer's usual pattern are examined.
Step5—ML-based detection	A trained classifier or anomaly-detection method assigns a score or class to the observed consumption pattern.
Step6—Edge/cloud processing	Initial processing or inference may run near the source to reduce latency; central systems can support longer term analysis and model management.
Step 7—Alert and verification	A suspicious pattern generates an alert with relevant readings and a timestamp. Utility personnel review the evidence before deciding on inspection or other action.
Step 8—Feedback and maintenance	Verified investigation outcomes maybe used to evaluate and update the model through a controlled process.

Conclusion

The intended outcome of the literature-based seminar is an appreciation of how connected meters and ML can be used to enable continuous monitoring, earlier detection of anomalous consumption and improved grid efficiency. It looks at a number of method families, the 6role of edge computing and the practical issues of data quality, class imbalance, latency, computing resources, privacy and security. IoT based monitoring and ML based analysis represents a very promising avenue for electricity theft detection in smart grids. Smart meters provide a continuous stream of consumption data and machine learning algorithms can be used to identify anomalies that might suggest theft. Edge computing could be used to improve response times by processing a subset of the data at the edge. The real-world effectiveness of such an approach will of course be contingent upon the availability of representative data, the ability of the model to generalise, the quality of the evaluation, the ease of deployment and of course robust privacy and security. The workflow outlined here forms a basis upon which further implementation and experimental evaluation could be carried out. It is not in any way claiming that such a system has been deployed or tested in the real world.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

References

1. N. M. Elshennay, D. M. Ibrahim, and A. M. Gab Allah, "An Efficient Electricity Theft Detection Based on Deep Learning" Scientific Reports, vol. 15, Article 12866, 2025.
2. T. Niu, H. Yue, T. AlSkaif, and M. Cui, "ETDS: A Novel Electricity Theft Detection System for Highly Unbalanced Data in Smart Grids," IEEE Transactions on Industrial Informatics, vol. 21, no. 3, pp. 2114–2123, 2025.
3. T.-D. Le, N. T. M. Duy, T. H. B. Huy, et al., "Advanced Deep Learning-Based Electricity Theft Detection in Smart Grids Using Multi-Dimensional Analysis with Convolutional Autoencoder and Transformer," Engineering Applications of Artificial Intelligence, 2025.
4. "Robust Resampling and Stacked Learning Models for Electricity Theft Detection in Smart Grid," Energy Reports, vol. 13, pp. 770–779, 2025.
5. K. Blazakis, N. Schetakis, M. M. Badr, et al., "Power Theft Detection in Smart Grids 7Using Quantum Machine Learning," IEEE Access, vol. 13, 2025.
6. H. Iftikhar, N. Khan, M. A. Raza, et al., "Electricity Theft Detection in Smart Grid Using Machine Learning," Frontiers in Energy Research, vol. 12, 2024.
7. S. Zhu, Z. Xue, and Y. Li, "Electricity Theft Detection in Smart Grids Based on Omni-Scale CNN and Auto XGB," IEEE Access, vol. 12, 2024.