

Scientiarum: A Multidisciplinary Journal
Vol. 2 No. S1 (2026): Supplement Issue
DOI: 10.54646/SAPARS.2026.81

SAPAR SERVICES LLP
Scholarly Articles Publication And Repository Services

A RESEARCH PAPER ON SMART METER MONITORING, ANOMALY DETECTION AND MACHINE LEARNING-BASED IDENTIFICATION OF NON-TECHNICAL LOSSES

1. Manjunatha S, 2. Adarshakumar V

1. Head of the Department, Master of Computer Applications (MCA), SEA College of Science, Commerce and Arts, K. R. Puram, Bangalore-560049. 2. Project Research Scholar Master of Computer Applications (MCA) SEA College of Science, Commerce and Arts, K. R. Puram,

Bangalore-560049.

Abstract

Electricity misuse, including unauthorized consumption, meter tampering, bypass connections, billing manipulation and other forms of non-technical loss, creates financial, operational and safety challenges for electricity distribution systems. Conventional inspection methods are often dependent on periodic field visits and predefined rules, which can be costly and may fail to identify changing consumption patterns. This paper presents a conceptual framework for monitoring and detecting electricity misuse using smart-meter data, statistical analysis, anomaly detection and machine learning. The proposed approach collects time-series consumption data, preprocesses and extracts meaningful features, identifies abnormal patterns, assigns a misuse risk score and prioritizes suspicious cases for authorized inspection. The study reviews classification, ensemble-learning, deep-

learning and pattern-based approaches reported in recent research. Existing studies show that machine-learning methods can improve the identification of suspicious customers, while also highlighting challenges such as class imbalance, false positives, privacy, changing consumer behavior and the need for field verification. The proposed framework emphasizes that automated detection should support—not replace—authorized inspection and due process. The paper concludes that combining smart metering, data analytics, explainable models and secure operational procedures can contribute to more efficient monitoring of electricity distribution networks.

Keywords

Electricity misuse, electricity theft, non-technical losses, smart meter, smart grid, anomaly detection, machine learning, energy monitoring, data analytics.

Introduction

Electricity is an essential resource for households, industries, commercial establishments and public services. Reliable distribution requires utilities to measure consumption accurately, maintain network infrastructure and recover the cost of electricity supplied. Electricity misuse is one of the factors that can reduce the accuracy of this process. It may include unauthorized connections, meter tampering, bypassing meters, manipulation of metering information and other forms of unaccounted consumption. Such activity is generally discussed within the broader category of non-technical losses (NTLs), which also includes some billing and measurement problems.

Traditional methods for identifying suspicious consumption commonly rely on inspections, complaints, meter testing and rule-based comparisons. These methods remain important because physical verification is necessary before enforcement

action. However, manual inspection of a large customer base is resource-intensive. The growth of smart meters and advanced metering infrastructure (AMI) provides utilities with more frequent consumption measurements that can be analyzed computationally. Recent research therefore investigates machine learning, deep learning, anomaly detection and pattern-recognition techniques for identifying customers whose consumption behavior differs from expected patterns.

Problem Statement

Electricity distribution utilities may receive large volumes of meter readings that are difficult to examine manually. A simple threshold-based method may also generate false alarms because legitimate consumption can change due to weather, occupancy, holidays, business activity, seasonal demand or changes in appliances. The research problem is therefore to develop a monitoring framework capable of identifying unusual electricity-consumption behavior while reducing unnecessary inspections. The system must distinguish between normal variation and patterns that warrant further investigation, while preserving consumer privacy and maintaining appropriate verification procedures.

Objectives

The major objectives of the proposed research are:

To monitor electricity-consumption patterns using meter data.

To identify abnormal or inconsistent consumption behavior.

To apply data preprocessing and feature extraction for reliable analysis.

To investigate machine-learning and anomaly-detection methods for misuse identification.

To reduce unnecessary manual inspection by prioritizing higher-risk cases.

To provide interpretable information that can assist authorized field personnel.

To consider privacy, security and fairness while processing consumer data.

Literature Review

Research on electricity misuse detection has increasingly moved from manual inspection toward data-driven approaches. Pattern-based methods use historical consumption profiles and contextual information to calculate anomaly or suspicion scores. One published approach combined Dynamic Time Warping with k-nearest neighbors and reported an F1-score of 94% with a false-positive rate of 1.1% on its evaluation setting. Such work demonstrates the value of comparing consumption patterns rather than relying only on fixed thresholds.

Machine-learning classification has also been widely investigated. A 2023 study proposed an XGBoost-based approach using consumption patterns and reported precision, recall and F1-score values of approximately 96%, 95% and 95% respectively in its experimental setup. Deep-learning methods have similarly been investigated to address complex temporal patterns and imbalanced datasets. A 2023 study described a two-stage deep-learning approach designed to narrow the scope of on-site inspections by identifying suspicious metering patterns.

More recent research continues to investigate ensemble models, feature selection and imbalance-handling techniques. A 2026 study evaluated several tree-based algorithms and reported strong results for Extra Trees with SMOTE and feature selection on its selected dataset. These results are promising, but performance values from different datasets and experimental designs should not be interpreted as directly comparable. A recurring challenge is that fraudulent examples are usually much less common than legitimate examples, making class imbalance an important consideration.

Existing System

The existing approach in many conventional distribution environments depends on periodic meter reading, customer complaints, physical inspection, billing analysis and rule-based identification. Inspectors may examine meters and service connections when there is evidence of irregularity. While physical inspection provides direct evidence, inspecting every customer at regular intervals is impractical for large networks. Rule-based systems can also struggle when misuse patterns change or when legitimate customers exhibit unusual demand.

Proposed System

The proposed system is a data-driven electricity misuse monitoring framework. Smart-meter or meter-reading data are collected and stored securely. The data are cleaned, normalized and transformed into analytical features such as daily consumption, rolling averages, consumption variability, peak-to-average ratio and changes from historical patterns. An anomaly-detection or classification model then produces a risk score. Cases above an operational threshold are prioritized for authorized inspection. The framework can also retain explanations, such as an unusually large deviation from a customer's historical profile, to support transparent investigation.

System Architecture

The conceptual architecture consists of six major layers: (1) Meter/Data Collection Layer, (2)

Secure Data Storage Layer, (3) Data Preprocessing Layer, (4) Feature Engineering Layer, (5) Detection and Risk Scoring Layer, and (6) Monitoring and Inspection Support Layer. The data collection layer receives meter readings. The storage layer maintains historical records. Preprocessing handles missing values, duplicates and inconsistent readings. Feature engineering converts raw readings into useful

indicators. The detection layer applies statistical or machine-learning methods. Finally, the monitoring layer presents prioritized cases to authorized personnel for review and field verification.

Methodology

The methodology begins with data acquisition from available meter readings or a suitable research dataset. Data preprocessing includes removal or treatment of invalid observations, missing-value handling, time alignment and normalization where appropriate. Feature engineering is then performed to capture both short-term and long-term behavior.

The next stage is model development. Supervised classification can be used when verified labels for legitimate and misuse cases are available. When reliable labels are limited, unsupervised or semi-supervised anomaly detection can identify observations that differ strongly from established normal patterns. Candidate models include Logistic Regression, Decision Trees, Random Forest, XGBoost, Support Vector Machines, k-nearest neighbors and neural-network models. The model should be evaluated using precision, recall, F1-score, ROC-AUC or PR-AUC as appropriate. Because misuse cases are often rare, accuracy alone should not be treated as the primary evaluation measure.

The final stage converts model outputs into a risk-prioritization list. The list should support authorized investigation rather than automatically declaring a customer guilty of misuse. Field findings can later be used as verified labels for model improvement.

Data Preprocessing and Feature Engineering

Data quality strongly affects detection performance. The preprocessing stage can identify missing readings, duplicate records, abnormal timestamps and inconsistent

meter values. Historical consumption can be aggregated into daily or weekly profiles. Useful features may include average consumption, minimum and maximum demand, standard deviation, coefficient of variation, peak-hour consumption, weekend-versus-weekday difference and deviation from seasonal baselines. Context variables such as weather, holidays or customer category may also help reduce false positives when legally and operationally appropriate.

Detection Techniques

Several analytical approaches can be combined. Statistical techniques can compare current consumption against historical baselines. Distance-based methods can identify customers whose patterns differ from similar customers. Classification algorithms can learn from verified examples. Ensemble models combine multiple decision structures and can perform well on tabular data. Neural networks and recurrent or convolutional architectures can model complex time-series relationships. Hybrid systems may combine an anomaly score with a supervised classification score. Explainability methods can be used to show which features contributed to a high-risk prediction.

Monitoring and Alert Generation

The monitoring dashboard can display customer or meter identifiers in a controlled environment, recent consumption trends, anomaly scores, model confidence and the reason for prioritization. Alerts should be categorized by severity and reviewed according to utility procedures. A high anomaly score should trigger investigation rather than immediate enforcement. This distinction is important because legitimate changes in consumption can resemble suspicious patterns. The system should also maintain an audit trail showing when an alert was generated, reviewed and resolved.

Performance Evaluation

A research implementation can be evaluated using a confusion matrix containing true positives, true negatives, false positives and false negatives. Precision indicates the proportion of flagged cases that are confirmed positive in the labeled test set, while recall measures the proportion of positive cases detected. The F1-score combines precision and recall. ROC-AUC summarizes ranking performance across thresholds, while PR-AUC can be particularly informative for imbalanced datasets. Evaluation should use a separated test set and, where possible, time-aware validation so that future consumption is not inadvertently used to predict the past.

Security and Privacy

Electricity consumption data can reveal sensitive information about usage patterns. A practical monitoring system should therefore implement access control, encryption, secure authentication, logging and data-minimization principles. Only authorized personnel should access personally identifiable information. Research datasets should be anonymized or pseudonymized where possible. Model outputs should be treated as decision-support information, and operational actions should follow applicable utility rules, consumer-protection requirements and legal procedures.

Challenges and Limitations

The major challenges include limited labeled misuse data, class imbalance, changing customer behavior, noisy measurements, seasonal effects and model drift. Attackers may also change their behavior after detection systems are deployed. A model trained on one geographic region or customer population may not perform equally well in another. Reported accuracy or F1-score from a research dataset therefore cannot be assumed to represent field performance. Another limitation is that algorithmic detection cannot by itself establish physical tampering or unauthorized

connection; inspection and appropriate evidence remain necessary.

Expected Outcomes

The proposed framework is expected to improve the ability of utilities to screen large quantities of consumption data and prioritize cases for inspection. It can reduce dependence on purely manual screening, provide earlier visibility into unusual patterns and create a structured feedback loop between analytics and field verification. The expected outcome is not simply a higher number of alerts, but a more useful prioritization process with controlled false positives and interpretable evidence.

Future Scope

Future work can explore real-time edge analytics, federated learning, explainable artificial intelligence, multimodal data fusion and continual learning. Research can also examine integration of smart-meter voltage and power-quality measurements with consumption data. Recent literature identifies real-time detection, continual learning, multimodal data and standardized benchmarking as important research directions. Future systems should also investigate robust privacy-preserving architectures and methods for adapting models when customer behavior changes over time.

Conclusion

Monitoring and detecting electricity misuse is an important application of smart-grid analytics. Traditional inspection remains essential, but large-scale meter data creates an opportunity to use statistical and machine-learning methods to identify unusual consumption patterns and prioritize investigations. A robust system should combine secure data collection, preprocessing, feature engineering, anomaly detection, explainable risk scoring and human verification. Research results demonstrate that

modern machine-learning techniques can achieve useful detection performance under controlled experimental conditions, while also showing that class imbalance and false positives remain important challenges. Therefore, electricity misuse detection should be treated as a continuous monitoring and decision-support process rather than a fully automated enforcement mechanism.

References

Ahir, R. K., & Chakraborty, B. (2022). Pattern-based and context-aware electricity theft detection in smart grid. *Energy Reports*.

Ali Abro, S., Laghari, J. A., Memon, S. A., et al. (2025). Non-technical loss detection in power distribution networks using machine learning. *Scientific Reports*.

Apata, O., Ratshitanga, M., & Davidson, I. E. (2026). Machine learning techniques for electricity theft detection in smart grids: A comprehensive review. *Energies*.

Aslam, S., et al. (2023). Deep learning-based meta-learner strategy for electricity theft detection. *Frontiers in Energy Research*.

Emadaleslami, M., Haghifam, M.-R., & Zangiabadi, M. (2023). A two-stage approach to electricity theft detection in AMI using deep learning. *International Journal of Electrical Power & Energy Systems*.

Kawoosa, A. I., Prashar, D., Faheem, M., Jha, N., & Khan, A. A. (2023). Using machine learning ensemble method for detection of energy theft in smart meters. *IET Generation, Transmission & Distribution*.

Pamir, M., et al. (2023). Electricity theft detection for energy optimization using deep learning models. *Energy Science & Engineering*.

Pengwah, A. B., Razzaghi, R., & Andrew, L. L. H. (2023). Model-less non-technical loss detection using smart meter data. *IEEE Transactions on Power Delivery*.