Eksploatacja i Niezawodnosc - Maintenance and Reliability

Volume 29 (2027), Issue 2

https://ein.org.pl

The Polish Maintenance Society

Reliability-oriented diagnosis of non-technical losses in electricity distribution networks using real AMR data

Celal Ateşşahin[1], Hüseyin Acar[2,*], Cem Haydaroğlu[2]

[1] AMR Systems Engineer, Dicle Electric Distribution Inc., Türkiye
[2] Department of Electrical and Electronics Engineering, Engineering Faculty, Dicle University, Türkiye
* Corresponding Author: hacar@dicle.edu.tr

Received: 17 June 2026
Revised: 3 July 2026
Accepted: 8 July 2026
Online: 13 August 2026

Abstract

Non-technical losses in distribution networks pose critical risks to utility revenue, system reliability, monitoring accuracy, and field inspection efficiency. This study proposes an expert-assisted hybrid diagnostic framework using real Automatic Meter Reading (AMR) data from 951 subscribers in the Dicle distribution region. Hourly standardized data were labeled using six expert rules to address limited field-verified annotations. After undersampling, a balanced dataset of 304 risky and 304 clean subscribers was constructed, and Random Forest (RF), LSTM, and CNN models were evaluated. Random Forest achieved the best performance with 96.27% accuracy, 0.9952 AUC, 96.97% precision, 95.52% recall, and a 96.24% F1-score, outperforming LSTM and CNN. The results demonstrate that the proposed framework offers an effective and interpretable decision-support tool for reliability-oriented diagnostics and risk-based field inspection planning in smart distribution networks.

Keywords: *non-technical losses, diagnostics, distribution networks, risk-based inspection, AMR data*

Article citation:
Ateşşahin C, Acar H, Haydaroğlu C, Reliability-oriented diagnosis of non-technical losses in electricity distribution networks using real AMR data, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2027: 29(2) http://doi.org/10.17531/ein/226532

Highlights

- Reliability oriented NTL diagnosis is developed using real AMR data.
- The framework supports utility decision making and inspection prioritization.
- Current based indicators improve diagnostic consistency and field relevance.
- Random Forest provides reliable and interpretable NTL diagnostics.
- Expert rules support risk based inspection under limited verified labels.

1. Introduction

The energy sector is essential for economic development and strategic competitiveness; however, losses during electricity delivery significantly reduce energy efficiency. In the literature, these losses are commonly classified as Technical Losses (TLs), caused by the physical resistance of transmission and distribution lines, and Non-Technical Losses (NTLs), which represent a major global challenge for power systems and lead to substantial revenue losses for distribution utilities [1]. In distribution networks, NTLs mainly result from electricity theft, meter tampering, unauthorized connections, and billing errors, creating serious economic, technical, and operational impacts [2]. Although NTLs constitute a major share of total distribution losses especially in developing countries, they have also become increasingly complex in developed countries with the expansion of smart grid infrastructures [3]. Recent studies based on World Bank reports indicate that NTLs cause annual financial losses exceeding USD 96 billion for distribution companies [4]. Beyond revenue loss, NTLs reduce grid reliability, trigger voltage fluctuations, and impair energy demand forecasting accuracy [5].

Smart meters and Advanced Metering Infrastructure (AMI) have opened a new era in the detection of non-technical losses by enabling the acquisition of high-resolution consumption, current, and voltage data in electricity distribution systems [6]. Owing to smart meter-based approaches, traditional methods relying on manual field inspections have gradually been replaced by more scalable and automated data-driven analyses. Review and survey studies in this field have shown that smart meter data offer strong potential for NTL detection, while also introducing new challenges such as data quality, cybersecurity, and algorithmic complexity [7].

In recent years, machine learning (ML) and supervised learning methods have been widely studied for NTL detection. Algorithms such as Support Vector Machines, Random Forest, Logistic Regression, and gradient boosting have been used to identify abnormal consumption patterns from smart meter data [8]. Buzău et al. proposed a supervised learning approach using high-resolution consumption, alarm, and electrical measurement data from smart meters. Their results showed that XGBoost achieved superior performance under imbalanced data conditions; however, the method strongly depends on labeled inspection data and offers limited physical interpretability, which may restrict its generalizability to different grid structures and theft patterns [9]. Sahoo et al. developed a predictive model combining smart meter and distribution transformer measurements, where technical losses were estimated using temperature-dependent line resistances and theft was detected by separating technical losses from total losses. Although effective at higher theft rates, the need for transformer-level measurements, dependence on technical loss modeling accuracy, and reduced sensitivity to low-rate theft cases limit its large-scale applicability [10]. Zahoor Ali Khan et al. proposed a supervised theft detection framework using ADASYN, VGG-16-based feature extraction, and Firefly-optimized XGBoost, achieving high recall and F1-score; however, its computational complexity may reduce scalability in real-time and resource-constrained field applications [11].

Tanveer Ahmad examined AMI-based smart meter data and data mining methods, particularly SVM and linear regression, for NTL detection in distribution networks. Although the study offers a practical monitoring framework using real consumption data, its applicability may be limited by infrastructure cost, data quality, and computational burden [12]. Wenlin Han et al. proposed an RLS-based method for detecting colluded NTL fraud, enabling the identification of multiple fraudsters with limited data and low computational complexity. However, its reliance on synthetic validation and observer meter infrastructure limits practical deployment [13]. Ejaz Ul Haq et al. developed a Deep-CNN-based electricity theft detection method using high-resolution smart meter data. While the method improves accuracy and reduces false positives without manual feature extraction, its need for large labeled datasets and high computational power may restrict real-world scalability [14].

Pengwah et al. introduced a model-less WLS-based NTL detection method using smart meter voltage data, achieving good accuracy without detailed network topology, although measurement errors and simultaneous theft cases may reduce performance [15]. Han and Xiao proposed a simple NFD approach that avoids complex customer behavior models and detailed grid parameters, but it depends on a central observer meter and may be less accurate under dynamic theft conditions [16]. Faqishafyee et al. showed that smart meter deployment reduced total losses and theft cases in a real distribution network; however, the approach mainly depends on hardware investment rather than advanced analytical detection [17]. Seepromting et al. combined ML-based NTL detection with economic feasibility analysis and showed that smart meter replacement can be viable in high-loss regions, though the study was limited by dataset specificity and simplified operational assumptions [18].

Although several methods in the literature report high detection accuracy, important limitations remain. Most supervised learning approaches depend on reliable and fully labeled datasets, whereas obtaining verified electricity theft labels in real distribution systems is highly challenging [19]. Moreover, since NTL cases are rare, available datasets are often highly imbalanced, which reduces the detection capability for the minority class. Overall, survey studies provide broad insights but offer limited directly deployable solutions. Pure ML and deep learning methods may achieve strong predictive performance, yet they are constrained by labeling requirements, class imbalance, and limited interpretability. Conversely, physics-based approaches provide better explainability but often suffer from modeling complexity and limited field applicability. Therefore, existing methods still face challenges related to additional hardware needs, reliable labeling, and real-time implementation. The main advantages and limitations of these approaches are summarized in Table 1.

Table 1. Advantages and limitations of smart meter-based NTL detection methods.

Ref	Approach	Advantages	Limitations
[20]	Hamming code + energy balance	Simultaneously addresses technical and non-technical losses; can estimate the amount of loss	Requires additional check meters; increases infrastructure cost
[21]	Intermediate monitoring meters (IMM)	Can distinguish bypassing and meter tampering; supports localization	Requires additional hardware; installation and maintenance cost
[22]	State-estimation-based analytical method	Validated on a real network; enables NTL localization	Requires detailed network parameter information; mainly suitable for offline analysis
[23]	Physics-based + data-driven approach	Does not require labeled data; provides high interpretability	Sensitive to assumptions; performance may decrease under complex load profiles
[24]	Causal/preventive analysis	Based on real field data; reveals policy impact	Does not provide a direct detection algorithm

1.1. Comparative analysis with studies in literature

When the obtained results are compared with the methods and findings reported in recent studies in the literature, it is observed that the proposed hybrid model provides a clear advantage, particularly in terms of non-technical loss detection performance. As shown in Table 2, the proposed method achieves high performance on real and noisy field data, with 96.27% accuracy and 95.52% recall, without requiring complex data transformations. Among the studies reporting higher accuracy, Türk et al. used data from a limited industrial zone dataset consisting of only 400 meters and transformed the one-dimensional data into two-dimensional matrices to apply a Convolutional Neural Network (CNN)-based algorithm [26]. Adil et al., who used data from the State Grid Corporation of China (SGCC), applied intensive preprocessing steps such as interpolation, normalization, and RUSBoost-based random undersampling to overcome missing data and class imbalance problems, achieving 87.90% accuracy and 91.09% recall [1]. In contrast, Kara and Aksu analyzed data from the same institution (SGCC) without applying intensive data transformations, and their results remained limited to 78.10% accuracy with CatBoost and 69.43% recall with LightGBM [25]. In conclusion, the proposed study offers a highly applicable solution for real field operations by using a lightweight Random Forest algorithm and physically meaningful expert rules that establish a clear reference ground truth, without requiring complex spatial data transformations or aggressive undersampling/oversampling procedures.

Table 2. Comparison of the proposed model with state-of-the-art studies in the literature.

Ref	Method	Data/Approach	Accuracy (%)	Recall (%)
This Study	Expert Rules + Random Forest	Real and noisy field data	96.27	95.52
[1]	LSTM-RUSBoost	State Grid Corporation of China (SGCC) dataset	87.90	91.09
[25]	CatBoost / LightGBM	State Grid Corporation of China (SGCC) dataset	78.10	69.43
[26]	CNN (2D Image)	Diyarbakır Organized Industrial Zone (OIZ) dataset	97.50	75.25

1.2. Research gap and motivation

Non-technical losses (NTLs) vary significantly with a country's development level and distribution infrastructure. While electricity theft-related losses are about 6% in European countries and approximately 14% in Turkey [27], they have been reported to reach 53%–72% in some eastern provinces, including the Dicle Electricity Distribution Region considered in this study [28]. Such high rates show that NTLs are not only an economic problem but also a serious operational issue affecting grid reliability, energy planning, and inspection efficiency [29]. Therefore, practical and data-driven detection mechanisms are essential for high-loss distribution regions.

Conventional NTL detection mainly relies on static expert rules or random field inspections, which are inadequate for large-scale distribution networks serving millions of customers and thousands of transformers [27,28]. The increasing use of AMI and smart grid technologies enables continuous collection of consumption, current, and voltage data, creating opportunities for machine learning and deep learning-based NTL detection. However, two key challenges remain: the scarcity of reliable labeled fraud data in real systems [30] and the severe class imbalance caused by the rare occurrence of NTL events [25,31]. These limitations motivate hybrid frameworks that can operate with limited labels, handle imbalance, and support risk-based inspection planning. The main contributions of this study are as follows:

- A field-oriented hybrid NTL detection framework is developed using real AMR data collected from 951 subscribers in the Dicle distribution region.
- An expert-assisted reference labeling mechanism is proposed based on six industrial rules derived from field experience.
- Class imbalance is addressed using an undersampling strategy, and machine learning/deep learning models are evaluated under comparable conditions.
- Random Forest, LSTM, and CNN models are compared, showing that Random Forest offers a practical balance between performance, interpretability, and field applicability.
- The proposed framework provides a reliability-oriented decision-support tool for risk-based field inspection prioritization in distribution utilities.

2. Material and method

This section describes the main stages of the proposed hybrid methodology for non-technical loss detection. The framework covers data collection, preprocessing, expert-based labeling, class balancing, input preparation, and comparative model assessment. The complete workflow is shown in Fig. 1.

2.1. Study area and dataset

This study uses real AMR data obtained from 951 subscribers in the Dicle Electricity Distribution Region, one of the highest-loss areas in Türkiye [27,28]. The data were recorded in January 2026 and organized using a 7-day (168-h) observation window for each subscriber. To support both statistical feature extraction and time-series modeling, all measurements were converted to hourly resolution, and the dataset was split into 80% training and 20% testing sets.

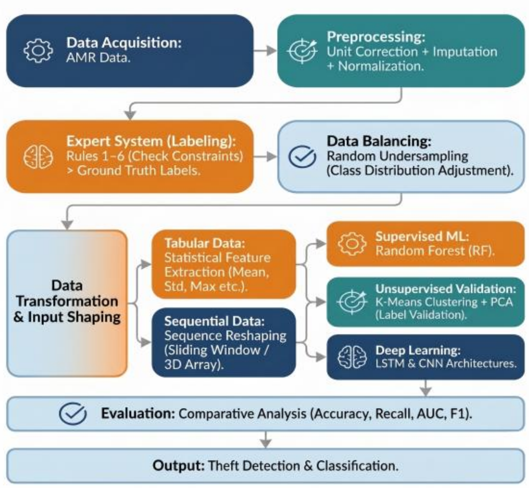

Figure 1. Overall workflow of the proposed hybrid framework for NTL detection.

Table 3. Description of the raw features obtained from the AMI system.

Feature Name	Symbol	Unit	Description	Resolution
Phase Voltage	V_{phase}	Volt (V)	Voltage measurements of R, S, T phases.	Hourly
Phase Current	I_{phase}	Ampere (A)	Current flowing through the phase line.	Hourly
Neutral Current	I_n	Ampere (A)	Current returning through the neutral line.	Hourly
Power Factor	$Cos\varphi$	-	The ratio of real power to apparent power.	Hourly
Relay Status	R_{st}	Binary (0/1)	0: Connected, 1: Cut-off	Event-based
Cumulative Index	E_total	kWh	Total energy consumption reading.	Hourly

To ensure a reliable evaluation, the dataset was divided into training and testing subsets on a subscriber basis rather than on a record basis. Since the dataset consists of continuous time-series measurements, a random record-based split could naturally result in consecutive data points from the same subscriber appearing in both the training and test sets, which would artificially inflate the predictive performance of the models. By keeping subscribers completely isolated between the training and test sets and applying stratified sampling based on risk labels, the models were tested on subscribers that had not been seen during training. This strategy ensures that the reported performance metrics better reflect the true diagnostic capability and field generalizability of the proposed framework. The raw data include load profile and meter readout variables such as current, voltage, power factor, relay status, and cumulative index values. The raw features used in the study are listed in Table 3.

2.2. Statistical analysis of the dataset

The raw dataset includes both single-phase and three-phase meters with different electrical characteristics. Among 176,151 records, 173,256 records (98.3%) belong to single-phase meters, while 2,895 records (1.7%) correspond to three-phase meters. Since the S and T phase measurements are structurally recorded as zero for single-phase meters, analyzing all records together may bias descriptive statistics such as mean and standard deviation. Therefore, the dataset was statistically evaluated separately by meter type, as shown in Tables 4 and 5.

Table 4. Descriptive statistics of single-phase meter data.

Variable	Mean	Min	Max	Std. Dev	Skewness	Kurtosis
Voltage R (V)	222.71	0.0	408.0	31.87	-6.41	**41.99**
Current R (A)	2.15	0.0	359.2	5.10	8.35	314.37
Current N (A)	**2.54**	0.0	359.2	5.21	5.91	159.08
Cosφ R	0.91	-1.0	1.0	0.37	-4.57	19.97
Index N (kWh)	6,849.03	0.0	66,495.2	9,485.68	1.79	3.74

Table 5. Descriptive statistics of three-phase meter data.

Variable	Mean	Min	Max	Std. Dev	Skewness	Kurtosis
Voltage R (V)	**242.77**	0.0	405.0	80.32	-0.41	2.74
Voltage S (V)	200.83	0.0	390.0	95.85	-1.14	0.85
Voltage T (V)	217.79	0.0	408.9	94.18	-0.62	1.48
Current R (A)	2.20	0.0	41.9	5.48	3.51	13.82
Current S (A)	1.09	0.0	31.4	3.37	4.13	20.25
Current T (A)	2.04	0.0	47.1	5.27	3.72	16.21
Cosφ (Avg)	0.75	-1.0	1.0	0.39	-1.66	1.84

The descriptive statistics reveal important dataset characteristics. For single-phase meters, the average neutral current (2.54 A) is higher than the average phase current (2.15 A), which may indicate abnormal patterns such as neutral-line manipulation or ground-return theft. The single-phase voltage data also show very high kurtosis (41.99), suggesting frequent interruptions, zero-voltage readings, and severe voltage sags, whereas the three-phase voltage profile is more stable with lower kurtosis (2.74). In addition, the maximum single-phase current reaches 359.2 A, indicating extreme outliers possibly related to abnormal loading or short-duration current shocks. Overall, these findings show that the raw data contain asymmetry, outliers, and meter-type-specific differences; therefore, preprocessing and feature engineering are necessary for reliable classification performance.

2.3. Data preprocessing

Raw AMR data may include noise and missing values caused by communication interruptions and sensor errors [32]. Therefore, a multi-stage preprocessing strategy was applied to improve data integrity and model reliability. First, current, voltage, and power measurements were standardized, and inconsistent records were removed. Physically impossible cases, such as negative consumption caused by meter replacement or reset events $\text{Index}_{\{t\}} < \text{Index}_{\{t-1\}}$, were also excluded. Missing values were handled using a hybrid imputation strategy. Short-term gaps (≤ 2 h), were completed with the Simple Moving Average (SMA) method to preserve local trends, whereas longer gaps (> 2 h), were estimated using regression-based models based on historical consumption patterns [33], [34]. Finally, mass-balance checks were performed to verify data consistency and reliability [35]. The overall preprocessing and imputation workflow is shown in Figure 2.

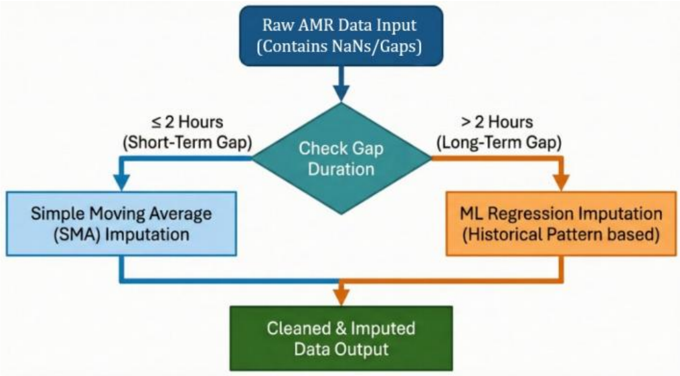

Figure 2. Hybrid preprocessing and imputation workflow for missing AMR data.

2.4. Labeling strategy: expert system rules

As stated in the Introduction, the lack of explicit “fraudulent” and “normal” labels in real field data is a major challenge for supervised NTL detection models [30]. To overcome this limitation, a deterministic expert system was developed using theft scenarios reported in the literature and field experience [5], [28]. The system evaluates AMR records based on predefined physical and operational inconsistencies. If at least one rule is satisfied, the subscriber is labeled as “Risky” (1); otherwise, the subscriber is labeled as “Clean” (0). These reference labels were then used for model training. The expert rules and their physical interpretations are summarized in Table 6.

Table 6. Expert system rules used for reference labeling and their physical interpretations.

Rule No.	Rule Name	Description	Risk Interpretation
1	Phase–Neutral Imbalance	In single-phase meters, the neutral current is significantly higher than the phase current.	Possible neutral-line manipulation or unauthorized current drawn from an external line
2	Voltage Limit Violation	Voltage values fall outside the critical operating range.	Possible neutral disconnection, voltage transformer intervention, or abnormal measurement behavior
3	Relay Bypass	Current continues to flow although the relay status is recorded as “disconnected.”	Possible relay bypass or direct meter circumvention
4	Shunting	Phase current is present while neutral current is absent or abnormally low.	Possible meter bypass through an external shunt path
5	Power Factor Manipulation	Physically inconsistent power factor behavior, especially in three-phase meters.	Possible phase-angle intervention or reactive manipulation causing under-recorded active power
6	Magnetic Interference and Saturation	Excessive asymmetry or abnormal deviations in voltage/current measurements that cannot be explained by normal operating conditions.	Possible magnetic saturation of meter sensing elements leading to distorted measurements

Table 7. Representative sample records of clean subscribers.

Record	Condition	VoltageR (V)	CurrentR (A)	CurrentN (A)	Cosφ	Relay
S1	Normal load	230.4	4.02	4.01	0.98	0
S2	No load	223.0	0.00	0.00	1.00	0
S3	Relay disconnected	220.5	0.00	0.00	0.00	1
S4	Normal voltage drop	215.0	5.10	5.08	0.95	0

Table 8. Representative sample records of risky subscribers.

Record	Condition	VoltageR (V)	CurrentR (A)	CurrentN (A)	Cosφ	Relay
K1	Phase–neutral imbalance	222.6	3.52	8.75	0.96	0
K2	Voltage limit violation	134.0	8.10	8.10	0.95	0
K3	Relay bypass	225.0	4.80	4.80	0.97	1
K4	Shunting	231.0	6.20	0.00	0.98	0
K5	Power factor manipulation	229.5	12.00	12.00	0.15	0
K6	Magnetic interference	45.0	350.00	350.00	0.50	0

The records in Table 7 represent typical non-risky operating conditions that were not flagged by the expert system. In these cases, phase and neutral currents are mutually consistent, the power factor remains within an acceptable range, and the relay-current relationship is physically plausible.

The examples in Table 8 illustrate typical anomalous patterns that were labeled as risky by the expert system. These records reflect possible theft-related or measurement-related abnormalities, including current imbalance, voltage deviation, relay-current inconsistency, shunting behavior, power factor anomalies, and extreme measurement deviations.

2.5. Data transformation and rule-based feature engineering

Although raw AMR data include basic electrical measurements such as voltage and current, these variables alone provide limited discriminative capability for NTL detection; therefore, a two-stage data transformation and feature engineering process was applied to improve the representation capacity of the machine learning models [26]. First, the “Next Logic” approach was used to evaluate changes between consecutive time steps, enabling the detection of abrupt and physically inconsistent variations in relay status, current flow, and power factor, and allowing anomalies such as relay bypass and power factor manipulation to be identified through temporal change patterns rather than isolated measurements [36]. Second, each subscriber's 168-h time-series data were converted into weekly statistical features, including mean neutral current, current standard deviation, and the number of triggered rule violations [25,28]. These features represent possible fraudulent consumption, consumption fluctuation, and the persistence of expert-defined anomalies, respectively, thereby transforming raw time-series data into a more meaningful and discriminative feature space for classification.

2.6. Random forest model

Random Forest (RF) is a powerful ensemble learning method widely used in classification and regression problems due to its high accuracy and strong generalization capability [32]. The method is based on training multiple decision trees independently on different subsets generated through bootstrap sampling from the original training dataset. In addition, at each split, only a randomly selected subset of features is evaluated rather than the full feature set, which reduces correlation among trees, improves generalization performance, and mitigates overfitting [37-39]. For an input x, the prediction of the RF model $(\hat{y})$, is obtained from the outputs of T decision trees, $h_{t(x)}$, and can be expressed as follows [40]:

$$\hat{y} = mod\left(h_{1(x)}, h_{2(x)}, \dots, h_{T(x)}\right) \tag{1}$$

In this study, Random Forest (RF) was used as the baseline model for tabular data obtained through statistical feature engineering, including mean, maximum, and standard deviation-based features. Its robustness to noise and resistance to overfitting make RF suitable for imbalanced real-world datasets. In the classification process, the final decision is determined by majority voting among individual decision trees. As shown in Figure 3, the RF-based framework consists of real AMR data collection, expert-based labeling, imbalance handling, and final classification output generation.

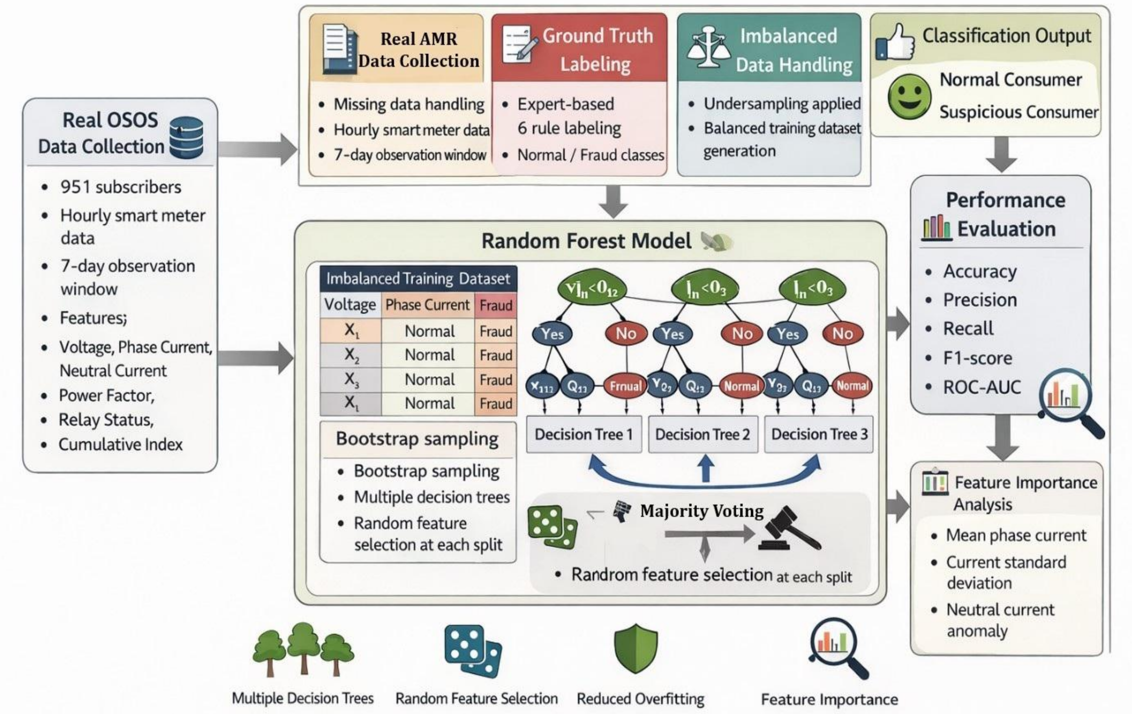

Figure 3. RF-based classification framework for NTLs detection using AMI data.

In this study, Random Forest (RF) was selected as a main supervised classifier for NTL detection due to its strong performance on tabular data, ability to model nonlinear relationships, robustness to noise, and interpretability through feature importance analysis. As shown in Figure 3, the RF-based framework integrates bootstrap sampling, multiple decision trees, random feature selection, majority voting, and evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. In addition, the feature importance ranking, including mean phase current, current standard deviation, and neutral current anomaly, supports the physical interpretation of the decision process. Overall, RF provides an effective and practical solution for detecting NTLs, particularly when statistical and rule-based features are used.

2.7. Setup and hyperparameter optimization

To ensure the reproducibility of the study and to enable a fair comparison between machine learning and deep learning architectures, the hyperparameters of all models were carefully configured. The time-series data were structured using a sliding-window approach with a 24-hour window size. Both the LSTM and CNN models were trained using the Adam optimizer with a learning rate of 0.001, a batch size of 64, and a maximum of 40 epochs. To prevent overfitting and ensure optimal generalization, an Early Stopping mechanism was applied by monitoring the validation loss with a patience value of 5 epochs and restoring the best model weights. The Random Forest classifier was configured as an ensemble consisting of 100 decision trees to provide stable majority voting. The detailed architectural configurations and hyperparameter settings of all evaluated models are summarized in Table 9.

Table 9. Hyperparameter configurations and model architectures.

Model	Parameter	Value
Global Settings	Time Steps (Window Size)	24 Hours
	Batch Size	64
	Max Epochs	40
	Validation Split	20%
	Early Stopping	Patience = 5, Monitor = Validation Loss
	Optimizer	Adam (Learning Rate = 0.001)
	Loss Function	Binary Crossentropy
Random Forest	Number of Trees (n_estimators)	100
	Splitting Criterion	Gini Impurity (Default)
LSTM	Layer 1 (Sequence Learning)	LSTM (64 Units), Return Sequences = False
	Regularization	Dropout (Rate = 0.20)
	Layer 2 (Dense)	Fully Connected (32 Units), ReLU Activation
	Output Layer	Fully Connected (1 Unit), Sigmoid Activation
CNN	Layer 1 (Feature Extraction)	Conv1D (32 Filters, Kernel Size = 3), ReLU, Same Padding
	Layer 2 (Pooling)	MaxPooling1D (Pool Size = 2)
	Layer 3 (Flattening)	Flatten
	Layer 4 (Dense)	Fully Connected (50 Units), ReLU Activation
	Output Layer	Fully Connected (1 Unit), Sigmoid Activation

3. Findings and discussion

This section evaluates the proposed hybrid framework under two scenarios: the original imbalanced dataset (Scenario B) and the undersampling-based balanced dataset (Scenario A). RF was compared with LSTM and CNN to assess classification performance and robustness under different data distributions. The results are also discussed in relation to similar studies in the literature.

3.1. Scenario A: performance on the balanced dataset

To reduce model bias, the training process was repeated on a balanced dataset consisting of 304 risky and 304 clean subscribers. Random Forest, LSTM, CNN, and K-Means models were comparatively evaluated on this dataset, and the corresponding test results are summarized in Table 10.

Table 10. Performance comparison of the models on the balanced dataset.

Model	Accuracy	AUC	Precision	Recall	F1-Score
Random Forest	0.9627	0.9952	0.9697	0.9552	0.9624
LSTM	0.9152	0.9302	0.9485	0.8717	0.9085
CNN	0.8800	0.9125	0.9087	0.8353	0.8705
K-Means	0.5075	0.5000	1.0000	0.0149	0.0294

As shown in Table 8, the Random Forest model achieved the best overall performance on the balanced dataset, with 96.27% accuracy and 95.52% recall, outperforming both LSTM (91.52% accuracy) and CNN (88.00% accuracy). In addition, Random Forest yielded the highest AUC (0.9952) and F1-score (0.9624), indicating a superior balance between discrimination capability and classification reliability. By contrast, the K-Means model produced very poor results, with near-random classification performance, confirming that unsupervised clustering is not suitable for this dataset structure.

The fact that K-Means remained at a random prediction level (AUC = 0.50) is an expected outcome due to the unsupervised nature of the model. Since the labeling process is based on logical conditions defined by expert rules, the class separation can be effectively captured by rule-based and supervised algorithms such as Random Forest, whereas K-Means, which relies solely on distance-based clustering, fails to distinguish these rule-driven patterns.

Figure 4 shows that the Random Forest model achieved balanced and reliable classification performance on the balanced dataset. Among 134 randomly selected samples, it correctly classified 65 of 67 normal cases and 64 of 67 risky cases, with only 2 false alarms and 3 missed risky cases. This low error rate indicates that the model can effectively identify risky subscribers while minimizing unnecessary inspections, making it a practical decision-support tool for field inspection planning.

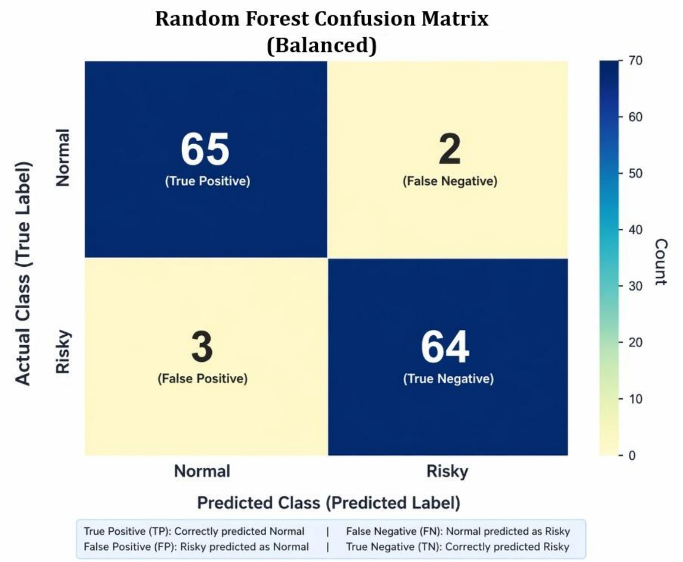

Figure 4. Confusion matrix of the Random Forest model on the balanced dataset.

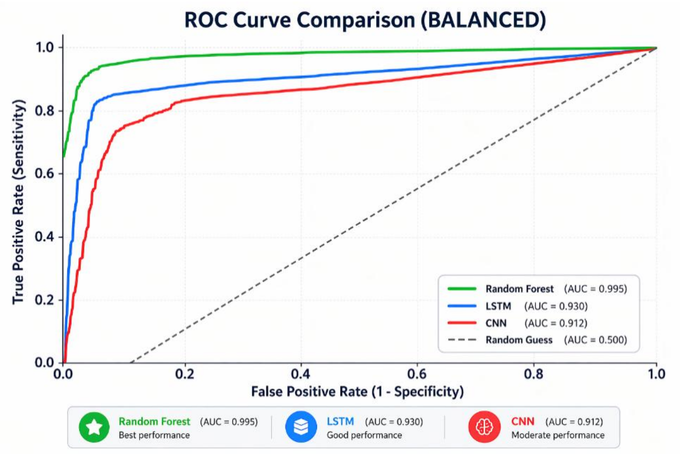

Figure 5. ROC curve comparison of the balanced dataset.

As shown in Figure 5, Random Forest achieved the highest discriminative performance on the balanced dataset, with an AUC of 0.995, indicating near-perfect separation between risky and clean subscribers. Although LSTM and CNN also showed acceptable performance with AUC values of 0.930 and 0.912, respectively, they remained below Random Forest. These results confirm that Random Forest is a more suitable and reliable classifier for balanced datasets based on statistical and rule-based features.

3.2. Scenario B: performance on the imbalanced dataset

In the first stage, the expert-labeled original dataset was used without any balancing procedure. The dataset included 68% risky samples as the majority class and 32% clean samples as the minority class. RF, LSTM, CNN, and K-Means models were trained and tested on this imbalanced dataset, and the results are summarized in Table 11.

Table 11. Performance comparison of the models on the imbalanced dataset.

Model	Accuracy	AUC	Precision	Recall	F1-Score
Random Forest	0.9843	0.9949	0.9840	0.9919	0.9880
CNN	0.9104	0.9408	0.9350	0.9213	0.9281
LSTM	0.9061	0.9336	0.9610	0.8863	0.9221
K-Means	0.6492	0.5000	0.6492	1.0000	0.7873

As shown in Table 11, Random Forest achieved the best overall performance, with 98.43% accuracy and 99.19% recall, outperforming LSTM and CNN. Although both deep learning models exceeded 90% accuracy, K-Means showed no meaningful discrimination, with an AUC of 0.50. Its recall of 1.00 suggests that most samples were assigned to the majority risky class, reducing its practical usefulness despite the seemingly acceptable F1-score.

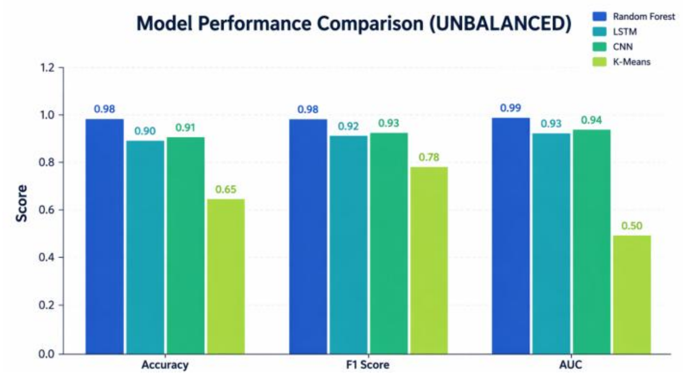

Figure 6. Performance comparison of the imbalanced dataset.

As shown in Figure 6, Random Forest achieved the highest performance on the imbalanced dataset, outperforming the other methods in accuracy, F1-score, and AUC. Although CNN and LSTM produced acceptable results, they remained below RF. In contrast, K-Means obtained an AUC of 0.50, indicating random-level discrimination. These results show that supervised learning methods are more effective for imbalanced NTL detection, with Random Forest providing the most reliable and practical classification performance.

3.3. Feature importance analysis

To improve model interpretability and assess the physical relevance of the decision process, the feature importance scores of the Random Forest classifier were analyzed. As shown in Fig. 7, mean phase current, current standard deviation, and maximum current were the most influential features. This indicates that the model mainly relies on current-based behavior rather than voltage-related variables to distinguish risky and clean subscribers. The dominance of mean and maximum current highlights the role of abnormal load intensity and excessive current draw, while current standard deviation reflects irregular consumption fluctuations as an important indicator of suspicious behavior.

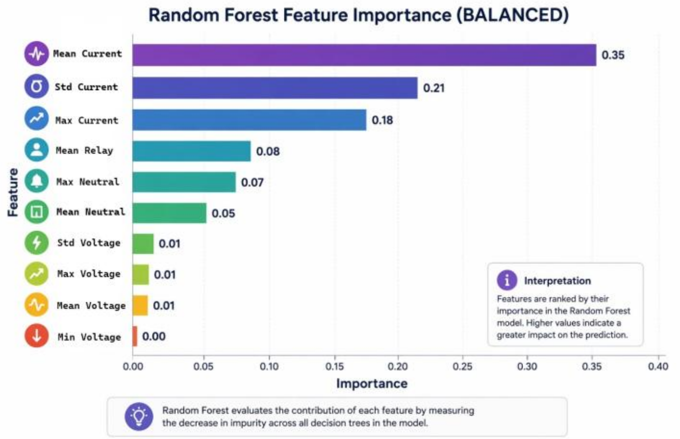

Figure 7. Feature importance ranking of the RF model on the balanced dataset.

In addition, the high importance of maximum and mean neutral current confirms the relevance of neutral-line anomalies in the classification process, consistent with the expert rules for phase–neutral imbalance and shunting behavior. In contrast, voltage-based features had lower importance scores, indicating that current-derived statistical descriptors are more informative for this dataset. Overall, the feature importance analysis shows that the Random Forest model achieves high performance while capturing physically interpretable patterns related to non-technical loss behavior.

Figure 8 shows strong positive correlations among current- and neutral current-based features, particularly between mean current and mean neutral current (0.86), current standard deviation and maximum current (0.82), and maximum current and maximum neutral current (0.77). In contrast, voltage-related features show weak correlations with current-based variables, indicating that they represent different physical behaviors. This explains the higher importance of current-based features in the RF model and supports the feature importance results.

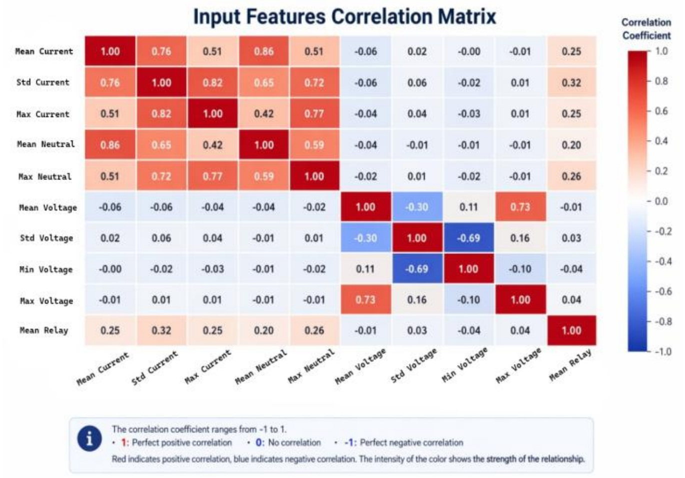

Figure 8. Correlation matrix of the extracted features used in the classification model.

4. Conclusion

This study proposed an expert-assisted hybrid detection framework for identifying non-technical losses in electricity distribution networks using real AMR data obtained from the Dicle distribution region. Rather than relying solely on predictive performance metrics, the study demonstrates how the integration of rule-based expert knowledge with machine learning can effectively address the critical challenges posed by imbalanced and unverified field data. The results show that Random Forest outperformed LSTM and CNN on the balanced dataset, achieving 96.27% accuracy, 0.9952 AUC, 96.97% precision, 95.52% recall, and a 96.24% F1-score. The comparative evaluation revealed that the Random Forest model provided the most robust and interpretable performance compared with deep learning architectures such as LSTM and CNN. Beyond achieving high discriminative capability, the dominance of indicators such as mean and maximum neutral current in the model decision process strongly confirms the physical validity of the proposed method.

From an operational perspective, the proposed expert-assisted hybrid method contributes not only to the economic dimension of non-technical loss detection but also directly to network safety and operational reliability. This early diagnostic mechanism provides an additional protection layer against transformer overloads, local power interruptions, and physical interventions in metering infrastructure. Moreover, the high-confidence predictions of the Random Forest model can guide field inspection teams directly toward high-risk subscribers, thereby optimizing inspection efficiency, reducing operational costs, improving the quality of network monitoring in critical areas, and contributing to the efficient use of national energy resources.

Despite its strong practical applicability, the study has certain limitations. The proposed method relies on reference labels and uses data from a limited period corresponding only to January. Therefore, the model may not fully capture consumption behaviors arising from seasonal conditions, which may affect its long-term generalizability. For this reason, future studies should focus on validating the proposed framework using one-year datasets in order to account for seasonal variations. In addition, cross-validating expert rules with physically verified field inspection records and integrating explainable real-time learning strategies will be essential steps toward developing a fully autonomous and adaptive diagnostic system for smart distribution networks.

Acknowledgment

This research is supported by the Research Projects Committee of Dicle University (DUBAP) with the project numbers MUHENDISLIK.26.024. We are grateful to DUBAP for the support.

References

1. Adil M, Javaid N, Qasim U, Ullah I, Shafiq M, Choi JG. LSTM and bat-based RUSBoost approach for electricity theft detection. *Applied Sciences* 2020; 10(12): 4378. https://doi.org/10.3390/app10124378.
2. Chandel P, Thakur T, Sawle BA, Sharma R. Power theft: Major cause of non technical losses in Indian distribution sector. *2016 IEEE 7th Power India International Conference (PIICON)*, p. 1-6. https://doi.org/10.1109/POWERI.2016.8077253.
3. Yang Z, Wei J, Wei X, Pan J, Jiang W, Wei H. A review of cybersecurity and non-technical loss model in the smart grid. *2025 CAA Symposium on Fault Detection, Supervision, and Safety for Technical Processes (SAFEPROCESS)*, p. 1-6. https://doi.org/10.1109/SAFEPROCESS67117.2025.11267897.
4. Lepolesa LJ, Achari S, Cheng L. Electricity theft detection in smart grids based on deep neural network. *IEEE Access* 2022; 10: 39638-39655. https://doi.org/10.1109/ACCESS.2022.3166146.
5. Depuru SSSR, Wang L, Devabhaktuni V. Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft. *Energy policy* 2011; 39(2): 1007-1015. https://doi.org/10.1016/j.enpol.2010.11.037.
6. Xia X, Xiao Y, Liang W, Cui J. Detection methods in smart meters for electricity thefts: A survey. *Proceedings of the IEEE* 2022; 110(2): 273-319. https://doi.org/10.1109/JPROC.2021.3139754.
7. Anas M, Javaid N, Mahmood A, Raza SM, Qasim U, Khan ZA. Minimizing electricity theft using smart meters in AMI. *2012 Seventh International conference on P2P, parallel, grid, cloud and internet computing*, p. 176-182. IEEE. https://doi.org/10.1109/3PGCIC.2012.42.
8. Chuwa MG, Wang F. A review of non-technical loss attack models and detection methods in the smart grid. *Electric Power Systems Research* 2021; 199: 107415. https://doi.org/10.1016/j.epsr.2021.107415.
9. Buzau MM, Tejedor-Aguilera J, Cruz-Romero P, Gómez-Expósito A. Detection of non-technical losses using smart meter data and supervised learning. *IEEE Transactions on Smart Grid* 2018; 10(3): 2661-2670. https://doi.org/10.1109/TSG.2018.2807925.
10. Sahoo S, Nikovski D, Muso T, Tsuru K. Electricity theft detection using smart meter data. *2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, p. 1-5, IEEE. https://doi.org/10.1109/ISGT.2015.7131776.
11. Khan ZA, Adil M, Javaid N, Saqib MN, Shafiq M, Choi JG. Electricity theft detection using supervised learning techniques on smart meter data. *Sustainability* 2020; 12(19): 8023. https://doi.org/10.3390/su12198023.
12. Ahmad T. Non-technical loss analysis and prevention using smart meters. *Renewable and Sustainable Energy Reviews* 2017; 72: 573-589. https://doi.org/10.1016/j.rser.2017.01.100.
13. Han W, Xiao Y. A novel detector to detect colluded non-technical loss frauds in smart grid. *Computer Networks* 2017; 117: 19-31. https://doi.org/10.1016/j.comnet.2016.10.011.
14. Haq EU, Pei C, Zhang R, Jianjun H, Ahmad F. Electricity-theft detection for smart grid security using smart meter data: A deep-CNN based approach. *Energy Reports* 2023; 9: 634-643. https://doi.org/10.1016/j.egyr.2022.11.072.
15. Pengwah AB, Razzaghi R, Andrew LL. Model-less non-technical loss detection using smart meter data. *IEEE Transactions on Power*

Delivery 2023; 38(5): 3469-3479. https://doi.org/10.1109/TPWRD.2023.3280551.

16. Han W, Xiao Y. NFD: A practical scheme to detect non-technical loss fraud in smart grid. *2014 IEEE International Conference on Communications (ICC)*, p. 605-609. IEEE. https://doi.org/10.1109/ICC.2014.6883385.
17. Faqishafyee NJ, Sadiq EH, Taha HM. Mitigating non-technical losses and electricity theft through smart meters: A case study of the Akre district power distribution network. *Journal of Intelligent Systems and Control* 2024; 3(3): 135-151. https://doi.org/10.56578/jisc030301.
18. Seepromting K, Rajitroj N, Michel J, Surawanitkun C. Strategic Smart Meter Deployment in Power Distribution: Leveraging Machine Learning and Economic Evaluation for Non-Technical Loss Reduction. *2025 60th International Universities Power Engineering Conference (UPEC)*, p. 1-7, IEEE. https://doi.org/10.1109/UPEC65436.2025.11279864.
19. Viegas JL, Esteves PR, Melicio R, Mendes VMF, Vieira SM. Solutions for detection of non-technical losses in the electricity grid: A review. *Renewable and Sustainable Energy Reviews* 2017; 80: 1256-1268. https://doi.org/10.1016/j.rser.2017.05.193.
20. Shah AL, Mesbah W, Al-Awami AT. An algorithm for accurate detection and correction of technical and nontechnical losses using smart metering. *IEEE Transactions on Instrumentation and Measurement* 2020; 69(11): 8809-8820. https://doi.org/10.1109/TIM.2020.2999175.
21. Kim JY, Hwang YM, Sun YG, Sim I, Kim DI, Wang X. Detection for non-technical loss by smart energy theft with intermediate monitor meter in smart grid. *IEEE Access* 2019; 7: 129043-129053. https://doi.org/10.1109/ACCESS.2019.2940443.
22. Raggi LM, Trindade FC, Cunha VC, Freitas W. Non-technical loss identification by using data analytics and customer smart meters. *IEEE Transactions on Power Delivery* 2020; 35(6): 2700-2710. https://doi.org/10.1109/TPWRD.2020.2974132.
23. Gao Y, Foggo B, Yu N. A physically inspired data-driven model for electricity theft detection with smart meter data. *IEEE Transactions on Industrial Informatics* 2019; 15(9): 5076-5088. https://doi.org/10.1109/TII.2019.2898171.
24. Otchere-Appiah G, Takahashi S, Yeboah MS, Yoshida Y. The impact of smart prepaid metering on non-technical losses in Ghana. *Energies* 2021; 14(7): 1852. https://doi.org/10.3390/en14071852.
25. Kara Y, Aksu A. Tüketicilerde Kaçak Elektrik Kullanımının Akıllı Sayaç Verisi Üzerinden Gradyan Artırmalı Karar Ağacı Tabanlı Makine Öğrenmesi Yöntemleriyle Tespiti. *Journal of Investigations on Engineering and Technology* 2023; 6(1): 1-12. https://dergipark.org.tr/en/pub/jiet/article/1278937.
26. Türk M, Haydaroglu C, Kılıç H. Machine learning-based detection of non-technical losses in power distribution networks. Firat University *Journal of Experimental and Computational Engi*neering 2025; 4(1): 192-205. https://doi.org/10.62520/fujece.1551601.
27. Kocaman B. Teknik olmayan enerji kayıplarının azaltılmasında PLC sayaçlarının önemi. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi* 2018; 7(2): 220-230. https://doi.org/10.17798/bitlisfen.420198.
28. Çelikpençe, M. Elektrik Dağıtım Şebekelerinde Teknik Olmayan Kayıp Kaçakların Makina Öğrenmesi İle Tespiti, PhD Thesis, İstanbul Sabahattin Zaim Üniversitesi 2023.
29. Depuru SSSR, Wang L, Devabhaktuni V. Support vector machine based data classification for detection of electricity theft. *2011 IEEE/PES Power Systems Conference and Exposition*, p. 1-8. IEEE. https://doi.org/10.1109/PSCE.2011.5772466.
30. Villar-Rodriguez E, Del Ser J, Oregi I, Bilbao MN, Gil-Lopez S. Detection of non-technical losses in smart meter data based on load curve profiling and time series analysis. *Energy* 2017; 137: 118-128. https://doi.org/10.1016/j.energy.2017.07.008.
31. Glauner P, Meira, JA, Valtchev, P, State R, Bettinger F. The challenge of non-technical loss detection using artificial intelligence: A survey. *International journal of computational intelligence systems* 2017; 10(1): 760-775. https://doi.org/10.2991/ijcis.2017.10.1.51.
32. Emmanuel T, Maupong T, Mpoeleng D, Semong T, Mphago B, Tabona O. A survey on missing data in machine learning. *Journal of Big data* 2021; 8(1): 140. https://doi.org/10.1186/s40537-021-00516-9.
33. Hou Z, Liu J. Enhancing smart grid sustainability: using advanced hybrid machine learning techniques while considering multiple influencing factors for imputing missing electric load data. *Sustainability* 2024; 16(18): 8092. https://doi.org/10.3390/su16188092.
34. Jung S, Moon J, Park S, Rho S, Baik SW, Hwang E. Bagging ensemble of multilayer perceptrons for missing electricity consumption data imputation. *Sensors* 2020; 20(6): 1772. https://doi.org/10.3390/s20061772.
35. Gardian H, Beck JP, Koch M, Kunze R, Muschner C, Hülk L, Bucksteeg M. Data harmonisation for energy system analysis–Example of multi-model experiments. *Renewable and Sustainable Energy Reviews* 2022; 162: 112472. https://doi.org/10.1016/j.rser.2022.112472.
36. Nagi J, Yap KS, Tiong SK, Ahmed SK, Mohamad M. Nontechnical loss detection for metered customers in power utility using support vector machines. *IEEE transactions on Power Delivery* 2009; 25(2): 1162-1171. https://doi.org/10.1109/TPWRD.2009.2030890.

37. Aydın EY, Önal K, Haydaroğlu C, Kılıç H, Yıldırım Ö, Katar O, Erdoğan H. A Novel Scenario-Based Comparative Framework for Short- and Medium-Term Solar PV Power Forecasting Using Deep Learning Models. *Applied Sciences* 2025; 15(24): 12965. https://doi.org/10.3390/app152412965.
38. Abro, SA, Laghari, JA, Memon, SA, Khan TA, Memon I, Nasir H, Fatima, K. Non-technical loss detection in power distribution networks using machine learning. *Scientific Reports* 2025; 15(1): 36189. https://doi.org/10.1038/s41598-025-20048-z.
39. Saeed MS, Mustafa MW, Sheikh UU, Jumani TA, Mirjat NH. Ensemble bagged tree based classification for reducing non-technical losses in multan electric power company of Pakistan. *Electronics* 2019; 8(8): 860. https://doi.org/10.3390/electronics8080860.
40. Ghori KM, Imran M, Nawaz A, Abbasi RA, Ullah A, Szathmary, L. Performance analysis of machine learning classifiers for non-technical loss detection. *Journal of Ambient Intelligence and Humanized Computing* 2023; 14(11): 15327-15342 https://doi.org/10.1007/s12652-019-01649-9.