Jurnal
Teknologi

Full Paper

SUPERVISED MACHINE LEARNING FOR ELECTRICITY THEFT DETECTION: A SYSTEMATIC REVIEW OF TRENDS, CHALLENGES, AND FUTURE DIRECTIONS

Mohammed Ateequr Rahaman*, Rasyidah Mohamad Idris

Faculty of Electrical Engineering, Engineering Campus, 81310, UTM Johor Bahru, Johor, Malaysia

Article history
Received
24 December 2024
Received in revised form
31 October 2025
Accepted
25 November 2025
Published Online
29 August 2026

*Corresponding author
mohammedrahaman@graduate.utm.my

Graphical abstract

Abstract

Electricity theft causes substantial financial losses and grid instability, requiring data-driven detection solutions, especially through supervised machine learning (SML) techniques. This review systematically examines 50 studies (2012–2024) from Scopus, IEEE Xplore, and Web of Science, following PRISMA guidelines to ensure a thorough and high-quality evaluation. It focuses on key themes such as algorithms, dataset challenges, performance metrics, and scalability, which are crucial for advancing electricity theft detection (ETD). SML methods such as decision trees, gradient boosting, and deep learning demonstrate high accuracy but encounter challenges, including inconsistent evaluation metrics, limited scalability for large-scale grids, and reliance on simulated datasets. Notably, only 28% of studies consider real-world scalability, and 54% lack robust evaluation metrics. Additionally, data issues such as missing values and class imbalance exacerbate these limitations. Future research should focus on developing hybrid models that effectively balance accuracy and interpretability, establishing standardized benchmarks for comparison, and incorporating real-world validation to enhance scalability and reliability in smart grid environments.

Keywords: Electricity Theft Detection, Supervised Machine Learning, Systematic Literature Review, Performance Evaluation Metrics, Scalability, Interpretability

Abstrak

Kecurian elektrik menyebabkan kerugian kewangan yang besar dan ketidakstabilan grid, sekali gus memerlukan penyelesaian pengesanan berasaskan data, terutamanya melalui teknik pembelajaran mesin terselia (SML). Kajian ini menilai secara sistematik sebanyak 50 kajian (2012–2024) daripada pangkalan data Scopus, IEEE Xplore dan Web of Science, dengan menggunakan garis panduan PRISMA bagi memastikan penilaian yang menyeluruh dan berkualiti tinggi. Kajian ini memberi tumpuan kepada tema utama seperti algoritma, cabaran set data, metrik penilaian prestasi, dan kebolehskalaan yang penting untuk memajukan pengesanan kecurian elektrik (ETD). Kaedah SML seperti pokok keputusan, penggalakan kecerunan dan pembelajaran mendalam menunjukkan ketepatan yang tinggi tetapi menghadapi beberapa cabaran termasuk ketidakselarasan metrik penilaian, kebolehskalaan terhad untuk grid berskala besar, serta kebergantungan terhadap set data simulasi. Secara khususnya, hanya 28% daripada kajian mempertimbangkan kebolehskalaan dunia sebenar dan 54% kekurangan kerangka penilaian yang kukuh. Selain itu, isu data seperti nilai hilang dan ketidakseimbangan kelas memburukkan lagi had tersebut. Penyelidikan masa hadapan perlu memberi tumpuan kepada model hibrid yang dapat mengimbangkan ketepatan dan kebolehtafsiran, mewujudkan penanda aras piawai untuk perbandingan, serta menggabungkan pengesahan dunia sebenar bagi meningkatkan kebolehskalaan dan kebolehpercayaan dalam persekitaran grid pintar.

Kata kunci: Pengesanan Kecurian Elektrik, Pembelajaran Mesin Terselia, Ulasan Literatur Sistematik, Metrik Penilaian Prestasi, Kebolehskalaan, Kebolehtafsiran.

1.0 INTRODUCTION

The shift from conventional power grids to smart grids has transformed the energy sector by enabling two-way communication, real-time monitoring, and improved reliability. These advancements have enhanced system performance, increased consumer engagement, and facilitated the integration of renewable energy sources. However, they have introduced new vulnerabilities, with electricity theft emerging as a critical challenge that undermines grid performance and economic stability [1, 2].

Electricity theft contributes significantly to non-technical losses (NTLs), causing an estimated annual global financial loss of nearly USD 96 billion. In developed countries such as the United States, NTLs account for around 6% of total power generation. The impact is even more pronounced in developing regions, where weak enforcement mechanisms and higher theft rates disrupt energy markets and hinder economic growth [3, 4, 5].

To address electricity theft, researchers have developed both hardware- and software-based detection approaches. Hardware-oriented solutions, such as tamper-proof meters, sensor networks, and radio frequency identification (RFID) monitoring, offer high accuracy but remain expensive and difficult to implement on a large scale [6, 7]. Conversely, software-based approaches rely on data-driven techniques, particularly unsupervised and supervised machine learning (USML and SML) methods. USML methods detect irregular consumption patterns without labeled data through clustering and anomaly detection, while SML techniques classify consumer behavior into legitimate or fraudulent categories using labeled datasets [8, 9]. Among these, algorithms such as decision trees, random forests, gradient boosting, and deep learning have demonstrated high detection accuracy in electricity theft detection (ETD) [10, 11]. Recent reviews further consolidate these developments, highlighting progress and remaining challenges in data-driven ETD and smart grid anomaly detection [12, 13].

Despite their strong performance, SML-based approaches encounter several persistent challenges, including missing or incomplete data, class imbalance, limited scalability, and lack of interpretability. These limitations hinder real-world deployment, particularly in heterogeneous and dynamic smart grid environments. To overcome these barriers, recent studies have explored hybrid ensemble frameworks [14, 15, 16, 17] and frontier architectures such as graph neural networks and transformers to improve feature representation, scalability, and generalization across diverse consumers [18, 19]. Addressing these issues requires advanced feature engineering, standardized evaluation protocols, and scalable, interpretable architectures suitable for real-time grid applications. Parallel advances also emphasize privacy-preserving and benchmark-oriented approaches, including standardized evaluation datasets and federated learning-based anomaly detection for energy systems [20, 21].

Given these emerging trends and persistent challenges, the present study systematically reviews 50 high-quality articles published between 2012 and 2024, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The review synthesizes the evolution of supervised ML techniques for ETD, identifies research gaps, and highlights future directions for developing scalable, interpretable, and field-deployable SML frameworks tailored to modern smart grid infrastructures.

2.0 LITERATURE REVIEW

Building upon the scope outlined in the introduction, this section reviews the application of SML techniques for ETD. It systematically examines four core themes: algorithmic progress, dataset challenges (missing values, class imbalance, and feature engineering), performance benchmarking, and scalability in deployment. By synthesizing current findings, the review identifies limitations, highlights recent advancements, and outlines research directions to enhance model robustness and real-world applicability.

2.1 Progress in Machine Learning Algorithms

SML algorithms have significantly advanced ETD by uncovering fraudulent consumption patterns in large-scale smart meter data. Core models such as extreme gradient boosting (XGBoost), random forest (RF), support vector machines (SVM), and k-nearest neighbors (KNN) serve as foundational classifiers [22, 23]. XGBoost manages high-dimensional data through regularization, while RF aggregates decision trees to yield stable predictions [24, 25]. The SVM performs well for binary classification but struggles with class imbalance. In contrast, KNN is simple and effective for smaller datasets but remains sensitive to noise and feature scaling [26, 27].

Despite their predictive strength, these models face scalability and computational constraints. XGBoost [28, 29, 30, 31] and RF [32, 33, 34] require significant resources for real-time use, whereas SVM and KNN degrade with noisy or skewed data [35, 36, 37]. Decision Trees (DT) are interpretable but prone to overfitting [38, 39, 40]. Deep learning models, such as artificial and convolutional neural networks (ANNs, CNNs), capture complex nonlinear patterns but depend on large labeled datasets [41, 42, 43, 44]. Ensemble methods, including bagging, boosting, and stacking, enhance accuracy by integrating multiple learners [45, 46, 47, 48]. However, they can be resource-intensive and vulnerable to noise. Probabilistic models such as Naïve Bayes offer efficiency but falter when features are correlated [49]. These trade-offs motivate the design of hybrid, scalable frameworks that address data quality and imbalance while maintaining interpretability [28, 29, 50]. Recent studies also explore graph- and transformer-based architectures for richer spatial-temporal learning [18, 19]. Table 1 summarizes the strengths and limitations of key SML algorithms.

Table 1 Summary of SML methods in ETD

Method	Strengths	Limitations
Decision Trees (DT)	Simple; interpretable	Overfits on complex data
Random Forest (RF)	Robust; high accuracy	Computationally intensive
XGBoost	High accuracy; regularized	Long training time
SVM	Effective for binary tasks	Weak on imbalanced data
KNN	Simple; works on small data	Sensitive to Noise and scaling
Deep Learning	Captures nonlinear patterns	Requires large labeled datasets
Ensemble methods	High accuracy; robust	Computationally expensive
Naïve Bayes	Fast; efficient	Fails with correlated features
Hybrid/Optimized Frameworks	High scalability and accuracy	Complex tuning; limited validation

2.2 Dataset Challenges

The effectiveness of SML models for ETD depends on the quality of input data, which commonly suffers from three major issues: missing values, class imbalance, and inadequate feature representation. Tackling these challenges is crucial for accurate preprocessing and reliable model performance.

2.2.1 *Addressing Missing Data*

Missing data—arising from unstable communication links, sensor faults, or maintenance outages—can bias estimates and degrade model accuracy [51]. Traditional remedies include simple imputations and deletions [52]. Mean, median, or mode replacement is fast but distorts distributions when outliers are present [53, 54, 55, 56]. Linear interpolation preserves temporal continuity but fails for irregularly sampled or multidimensional data [57, 58]. Time-series approaches such as autoregressive integrated moving average (ARIMA) predict short gaps but are sensitive to long interruptions [59], whereas historical averages tend to underestimate variance [60].

Advanced imputation methods such as regression models [61, 62], Gaussian mixture models (GMMs) [63, 64], SuperMICE [65], and random forest imputation (RFI) [66, 67] better capture feature dependencies but are computationally demanding. Table 2 compares these methods and highlights the need for hybrid, scalable imputation strategies suitable for large-scale smart grid applications.

Table 2 Comparison of imputation methods for handling missing data

Method	Limitations
Mean/Median Imputation	Produces skewed results in the presence of outliers
Linear Interpolation	Poor performance with high-dimensional data
ARIMA	Sensitive to outliers and large data gaps
Historical Average Modeling	Underestimates variance; yields unrealistic imputations
Data Deletion	Causes loss of critical information
Regression Models	Ignore complex feature interactions.
GMM-based	Struggle with outliers
RF Imputation	Computationally expensive for large datasets

2.2.2 *Handling Data Imbalance*

Class imbalance, where fraudulent consumption records form only a small fraction of total data, biases models toward the majority (legitimate) class [68]. To address this, researchers apply sampling strategies such as oversampling, undersampling, and hybrid techniques [69]. Oversampling techniques, such as the synthetic minority oversampling technique (SMOTE) and adaptive synthetic sampling

(ADASYN), address class imbalance by generating synthetic minority samples, although they may introduce noise and overfitting [70, 71]. Conversely, undersampling methods such as random undersampling (RUS) improve class balance but risk discarding valuable data [72, 73]. Hybrid strategies, including RUSBoost and random hybrid sampling (RHSBoost), combine resampling with boosting to enhance minority detection, though they remain computationally intensive and require further real-world validation [29, 74]. Table 3 summarizes major imbalance-handling methods and their limitations.

2.2.3 *Feature Engineering*

Effective feature engineering is crucial for extracting meaningful consumption patterns. Raw smart meter data are often high-dimensional and noisy, limiting model efficiency [75]. Statistical descriptors (mean, variance, skewness) capture aggregate behavior but ignore temporal dynamics [9]. Dimensionality reduction techniques such as principal component analysis (PCA) and Kernel PCA improve interpretability but may discard critical information in the presence of outliers [76, 77]. Advanced deep representation methods—such as CNN, long short-term memory (LSTM), and autoencoders—effectively capture nonlinear temporal structures, but they also increase computational cost and the risk of overfitting [16, 28, 29]. Table 4 summarizes the main feature-engineering methods and their limitations.

2.2.4 *Comparative Summary*

The reliability of SML-based ETD frameworks depends on systematically addressing missing data, class imbalance, and feature redundancy. Table 5 synthesizes recent studies, underscoring overlapping strategies and remaining gaps. The findings reinforce the need for hybrid, scalable data-repair frameworks that unify advanced imputation, balanced resampling, and feature-reduction techniques to enhance detection accuracy, interpretability, and real-world generalizability.

2.3 Performance Metrics and Benchmarking

Evaluating ETD models requires a range of performance metrics to assess accuracy, reliability, and discriminative capability. Common measures include accuracy, precision, recall, F1-score, and AUC-ROC. While accuracy reflects overall success, it can be misleading for imbalanced datasets dominated by genuine consumers. Precision measures the proportion of predicted thefts that are truly fraudulent, recall quantifies the proportion of actual thefts correctly identified, and their harmonic mean—the F1-score—balances both aspects [36]. The AUC-ROC summarizes the model's discrimination power across thresholds.

Relying solely on accuracy, as in Selvam *et al.* [39], may lead to an overestimation of performance. Complementary indicators such as detection rate and loss probability, used by Faria *et al.* [78], provide broader insight but remain inconsistent across studies. Standardized benchmark datasets, such as TDD2022, have improved reproducibility—exemplified by the tuned BiLSTM model of Krishnamoorthy and Albert [79]. Nevertheless, inconsistent evaluation criteria and limited real-world validation still hinder cross-study comparability. Future research should prioritize benchmark datasets and standardized performance metrics reflecting deployment conditions.

Table 3 Methods for addressing data imbalance

Method	Limitations
SMOTE	Risk of overfitting; noisy synthetic data
ADASYN	Sensitivity to outliers; overlapping minority samples
RUS	Loss of informative samples; potential bias
RUSBoost	Computational complexity, possible underfitting
RHSBoost	High computational cost; information loss
Ensemble Methods	Resource-intensive; minority classes may be underrepresented

Table 4 Summary of feature engineering techniques and their limitations

Method	Limitation
Statistical Features	Limited to simple aggregate patterns
PCA/Kernel PCA	Sensitive to outliers; possible data loss
CNN, LSTM, Autoencoders	Computationally demanding; prone to overfitting

Table 5 Checklist of SML techniques addressing major data challenges

Ref	Method	Missing Values	Data Imbalance	Feature Extraction	Feature Selection
[8]	Ensemble ML	✓	✓	✗	✗
[10]	Boosted DT	✗	✗	✓	✓
[17]	Ensemble ML	✗	✗	✗	✗
[30]	FA-XG Boost	✓	✓	✓	✗
[31]	FE-XG Boost	✗	✓	✓	✗
[32]	RF	✗	✓	✗	✗
[33]	Autoencoder	✗	✓	✓	✗
[34]	CNN-RF	✓	✓	✓	✗
[36]	SVM-DT	✗	✗	✓	✓
[39]	DT, RF	✗	✗	✗	✗
[41]	DNN, DT	✓	✗	✗	✗
[43]	Wide-CNN	✓	✗	✓	✗

Ref	Method	Missing Values	Data Imbalance	Feature Extraction	Feature Selection
[44]	Neuro-Fuzzy	✗	✗	✓	✓
[48]	Ensemble ML	✓	✓	✓	✗
[78]	STE	✗	✗	✓	✗
[37]	CPD	✗	✓	✓	✗
[82]	RNN	✓	✗	✗	✗
[83]	CAT Boost	✓	✓	✓	✓
[84]	NG Boost	✓	✓	✗	✓

2.4 Scalability and Deployment Challenges

The scalability and deployment of SML models for ETD remain major obstacles. In many developing regions, inadequate infrastructure restricts efficient processing of high-resolution smart meter data, while the heavy computational demands of deep learning models raise operational costs and reduce interpretability.

Managing heterogeneous consumption patterns and massive data volumes requires extensive preprocessing, and batch-trained models often fail to meet real-time detection needs.

Proposed solutions include edge computing to decentralize computation [23, 29], cloud-based pipelines for scalable storage and parallel processing [33], and algorithmic optimization to balance detection accuracy with processing speed. Hybrid and adaptive methods, such as transfer learning, enhance scalability by reducing computational overhead while enabling continuous adaptation to evolving consumption behaviors.

Emerging graph and transformer-based frameworks further enhance real-time adaptability through edge–cloud collaboration but still require lightweight optimization to reduce latency and energy usage [18, 19, 48]. Deployment also faces challenges related to data governance, interoperability, and latency, as reported in recent utility case studies [80, 81, 82]. Consequently, future research should emphasize resource-efficient architectures, standardized deployment workflows, and cross-platform reproducibility to ensure scalable and practical ETD solutions [83, 84].

3.0 METHODOLOGY

This study employs a systematic literature review (SLR) methodology to rigorously identify, evaluate, and synthesize research on SML techniques for ETD. The review follows the preferred reporting items for systematic reviews and meta-analyses (PRISMA 2020) guidelines [85], while integrating quality-assessment principles from Kitchenham's (2015) framework [86], to ensure transparency, reproducibility, and methodological rigor. The review is structured around four key dimensions addressing algorithmic diversity, data challenges, evaluation metrics, and scalability considerations. The SLR methodology is organized into four sequential stages: planning, literature selection, data extraction and thematic synthesis, and quality assessment. Figure 1 illustrates the overall workflow, and the subsequent subsections detail the search strategy, inclusion and exclusion criteria, and thematic mapping aligning with research objectives.

3.1 Literature Search and Screening

3.1.1 Search Strategy

Keywords were carefully formulated to cover both the domain and methodological scope. Primary terms included "electricity theft detection," "energy fraud detection," "non-technical losses," and "power theft." Secondary terms captured algorithmic aspects such as "machine learning," "deep learning," "artificial intelligence," and the names of specific models (e.g., decision tree, random forest, support vector machine, and logistic regression). Boolean operators and database-specific syntax were used to enhance precision and comprehensiveness.

Figure 1 Overview of the systematic literature review (SLR) process showing planning, search, screening, synthesis, and quality-assessment stages

The search was conducted across Scopus, Web of Science, and IEEE Xplore and was limited to peer-reviewed studies published between January 2012 and April 2024. This period reflects the evolution of SML techniques following the widespread deployment of AMI in modern smart grid systems.

3.1.2 Screening Process and Criteria

The initial search retrieved 1,712 articles: 1,059 from Scopus, 437 from Web of Science, and 216 from IEEE

Xplore. After removing duplicates, 622 unique articles underwent a three-stage screening process. Title screening reduced this number to 437, abstract screening narrowed it to 123, and a full-text review identified 50 high-quality articles for detailed analysis. The PRISMA flow diagram shown in Figure 2 outlines this process, while Table 6 summarizes the inclusion and exclusion criteria applied during article selection.

Table 6 Selection criteria for article inclusion and exclusion

Criteria Type	Criteria
Inclusion	Studies on SML techniques for ETD; quantitative evaluation metrics (e.g., accuracy, F1-score); studies addressing challenges such as scalability, class imbalance, or missing data; English-language peer-reviewed articles published 2012–2024.
Exclusion	Solely hardware-focused studies; qualitative or purely theoretical papers lacking algorithmic details; non-peer-reviewed sources; duplicates; publications before 2012.

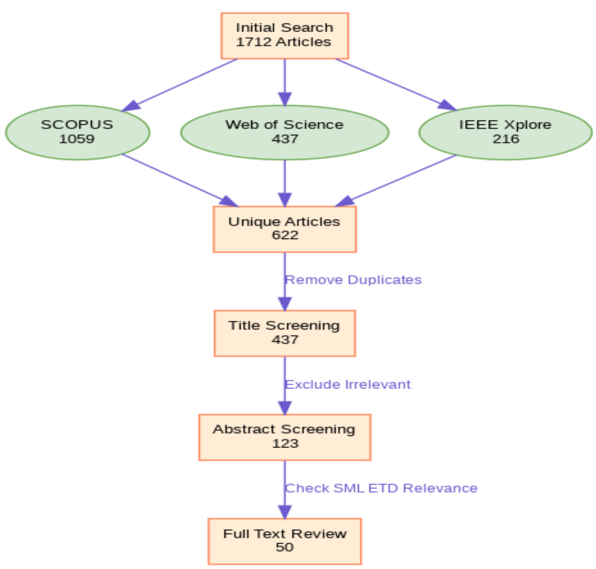

Figure 2 PRISMA flow diagram illustrating identification, screening, and inclusion stages

3.2 Data Extraction and Thematic Synthesis

The extracted data were categorized into four themes to provide a systematic analysis of the challenges and methodologies in ETD. The first theme explores diverse machine learning approaches, including decision trees, random forests, gradient boosting, and deep learning architectures such as convolutional and recurrent neural networks. These methods enhance detection accuracy but often reduce interpretability. The second theme examines datasets, with a focus on missing data, class imbalance, and the need for effective preprocessing strategies. The third theme highlights evaluation metrics—accuracy, recall, precision, F1-score, and ROC-AUC—emphasizing the need for standardized reporting. The fourth theme addresses scalability and deployment, analyzing how models can be adapted for large-scale grids and real-time applications. This synthesis provides a comprehensive overview of ETD research, identifying key advancements and persistent gaps in scalability, interpretability, and data quality.

3.3 Quality Assessment

The quality-assessment framework evaluated each study using four predefined criteria: (C1) relevance to ETD, (C2) clarity of methodology, (C3) robustness of evaluation metrics, and (C4) real-world applicability. Each criterion was scored on a scale from 0 to 3 (Table 7), yielding a maximum of 12 points per study. Studies scoring 7 or higher were classified as high-quality, reflecting both methodological rigor and practical significance. This integrated thematic synthesis and quality-assessment process provided a robust foundation for analyzing research trends and assessing the overall maturity of SML-based ETD studies.

Table 7 Evaluation criteria for ETD studies

Criterion	3	2	1	0
C1	Direct relevance	Indirect relevance	Marginal relevance	Irrelevant
C2	Clear methodology	Moderate clarity	Vague methods	Unclear methods
C3	Comprehensive metrics	Limited metrics	Weak metrics	No metrics
C4	Real-world validation	Potential applications	Minimal relevance	Lack of relevance

4.0 RESULTS AND DISCUSSION

This section presents the findings of the systematic review, focusing on the thematic distribution, quality assessment, and key trends in ETD using SML techniques.

4.1 Thematic Distribution of Studies

The review categorized 50 studies into four themes: SML algorithms, dataset challenges, performance metrics, and scalability and deployment. These themes represent the core methodologies and challenges in ETD research. Table 8 summarizes the thematic distribution and links study counts to each theme.

Table 8 Thematic distribution of studies

Theme	Count (Out of 50)	% of Studies
Machine Learning Algorithms	33	66%
Datasets and Challenges	24	48%
Performance Metrics	23	46%
Scalability and Deployment	14	28%

The analysis reveals that 66% of studies focus on SML algorithms, highlighting advancements in models such as decision trees, random forests, and deep learning. About 48% of studies address datasets and challenges, emphasizing preprocessing for missing data and class imbalance. Only 28% examine scalability and deployment, indicating gaps in real-time and large-scale implementations, and 46% investigate performance metrics, focusing on evaluation criteria such as precision, recall, and AUC-ROC.

The heatmap (Figure 3) visualizes thematic overlaps, highlighting intersections between SML algorithms and performance metrics (reflecting a strong emphasis on algorithm evaluation). Similarly, overlaps between dataset challenges and scalability underscore efforts to address preprocessing issues for robust, scalable applications. These findings emphasize the importance of integrative research approaches for advancing ETD solutions.

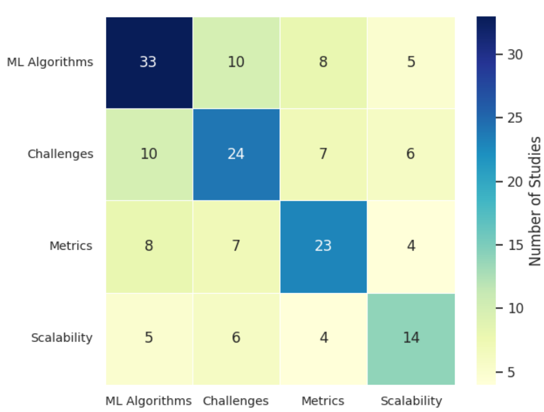

Figure 3 Heatmap of thematic distribution and overlaps

4.2 Quality Assessment Results

The quality assessment analyzed 50 selected studies based on four criteria: relevance to ETD (C1), methodological clarity (C2), robustness of evaluation metrics (C3), and real-world applicability (C4). Each study received a score from 0 to 3 per criterion (maximum 12). Studies scoring 7 or higher were classified as high-quality contributions.

As summarized in Table 9, relevance (C1) and methodological clarity (C2) emerged as strengths, with average scores of 2.66 and 2.68, respectively. A substantial share (72% for C1; 70% for C2) achieved the highest score, reflecting alignment with ETD objectives and sound methodology. In contrast, evaluation metrics (C3) and real-world applicability (C4) were weaker, with averages of 2.34 and 2.32. Over 54% of studies scored ≤2 for C3, highlighting inconsistent benchmarking, while 66% scored ≤2 for C4, indicating gaps in scalability and real-world deployments.

Figures 4 and 5 provide a visual summary of these findings. Figure 4 highlights strong performance in C1 and C2, while Figure 5 shows that most studies cluster near the threshold score of 7, underscoring the need for improved evaluation frameworks and practical validations.

Table 9 Quality assessment results

Evaluation Metric	Average Score	Score = 3	Score ≤ 2
C1	2.66	36	14
C2	2.68	35	15
C3	2.34	23	27
C4	2.32	17	33

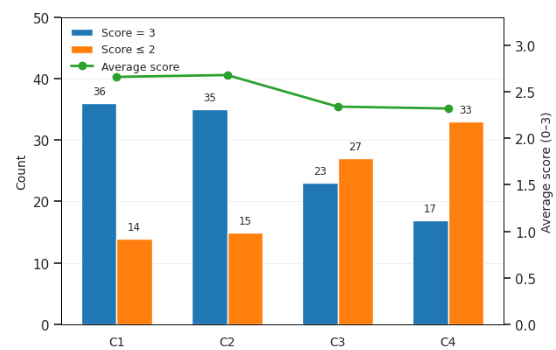

Figure 4 Evaluation summary across criteria

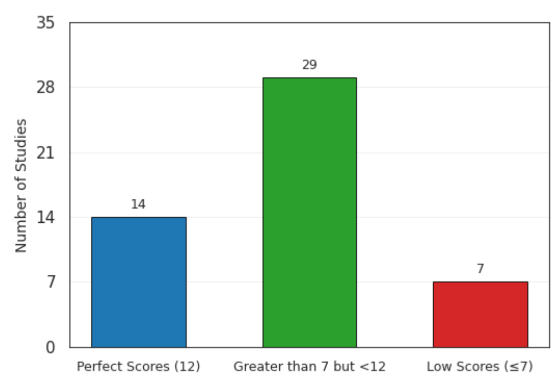

Figure 5 Distribution of overall quality scores across studies

4.3 Discussion

The results show substantial progress in ETD using SML, alongside persistent challenges. Most studies demonstrate strong relevance (C1) and methodological clarity (C2), with over 70% achieving the top score, which indicates close alignment with ETD objectives and replicable methods.

Despite these strengths, gaps in evaluation metrics (C3) and real-world applicability (C4) persist, with 54% and 66% of studies scoring ≤2, respectively. These shortcomings hinder cross-study comparability and impede scalable deployment. For example, although 66% of studies emphasize algorithmic advances—such as decision trees, random forests, and deep learning—only 28% address scalability, which is essential for real-world implementation.

Key trends highlight the importance of hybrid models that integrate complementary algorithms to improve robustness and interpretability. Addressing data issues—such as missing values and class imbalance—through advanced preprocessing is also vital for building scalable and reliable ETD systems. Future research must also prioritize standardizing evaluation frameworks (e.g., F1-score, ROC-AUC) to facilitate meaningful comparisons and validate solutions in diverse, real-world environments.

5.0 CONCLUSION

This systematic literature review analyzed 50 studies on ETD using SML techniques, providing a structured evaluation of the research landscape. The review identified significant advancements in machine learning methods, particularly decision trees, boosting algorithms, and deep learning architectures, which highlight their effectiveness in detecting fraudulent consumption patterns.

However, persistent challenges remain. Dataset limitations, inconsistent evaluation metrics, and scalability issues continue to hinder the practical deployment of ETD systems. The quality assessment revealed strong methodological clarity and alignment with ETD objectives, but real-world applicability emerged as a critical gap, with limited validation of solutions in large-scale grid environments.

Addressing these challenges requires a focus on developing hybrid models that balance accuracy and interpretability, standardizing performance metrics to enhance comparability, and validating models on diverse datasets for scalability. These efforts are essential to ensuring robust and scalable ETD solutions for dynamic smart grid environments.

Acknowledgment

The authors gratefully acknowledge the academic and research community for their valuable contributions, which form the foundation of this review.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] Kumar, P., Lin, Y., Bai, G., Paverd, A., Dong, J. S., & Martin, A. 2019. Smart Grid Metering Networks: A Survey on Security, Privacy and Open Research Issues. *IEEE Communications Surveys & Tutorials*. 21(3): 2886–2927. http://dx.doi.org/10.1109/comst.2019.2899354.

[2] Liu, Y., Yu, Y., Gao, N., & Wu, F. 2020. A Grid as Smart as the Internet. *Engineering*. 6(7): 778-788. https://doi.org/10.1016/j.eng.2019.11.015.

[3] Kim, S., Sun, Y., Lee, S., Seon, J., Hwang, B., Kim, J., Kim, J., Kim, K., & Kim, J. 2024. Data-Driven Approaches for Energy Theft Detection: A Comprehensive Review. *Energies*. 17(12): 3057. https://doi.org/10.3390/en17123057.

[4] Butt, O. M., Zulqarnain, M., & Butt, T. M. 2021. Recent Advancement in Smart Grid Technology: Future Prospects in the Electrical Power Network. *Ain Shams Engineering Journal*. 12(1): 687–695. http://dx.doi.org/10.1016/j.asej.2020.05.004.

[5] Guerrero, J. I., Monedero, I., Biscarri, F., Biscarri, J., Millan, R., & Leon, C. 2018. Non-Technical Losses Reduction by Improving the Inspections Accuracy in a Power Utility. *IEEE Transactions on Power Systems*. 33(2): 1209–1218. http://dx.doi.org/10.1109/TPWRS.2017.2721435

[6] Dineshkumar, K., Ramanathan, P., Ramasamy, S. 2015. Development of ARM Processor-based Electricity Theft Control System using GSM Network. *Proceedings of the IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. 1–6. https://doi.org/10.1109/ICECCT.2015.7226135.

[7] Kang, C., Wang, Y., Xue, Y., Mu, G., & Liao, R. 2018. Big Data Analytics in China's Electric Power Industry: Modern Information, Communication Technologies, and Millions of Smart Meters. *IEEE Power and Energy Magazine*. 16(3): 54–65. http://dx.doi.org/10.1109/mpe.2018.2790819.

[8] Gunturi, S. K., & Sarkar, D. 2020. Ensemble Machine Learning Models for Energy Theft Detection. *Electric Power Systems Research*. 180: 106904. http://dx.doi.org/10.1016/j.epsr.2020.106904.

[9] Punmiya, R., & Choe, S. 2019. Energy Theft Detection Using Gradient Boosting Theft Detector with Feature Engineering-Based Preprocessing. *IEEE Transactions on Smart Grid*. 10(2): 2326–2334. http://dx.doi.org/10.1109/tsg.2019.2892595.

[10] Saeed, M. S., Mustafa, M. W., Sheikh, U. U., Jumani, T. A., Khan, I., Atawneh, S., & Hamadneh, N. N. 2020. An Efficient Boosted C5.0 Decision-Tree-Based Classification Approach for Detecting Non-Technical Losses. *Energies.* 13(12): 3242. http://dx.doi.org/10.3390/en13123242.

[11] Ullah, A., Javaid, N., Shoaib, M., & Sultana, T. 2021. Hybrid Deep Neural Network for Electricity Theft Detection Using Intelligent Antenna-Based Smart Meters. *Wireless Communications and Mobile Computing*. http://dx.doi.org/10.1155/2021/6643901.

[12] Stracqualursi, E., Rosato, A., Di Lorenzo, G., Panella, M., & Araneo, R. 2023. Systematic Review of Energy Theft Practices and Autonomous Detection through Artificial Intelligence Methods. *Renewable and Sustainable Energy Reviews*. 184: 113544.

http://dx.doi.org/10.1016/j.rser.2023.113544.

[13] Guato Burgos, M. F., Morato, J., & Vizcaino Imacaña, F. P. 2024. A Review of Smart Grid Anomaly Detection Approaches Pertaining to Artificial Intelligence. *Applied Sciences*. 14(3): 1194. http://dx.doi.org/10.3390/app14031194.

[14] Shehzad, F., Javaid, N., Almogren, A., Ahmed, A., Gulfam, S. M., Radwan, A. 2021. A Robust Hybrid Deep Learning Model for Detection of Non-technical Losses to Secure Smart Grids. *IEEE Access*. 9: 128663–128678. https://doi.org/10.1109/ACCESS.2021.3113592.

[15] Hashim, M., Khan, L., Javaid, N., Ullah, Z., Javed, A. 2024. Stacked Machine Learning Models for Non-technical Loss Detection in Smart Grid: A Comparative Analysis. *Energy Reports*. 12: 1235–1253. https://doi.org/10.1016/j.egyr.2024.06.015.

[16] Adil, M., Javaid, N., Ullah, A., Shafiq, M., & Choi, J.-G. 2020. LSTM and Bat-Based RUS Boost Approach for Electricity Theft Detection. *Applied Sciences*. 10(12): 4378. http://dx.doi.org/10.3390/app10124378.

[17] Javaid, N., Gul, H., Baig, S., Shehzad, F., Xia, C., Guan, L., & Sultana, T. 2021. Using GANCNN and ERNET for Detection of Non-Technical Losses to Secure Smart Grids. *IEEE Access*: 9: 98679-98700. http://dx.doi.org/10.1109/access.2021.3099289.

[18] Xia, W., Li, Y., Yu, L., and He, D. 2024. Locational Detection of False Data Injection Attacks in the Edge Space via Hodge Graph Neural Network for Smart Grids. *IEEE Transactions on Smart Grid.* 15(5): 5102–5114. https://doi.org/10.1109/TSG.2024.3389948.

[19] Saeed, F., and Aldera, S. 2024. Adaptive Renewable Energy Forecasting Utilizing a Data-Driven PCA–Transformer Architecture. *IEEE Access*. 12: 109269–109280. http://dx.doi.org/10.1109/ACCESS.2024.3440226.

[20] Zidi, S., Mihoub, A., Qaisar, S. M., Krichen, M., & Abu Al-Haija, Q. 2023. Theft Detection Dataset for Benchmarking and Machine Learning-Based Classification in a Smart Grid Environment. *Journal of King Saud University – Computer and Information Sciences*. 35(1): 13–25. http://dx.doi.org/10.1016/j.jksuci.2022.05.007.

[21] Alshehri, A., Badr, M. M., Baza, M., & Alshahrani, H. 2024. Deep Anomaly Detection Framework Utilizing Federated Learning for Electricity Theft Zero-Day Cyberattacks. *Sensors*. 24(10): 3236. http://dx.doi.org/10.3390/s24103236.

[22] Hastie, T., Tibshirani, R., & Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

[23] Hussain, S., Mustafa, M. W., Al-Shqeerat, K. H. A., Al-Rimy, B. A. S., & Saeed, F. 2022. Electric Theft Detection in Advanced Metering Infrastructure Using Jaya Optimized Combined Kernel-Tree Boosting Classifier—A Novel Sequentially Executed Supervised Machine Learning Approach. *IET Generation, Transmission & Distribution*. 16(6): 1257–1275. https://doi.org/10.1049/gtd2.12386.

[24] Nawaz, A., Ali, T., Mustafa, G., Rehman, S. U., & Rashid, M. R. 2023. A Novel Technique for Detecting Electricity Theft in Secure Smart Grids Using CNN and XG-Boost. *Intelligent Systems with Applications*. 17: 200168. http://dx.doi.org/10.1016/j.iswa.2022.200168.

[25] Iranzad, R., & Liu, X. 2024. A Review of Random Forest-Based Feature Selection Methods for Data Science Education and Applications. *International Journal of Data Science and Analytics*. 8(1): 15–32. https://doi.org/10.1007/s41060-024-00509-w.

[26] Abro, S. A., Hua, L. G., Laghari, J. A., Bhayo, M. A., & Memon, A. A. 2024. Machine Learning-Based Electricity Theft Detection Using Support Vector Machines. *International Journal of Electrical and Computer Engineering*. 14(2): 1240–1250. http://doi.org/10.11591/ijece.v14i2.pp1240-1250.

[27] Peterson, L. E. 2009. K-nearest neighbor. *Scholarpedia*. 4(2): 1883. http://dx.doi.org/10.4249/scholarpedia.1883.

[28] Iftikhar, H., Khan, N., Raza, M. A., Abbas, G., Khan, M., Aoudia, M., Touti, E., & Emara, A. 2024. Electricity Theft Detection in Smart Grid using Machine Learning. *Frontiers in Energy Research*. 12: Article 1383090. https://doi.org/10.3389/fenrg.2024.1383090.

[29] Mujeeb, S., Javaid, N., Ahmed, A., Gulfam, S. M., Qasim, U., Shafiq, M., & Choi, J.-G. 2021. Electricity Theft Detection with Automatic Labeling and Enhanced RUS Boost Classification Using Differential Evolution and Jaya Algorithm. *IEEE Access*. http://dx.doi.org/10.1109/ACCESS.2021.3102643.

[30] Khan, Z. A., Adil, M., Javaid, N., Saqib, M. N., Shafiq, M., & Choi, J.-G. 2020. Electricity Theft Detection Using Supervised Learning Techniques on Smart Meter Data. *Sustainability*. 12(19): 8023. https://doi.org/10.3390/su12198023.

[31] Cheng, G., Zhang, Z., Li, Q., Li, Y., and Jin, W. 2021. Energy Theft Detection in an Edge Data Center Using Deep Learning. *Mathematical Problems in Engineering*. 2021. https://doi.org/10.1155/2021/5591769.

[32] Qu, Z., Li, H., Wang, Y., Zhang, J., Abu-Siada, A., & Yao, Y. 2020. Detection of Electricity Theft Behavior based on Improved Synthetic Minority Oversampling Technique and Random Forest Classifier. *Energies*. 13(8): 2039. http://dx.doi.org/10.3390/en13082039.

[33] Lin, G., Feng, X., Guo, W., Cui, X., Liu, S., & Jin, W. 2021. Electricity Theft Detection Based on Stacked Autoencoder and the Undersampling and Resampling Based Random Forest Algorithm. *IEEE Access*. 9: 124044–124058. http://dx.doi.org/10.1109/ACCESS.2021.3110510

[34] Li, S., Han, Y., Yao, X., Yingchen, S., Wang, J., & Zhao, Q. 2019. Electricity Theft Detection in Power Grids with Deep Learning and Random Forests. *Journal of Electrical and Computer Engineering*. 2019: Article ID 4136874. https://doi.org/10.1155/2019/4136874.

[35] Nagi, J., Yap, K. S., Tiong, S. K., Ahmed, S. K., & Mohamad, M. 2009. Nontechnical Loss Detection for Metered Customers in Power Utility using Support Vector Machines. *IEEE Transactions on Power Delivery*. 25(2): 1162–1171. http://dx.doi.org/10.1109/TPWRD.2009.2036519.

[36] Kosut, J. P., Santomauro, F., Jorysz, A., Fernández, A., Lecumberry, F., and Rodríguez, F. 2015. Abnormal Consumption Analysis for Fraud Detection: UTE–UDELAR joint Efforts. *In Proceedings of the 2015 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT LATAM)*, Montevideo, Uruguay. 887–892. https://doi.org/10.1109/ISGT-LA.2015.7381272.

[37] Jokar, P., Arianpoo, N., & Leung, V. C. M. 2016. Electricity Theft Detection in AMI using Customers' Consumption Patterns. *IEEE Transactions on Smart Grid*. 7(1): 216–226. http://dx.doi.org/10.1109/TSG.2015.2425222.

[38] Jindal, A., Dua, A., Kaur, K., Singh, M., Kumar, N., & Mishra, S. 2016. Decision Tree and SVM-based Data Analytics for Theft Detection in Smart Grid. *IEEE Transactions on Industrial Informatics*. 12(3): 1005–1016. http://dx.doi.org/10.1109/TII.2016.2543146.

[39] Nallathambi, S., and Ramasamy, K. 2017. Prediction of Electricity Consumption based on DT and RF: An Application on USA Country Power Consumption. In *Proceedings of the 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE)*, Karur, India. 1–7. https://doi.org/10.1109/ICEICE.2017.8191939.

[40] Kong, X., Zhao, X., Liu, C., Li, Q., Dong, D., & Li, Y. 2021. Electricity Theft Detection in Low-voltage Stations based on Similarity Measure and DT-KSVM. *International Journal of Electrical Power & Energy Systems*. 125: 106544. http://dx.doi.org/10.1016/j.ijepes.2020.106544.

[41] Bohani, F. A., Suliman, A., Saripuddin, M., Sameon, S. S., Md Salleh, N. S., & Nazeri, S. 2021. A Comprehensive Analysis of Supervised Learning Techniques for Electricity Theft Detection. *Journal of Electrical and Computer Engineering.* 2021: Article ID 2090-0147.

http://dx.doi.org/10.1155/2021/2090147.
[42] Fei, K., Li, Q., & Zhu, C. 2022. Non-technical losses Detection using Missing Values' Pattern and Neural Architecture Search. *International Journal of Electrical Power & Energy Systems*. 134: 107410. http://dx.doi.org/10.1016/j.ijepes.2021.107410.
[43] Zheng, Z., Yang, Y., Niu, X., Dai, H.-N., & Zhou, Y. 2018. Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids. *IEEE Transactions on Industrial Informatics*. 14(4): 1606–1615. http://dx.doi.org/10.1109/TII.2017.2785963.
[44] Blazakis, K. V., Kapetanakis, T. N., & Stavrakakis, G. S. 2020. Effective Electricity Theft Detection in Power Distribution Grids Using an Adaptive Neuro Fuzzy Inference System. *Energies*. 13(12): 3110. http://dx.doi.org/10.3390/en13123110.
[45] Barros, R. M. R., da Costa, E. G., & Araujo, J. F. 2021. Evaluation of Classifiers for Non-technical Loss Identification in Electric Power Systems. *International Journal of Electrical Power & Energy Systems*. 132: 107173. http://dx.doi.org/10.1016/j.ijepes.2021.107173.
[46] Aslam, Z., Javaid, N., Ahmad, A., Ahmed, A., & Gulfam, S. M. 2020. A Combined Deep Learning and Ensemble Learning Methodology to Avoid Electricity Theft in Smart Grids. *Energies*. 13(21): 5599. http://dx.doi.org/10.3390/en13215599.
[47] Saeed, M. S., Mustafa, M. W., Sheikh, U. U., Jumani, T. A., & Mirjat, N. H. 2019. Ensemble Bagged Tree-based Classification for Reducing Non-technical Losses in Multan Electric Power Company of Pakistan. *Electronics*. 8(8): 860. http://dx.doi.org/10.3390/electronics8080860.
[48] Arif, A., Javaid, N., Aldegheishem, A., & Alrajeh, N. 2021. Big Data Analytics for Identifying Electricity Theft using Machine Learning Approaches in Microgrids for Smart Communities. *Concurrency and Computation: Practice and Experience*. e6316. http://dx.doi.org/10.1002/cpe.6316.
[49] Nizar, A. H., Dong, Z. Y., Zhao, J. H., & Zhang, P. 2007. A Data Mining-based NTL Analysis Method. *IEEE*. 1–8. http://dx.doi.org/10.1109/PES.2007.385630.
[50] Rahaman, M. A., and Idris, R. M. 2024. Enhancing Energy Fraud Detection with Ensemble Learning Techniques. In *Proceedings of the IEEE International Conference on Advanced Power Engineering and Energy (APEE),* Johor Bahru, Malaysia. 159–164. https://doi.org/10.1109/APEE60256.2024.10790893.
[51] Elhassan, A., Abu-Soud, S. M., Alghanim, F., and Salameh, W. 2022. ILA4: Overcoming Missing Values in Machine Learning Datasets—An Inductive Learning Approach. *Journal of King Saud University – Computer and Information Sciences*. 34(7): 4284–4295. https://doi.org/10.1016/j.jksuci.2021.02.011.
[52] Garciarena, U., and Santana, R. 2017. An extensive Analysis of the Interaction between Missing Data Types, Imputation Methods, and Supervised Classifiers. *Expert Systems with Applications.* 89: 52–65. https://doi.org/10.1016/j.eswa.2017.07.026.
[53] Zhang, Y., Zhou, B., Cai, X., Guo, W., Ding, X., and Yuan, X. 2021. Missing Value Imputation in Multivariate Time Series with End-to-end Generative Adversarial Networks. *Information Sciences.* 551: 67–82. https://doi.org/10.1016/j.ins.2020.11.003.
[54] Huang, M. W., Lin, W. C., Chen, C. W., Ke, S. W., Tsai, C. F., and Eberle, W. 2016. Data Preprocessing Issues for Incomplete Medical Datasets. *Expert Systems.* 33(5): 432–438. https://doi.org/10.1111/exsy.12159.
[55] Xia, J., Zhang, S., Cai, G., Li, L., Pan, Q., Yan, J., and Ning, G. 2017. Adjusted Weight Voting Algorithm for Random Forests in Handling Missing Values. *Pattern Recognition.* 69: 52–60. https://doi.org/10.1016/j.patcog.2017.04.011.
[56] Pati, S. K., and Das, A. K. 2017. Missing Value Estimation for Microarray Data through Cluster Analysis. *Knowledge and Information Systems.* 52(3): 709–750. https://doi.org/10.1007/s10115-016-0980-6.
[57] Noor, N. M., Al Bakri Abdullah, M. M., Yahaya, A. S., and Ramli, N. A. 2015. Comparison of Linear Interpolation Method and Mean Method to Replace the Missing Values in Environmental Data Set. *Applied Mechanics and Materials.* 803: 278–281. https://doi.org/10.4028/www.scientific.net/AMM.803.278.
[58] Kornelsen, K., and Coulibaly, P. 2014. Comparison of Interpolation, Statistical, and Data-driven Methods for Imputation of Missing Values in a Distributed Soil Moisture Dataset. *Journal of Hydrologic Engineering.* 19(1): 26–43. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000767
[59] Saputra, M. D., Hadi, A. F., Riski, A., and Anggraeni, D. 2021. Handling Missing Values and Unusual Observations in Statistical Downscaling Using Kalman Filter. *Journal of Physics: Conference Series.* 1863: 012035. https://doi.org/10.1088/1742-6596/1863/1/012035
[60] Li, L., Du, B., Wang, Y., Qin, L., and Tan, H. 2020. Estimation of Missing Values in Heterogeneous Traffic Data: Application of Multimodal Deep Learning Model. *Knowledge-Based Systems.* 194: 105592. https://doi.org/10.1016/j.knosys.2020.105592.
[61] Gao, Y., Foggo, B., and Yu, N. 2019. A Physically Inspired Data-driven Model for Electricity Theft Detection with Smart Meter Data. *IEEE Transactions on Industrial Informatics.* 15(9): 5076–5088. https://doi.org/10.1109/TII.2019.2898171.
[62] Shao, J., Meng, W., and Sun, G. 2017. Evaluation of Missing Value Imputation Methods for Wireless Soil Datasets. *Personal and Ubiquitous Computing.* 21(1): 113–123. https://doi.org/10.1007/s00779-016-0980-3.
[63] Sovilj, E. E., Miche, D., Björk, Y., Nian, K. M., Akusok, R., and Lendasse, A. 2016. Extreme Learning Machine for Missing Data using Multiple Imputations. *Neurocomputing.* 174: 220–231. https://doi.org/10.1016/j.neucom.2015.09.076.
[64] García-Laencina, P. J., Sancho-Gómez, J.-L., and Figueiras-Vidal, A. R. 2013. Classifying patterns with Missing Values using Multi-task Learning Perceptrons. *Expert Systems with Applications.* 40(4): 1333–1341. https://doi.org/10.1016/j.eswa.2012.08.040.
[65] Laqueur, H. S., Shev, A. B., & Kagawa, R. M. C. 2022. SuperMICE: An Ensemble Machine Learning Approach to Multiple Imputation by Chained Equations. *American Journal of Epidemiology*. 191(3): 516–525. http://dx.doi.org/10.1093/aje/kwab271.
[66] Stekhoven, D. J., and Bühlmann, P. 2012. MissForest—non-parametric Missing Value Imputation for Mixed-type Data. *Bioinformatics.* 28(1): 112–118. https://doi.org/10.1093/bioinformatics/btr597.
[67] Tang, F., and Ishwaran, H. 2017. Random Forest Missing Data Algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal.* 10(6): 363–377. https://doi.org/10.1002/sam.11348.
[68] Depuru, S. S. S. R., Wang, L., Devabhaktuni, V., and Green, R. C. 2013. High-performance Computing for Detection of Electricity Theft. *International Journal of Electrical Power & Energy Systems*. 47: 21–30. https://doi.org/10.1016/j.ijepes.2012.10.031
[69] Pereira, J., and Saraiva, F. 2020. A Comparative Analysis of Unbalanced Data Handling Techniques for Machine Learning Algorithms to Electricity Theft Detection. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC),* Glasgow, UK, 19–24 July. 1–8. https://doi.org/10.1109/CEC48606.2020.9185822.
[70] Kumar, S., Sharma, R., & Singh, A. K. 2024. A Novel Hybrid SMOTE-ENN Sampling Framework for Fraud Detection in Power Systems. *Applied Soft Computing*. 155: 111276. http://dx.doi.org/10.1016/j.asoc.2023.111276.

[71] Yao, J., Zhang, T., & Hu, H. 2023. Adaptive Synthetic Oversampling with Feature Weighting for Imbalanced Data Classification. *Knowledge-Based Systems*. 284: 111252. http://dx.doi.org/10.1016/j.knosys.2023.111252.

[72] Nguyen, H. M., Cooper, E. W., and Kamei, K. 2012. A Comparative Study on Sampling Techniques for Handling Class Imbalance in Streaming Data. In *Proceedings of the IEEE ICDMW*. 1762–1767. https://doi.org/10.1109/icdmw.2012.128.

[73] Buzau, M. M., Tejedor-Aguilera, J., Cruz-Romero, P., & Gómez-Expósito, A. 2019. Detection of Non-Technical Losses Using Smart Meter Data and Supervised Learning. *IEEE Transactions on Smart Grid*. 10(3): 2661–2670. https://doi.org/10.1109/tsg.2018.2807925.

[74] Gong, J., and Kim, H. 2017. RHSBoost: Improving Classification Performance in Imbalance Data. *Computational Statistics & Data Analysis*. 111: 1–13. https://doi.org/10.1016/j.csda.2017.02.009.

[75] Razavi, R., Gharipour, A., Fleury, M., and Akpan, I. J. 2019. A Practical Feature-engineering Framework for Electricity Theft Detection in Smart Grids. *Applied Energy*. 238: 481–494. https://doi.org/10.1016/j.apenergy.2019.01.076.

[76] Lee, L. C., and Jemain, A. A. 2021. On overview of PCA Application Strategy in Processing High Dimensionality Forensic Data. *Microchemical Journal*. 106608. https://doi.org/10.1016/j.microc.2021.106608.

[77] Gul, H., Javaid, N., Ullah, I., Qamar, A. M., Afzal, M. K., and Joshi, G. P. 2020. Detection of Non-technical Losses using SOSTLink and Bidirectional Gated Recurrent Unit to Secure Smart Meters. *Applied Sciences*. 10(9): 3151. https://doi.org/10.3390/app10093151

[78] Faria, L. T., Melo, J. D., and Padilha-Feltrin, A. 2015. Spatial-temporal Estimation for Nontechnical Losses. *IEEE Transactions on Power Delivery*. 31(1): 362–369. https://doi.org/10.1109/TPWRD.2015.2430377.

[79] Krishnamoorthy, M., & Albert, J. R. 2024. Electricity Theft Detection in IoT-Based Smart Grids Using a Parameter-Tuned Bidirectional LSTM with Pre-Trained Feature Learning Mechanism. *Electrical Engineering*. 106: 5987–6001. https://doi.org/10.1007/s00202-024-02342-7.

[80] Lepolesa, L. J., Achari, S., & Cheng, L. 2022. Electricity Theft Detection in Smart Grids Based on Deep Neural Network. *IEEE Access*. 10: 39638–39655. http://dx.doi.org/10.1109/ACCESS.2022.3166146.

[81] Tehrani, S. O., Shahrestani, A., & Yaghmaee, M. H. 2022. Online Electricity Theft Detection Framework for Large-Scale Smart Grid Data. *Electric Power Systems Research*. 208: 107895. http://dx.doi.org/10.1016/j.epsr.2022.107895.

[82] Ullah, A., Munawar, S., Asif, M., Kabir, B., and Javaid, N. 2021. Synthetic Theft Attacks Implementation for Data Balancing and a Gated Recurrent Unit based Electricity Theft Detection in Smart Grids. *In Barolli, L., Yim, K., and Enokido, T. (eds.) Complex, Intelligent and Software Intensive Systems (CISIS 2021)*. Lecture Notes in Networks and Systems, vol. 278. Cham: Springer. https://doi.org/10.1007/978-3-030-79725-6_39.

[83] Hussain, S., Mustafa, M. W., Jumani, T. A., Baloch, S. K., Alotaibi, H., Khan, I., and Khan, A. 2021. A Novel Feature Engineered–CatBoost-based Supervised Machine Learning Framework for Electricity Theft Detection. *Energy Reports*. 7: 4425–4436. https://doi.org/10.1016/j.egyr.2021.07.008.

[84] Hussain, S., Mustafa, M. W., Al-Shqeerat, K. H. A., Saeed, F., and Al-rimy, B. A. S. 2021. A Novel Feature-Engineered–NGBoost Machine-Learning Framework for Fraud Detection in Electric Power Consumption Data. *Sensors*. 21(24): 8423. https://doi.org/10.3390/s21248423.

[85] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. 2021. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ*. 372: n71. https://doi.org/10.1136/bmj.n71.

[86] Kitchenham, B., Budgen, D., and Brereton, P. 2015. *Evidence-Based Software Engineering and Systematic Reviews*. Boca Raton: CRC Press.