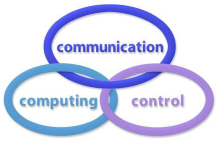
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 21, Issue: 1, Month: February, Year: 2026
Article Number: 7409, https://doi.org/10.15837/ijccc.2026.1.7409



CCC Publications

Interpretable Deep Learning using Kolmogorov–Arnold Networks for Energy Theft Detection

B. C. Neagu, A. A. Gugiuman, Gh. Grigoraș, A. Simo, S. Dzitac

Bogdan-Constantin Neagu
Gheorghe Asachi Technical University of Iași
700050 Iași, Bd. Dimitrie Mangeron No. 67, Romania
*Corresponding author: bogdan-constantin.neagu@academic.tuiasi.ro

Adi-Aurelian Gugiuman
Gheorghe Asachi Technical University of Iași
700050 Iași, Bd. Dimitrie Mangeron No. 67, Romania
adi-aurelian.gugiuman@student.tuiasi.ro

Gheorghe Grigoraș
Gheorghe Asachi Technical University of Iași
700050 Iași, Bd. Dimitrie Mangeron No. 67, Romania
gheorghe.grigoras@academic.tuiasi.ro

Attila Simo
Politehnica University Timişoara
300223 Timişoara, Piata Victoriei No. 2, Romania
attila.simo@upt.ro

Simona Dzitac
University of Oradea
410087 Oradea, 1 Universitatii No. 1, Romania
sdzitac@uoradea.ro

Abstract

The increasing availability of high-resolution metering data has positioned deep learning as a powerful tool for energy theft detection; however, most existing approaches rely on black-box models that lack interpretability and require large, labeled datasets, limiting their applicability in regulated and safety-critical environments. This paper proposes an interpretable deep learn-ing approach for energy theft detection based on Kolmogorov–Arnold Net-works (KAN), explicitly de-signed to combine non-linear modeling capability with feature-level transparency. The novelty of the proposed method lies in embedding explainability directly into the learning architecture, rather than relying on post hoc explanation techniques. By representing the detection function as a com-position of learnable univariate spline functions, the KAN-based model enables direct visualization and quantitative interpretation of each input feature's contribution to the detection outcome. This property al-lows energy theft detection to be formulated as a transparent decision-support process,

suitable for operational deployment and regulatory auditing. The proposed approach integrates engineered consumption features with KAN-based inference and supports deployment within a distributed edge in-telligence architecture, enabling low-latency detection and privacy-aware processing. Experimental evaluation on representative low-voltage network data demonstrates that the proposed approach achieves competitive detection accuracy while significantly improving interpretability and robustness under prosumer-induced variability. The results confirm that inherently interpreta-ble deep learning models represent a viable and effective alternative to con-ventional black-box techniques for energy theft detection.

Keywords: Interpretable Deep Learning, Energy Theft Detection, Kolmogorov–Arnold Networks.

1 Introduction

1.1 Context and Motivation

The rapid digitalization of energy systems has transformed electrical distribution networks into large-scale cyber–physical systems characterized by pervasive sensing, bidirectional information exchange, and increasingly autonomous decision-making processes. Within this context, artificial intelligence (AI) has emerged as a key technological enabler for monitoring, control, and anomaly detection in complex and highly dynamic environments. Deep learning (DL) techniques have demonstrated remarkable capabilities in modeling non-linear relationships and extracting high-level representations from large volumes of data generated by smart meters, sensors, and communication infrastructures. However, despite their strong predictive performance, the practical deployment of deep learning models in operational power systems remains constrained by fundamental challenges related to interpretability, data quality, and architectural scalability.

One of the most critical application domains where these challenges converge is the detection of non-technical losses (NTLs), with electricity theft representing the dominant component. NTLs arise from a wide range of illicit activities, including meter tampering, bypassing, illegal connections, and unregistered withdrawals, and they directly affect both the technical efficiency and the economic sustainability of distribution networks. According to Depuru et al. [1], such losses may account for up to 30% of total distributed energy in developing regions, affecting both technical performance and financial stability. From a system-level perspective, persistent NTLs distort feeder energy balances, degrade voltage quality, increase thermal loading, and reduce the effective hosting capacity for distributed energy resources. From an economic standpoint, they erode utility revenues, limit reinvestment capacity, and introduce tariff distortions that disproportionately penalize compliant consumers. Non-technical losses (NTL), particularly electricity theft, represent one of the most persistent challenges in modern distribution systems.

Although the magnitude of NTLs varies significantly across regions, their impact remains substantial even in highly regulated and technologically advanced power systems. In low-voltage (LV) distribution networks, where consumption behavior is heterogeneous and increasingly influenced by distributed generation, storage systems, and flexible loads, the identification of anomalous patterns becomes particularly challenging. The growing penetration of photovoltaic (PV) systems, electric vehicles (EVs), and prosumer-oriented energy services has fundamentally altered load profiles, introducing bidirectional power flows, strong temporal variability, and non-stationary consumption patterns. In such environments, legitimate behavioral variability can closely resemble classical theft signatures, blurring the boundary between normal operation and fraudulent activity. Even in European networks, NTL typically range between 5% and 8% of total energy, especially in low-voltage feeders with mixed consumer behavior [2]. These losses undermine investment capacity and create tariff distortions that penalize compliant consumers [3].

Within the broader AI landscape, deep learning has gained particular attention due to its ability to model high-dimensional data and uncover latent structures in large datasets. Convolutional neural networks, recurrent architectures, and hybrid temporal models have been successfully applied to a wide range of anomaly detection tasks, including network intrusion detection, industrial fault diagnosis, and financial fraud identification. In the context of energy systems, deep learning models have demonstrated strong performance in load forecasting, demand classification, and consumption

pattern recognition. Nevertheless, their application to NTL detection introduces specific constraints that distinguish this domain from conventional pattern recognition problems. Traditional detection through field inspections and manual audits is costly and reactive [4]. The widespread deployment of smart meters has generated large volumes of consumption data, creating opportunities for automated, data-driven detection [5]. However, the diversification of load patterns caused by distributed generation, electric-vehicle charging, and prosumer participation complicates the distinction between legitimate variation and fraud [6]. Europe-an regulatory bodies now encourage the digitalisation of energy systems [7] and promote AI-assisted supervision to improve transparency and efficiency [8].

First, NTL detection is inherently a high-stakes decision-making task, where false positives may lead to unjustified inspections, customer disputes, and reputational damage, while false negatives allow fraudulent behavior to persist. Second, con-firmed theft cases typically represent only a small fraction of the available data, resulting in severe class imbalance and limited availability of labeled samples. Third, regulatory frameworks governing energy systems impose strict requirements on transparency, auditability, and fairness, which are difficult to reconcile with opaque black-box models. As a result, purely data-driven deep learning approaches, despite their accuracy, often face resistance from operators and regulators due to their lack of explainability.

The issue of interpretability has therefore emerged as a central challenge in the application of AI to energy theft detection. Explainable Artificial Intelligence (XAI) techniques have been proposed to bridge the gap between predictive performance and human understanding, offering post hoc explanations of model outputs through feature importance analysis, local surrogate models, or gradient-based visualization methods. While such techniques provide valuable insights, they are typically applied after model training and remain external to the learning process itself. Consequently, explanations may be qualitative, computationally expensive, or inconsistent across operating conditions, limiting their usefulness in real-time monitoring and regulatory auditing.

These limitations have motivated increasing interest in inherently interpretable deep learning models, in which transparency is embedded directly into the mathematical structure of the learning algorithm. From an AI perspective, such models aim to preserve the expressive power of deep learning while enabling explicit interpretation of feature contributions and decision pathways. Among the emerging approaches in this area, Kolmogorov–Arnold Networks (KAN) have attracted attention as a promising alternative to conventional multilayer perceptrons. Grounded in the Kolmogorov–Arnold superposition theorem, KAN architectures approximate multi-variate functions as compositions of learnable univariate spline functions, allowing each input variable to be associated with an explicit, visualizable contribution to the model output.

In the context of energy theft detection, this property offers a significant conceptual advantage. Instead of treating consumption features as opaque inputs to a black-box classifier, KAN-based models enable a transparent mapping between engineered indicators and anomaly scores. This facilitates direct validation of AI decisions against physical reasoning, such as assessing whether a detected anomaly arises from feeder-level energy imbalance, abnormal temporal recurrence, or peer-group inconsistency. As a result, KANs provide a natural bridge between classical engineering knowledge and modern deep learning, aligning AI-driven detection with operational and regulatory requirements.

Beyond model interpretability, architectural considerations play a crucial role in the deployment of AI-based detection systems. Centralized processing architectures, in which all data are transmitted to a remote cloud for analysis, face scalability, latency, and privacy challenges, particularly in LV networks. The continuous trans-mission of high-resolution metering data places significant strain on communication infrastructures and raises concerns regarding data protection and customer trust. Moreover, centralized inference may introduce unacceptable delays in detection, limiting the ability to respond promptly to persistent irregularities. To address these challenges, recent research has increasingly explored edge intelligence as a deployment paradigm for AI in cyber–physical systems. By executing lightweight inference and feature extraction tasks closer to the data source, edge-based architectures enable low-latency, context-aware decision-making while reducing communication overhead. From an AI and control perspective, this distributed intelligence paradigm aligns with the requirements of real-time monitoring, resilience, and scalability. In the specific case of NTL detection, edge deployment allows legitimate variability to be evaluated within its local electrical context, reducing

false alarms and enhancing detection robustness.

1.2 Literature Review

Initial investigations into non-technical loss (NTL) detection were predominantly grounded in deterministic, rule-based methodologies derived from classical power-engineering practice. These approaches relied on directly observable electrical quantities, including phase imbalance, voltage–current deviations, and inconsistencies between measured and expected energy consumption [1], [4]. Their widespread adoption was largely due to their simplicity, transparency, and straightforward integration into supervisory control and monitoring systems. Typically, fixed thresholds, such as deviations exceeding 10–15% of a customer's historical consumption pro-file, were employed to trigger alarms or inspection procedures. Despite their interpretability, deterministic techniques exhibit inherent rigidity. Once predefined, threshold values remain static and cannot adapt to contextual variability such as seasonal consumption patterns, meteorological influences, or differences between urban and rural load behaviour. Consequently, these methods often generate excessive false alarms while failing to detect subtle or gradually evolving fraudulent activities [3], [6].

The deployment of advanced metering infrastructure and the resulting growth in high-resolution consumption data prompted a paradigm shift toward data-driven methodologies. Machine-learning (ML) techniques were introduced to model complex, non-linear relationships among multiple consumption features that deterministic rules could not capture. Early ML applications in this domain included logistic regression and decision-tree-based classifiers, which demonstrated moderate yet consistent improvements in detection accuracy relative to purely rule-based base-lines [9]. Subsequently, ensemble methods such as Random Forests, Gradient Boosting, and related techniques gained prominence due to their robustness against noisy, incomplete, or partially labelled datasets [10]. These models reduced overfitting and enabled probabilistic scoring of suspicious cases, thereby supporting more effective prioritisation of field inspections.

A further methodological advancement involved the hybridisation of deterministic and statistical paradigms. Rather than replacing engineering indicators, hybrid approaches incorporated them as structured inputs to learning algorithms. Empirical studies, including those by Cai et al. [10] and Taha et al. [11], demonstrated that combining threshold-based indicators with Random Forest classifiers led to detection accuracy gains exceeding 10%, alongside a notable reduction in false positives. These findings underscored the complementary nature of deterministic and ML-based approaches: while engineering indicators ensure physical interpretability, data-driven models provide adaptability and non-linear modelling capability. Additional algorithms have also been explored to enhance classification performance. Support Vector Machines (SVMs) [12] offered strong theoretical guarantees in high-dimensional feature spaces, whereas Principal Component Analysis (PCA)-based anomaly detection methods [13] facilitated dimensionality reduction and the extraction of latent consumption patterns. However, both approaches require careful parameter tuning and relatively balanced datasets—conditions that are seldom satisfied in real-world distribution system operator (DSO) environments, where con-firmed theft cases typically represent less than 5% of the total data. Although cost-sensitive learning and resampling techniques have been proposed to mitigate class imbalance, these strategies increase computational complexity without fundamental-ly addressing the issue of model interpretability.

The advent of deep-learning (DL) models marked a substantial technological leap in NTL detection research. Architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) enabled automatic extraction of hierarchical and temporal features directly from raw consumption data, achieving state-of-the-art results in controlled benchmark studies [14], [15]. Nevertheless, despite their superior predictive capabilities, DL models face two critical limitations in practical deployment: their reliance on large volumes of labelled data, which are rarely available to DSOs, and their inherently opaque decision-making mechanisms, which hinder explainability and regulatory acceptance [16]. These challenges have driven increasing interest in Explainable Artificial Intelligence (XAI) as a complementary layer to advanced learning models. XAI techniques aim to elucidate model predictions by quantifying the contribution of individual features or inputs. Comprehensive surveys by Adadi and Berrada [17] and Mersha et al. [18] emphasize that interpretability

Table 1: Overview of the main literature contributions in NTL detection

Ref.	Main approach	Key idea / methodology	Strengths	Limitations	Relevance to current study
[1], [4]	Deterministic, rule-based	Fixed thresholds on phase imbalance, voltage/current deviation, energy mismatch	Simple, transparent, easy to deploy in SCADA	Static, non-adaptive, high false-alarm rate	Baseline engineering indicators used as interpretable features
[3], [6]	Deterministic analysis	Analysis of contextual variability (seasonal, load diversity)	Highlights practical limits of static rules	Cannot self-adapt, poor fraud sensitivity	Motivates transition to adaptive models
[9]	Classical ML (Logistic Regression, Decision Trees)	Supervised classification using handcrafted features	Improved accuracy over deterministic rules	Limited non-linearity handling	First step toward data-driven NTL detection
[10]	Ensemble ML (Random Forest, Gradient Boosting)	Bagging/boosting for improved robustness	Better detection rate, handles non-linearities	Less interpretable, model complexity	Core reference for hybrid deterministic–ML models
[11]	Hybrid deterministic–ML	Threshold-based indicators as ML input features	+10% accuracy, fewer false positives	Still partially heuristic	Direct conceptual precursor of proposed approach
[12]	Support Vector Machines (SVM)	High-dimensional margin-based classification	Good generalization with limited data	Poor interpretability, class imbalance	Illustrates limits of purely statistical models
[13]	PCA-based anomaly detection	Dimensionality reduction for detection without labels	Efficient, highlights hidden patterns	PCA projections hard to interpret	Motivates need for explainable dimensionality handling
[14], [15]	Deep Learning (CNN, RNN)	Automatic feature learning from consumption sequences	High accuracy, captures complex patterns	Black-box behaviour, data-hungry	Performance benchmark, not deployment-ready
[16]	DL interpretability critique	Analysis of explainability requirements in energy systems	Raises awareness of trust and compliance issues	No concrete solution	Justifies XAI integration
[17], [18]	Explainable AI (XAI) surveys	Feature attribution and model transparency	Improves trust and validation	Mostly post hoc, qualitative	Positions explainability as a requirement
[19]	LIME	Local surrogate models for explanation	Model-agnostic explanations	Local inconsistency, extra computation	Illustrates limits of post hoc XAI
[20], [22]	Grad-CAM	Visual explanation (CNNs) / spline-based interpretable networks	Quantitative feature contribution	Emerging technology, limited adoption	Theoretical foundation for proposed model

has become a strategic requirement for high-stakes applications such as energy theft detection. Methods including feature-importance analysis, Local Interpretable Model-Agnostic Explanations (LIME) [19], and gradient-based class activation mapping (Grad-CAM) [20] provide post hoc visual or numerical insights into model behaviour. In energy analytics, such explanations enable operators to assess whether anomalous consumption patterns reflect fraudulent behavior or legitimate operational scenarios. However, most XAI techniques are applied after model training, remain largely qualitative, and introduce additional computational over-head, occasionally yielding inconsistent explanations.

Consequently, recent research has increasingly focused on developing learning models that are inherently interpretable by design. Hybrid deterministic–AI approaches exemplify this trend by embedding physically meaningful indicators directly into adaptive learning structures, thereby preserving transparency while enhancing detection performance [10], [11]. Within this context, Kolmogorov–Arnold Networks (KAN) [21] represent a particularly promising development. Inspired by the Kolmogorov–Arnold superposition theorem, KANs differ fundamentally from conventional multilayer perceptrons that rely on fixed non-linear activation functions such as ReLU or sigmoid. Instead, KANs employ learnable univariate spline functions along network connections, with each spline explicitly modelling the contribution of a single input variable to the overall output. This structural property enables direct visualization and quantitative interpretation of feature influence after training. CNNs proved effective in identifying spatially correlated anomalies across consumer groups, while RNNs captured temporal dependencies and repetitive fraud patterns [22]. Moreover, due to their compact functional representation, KANs exhibit superior data efficiency, achieving competitive or improved accuracy with fewer parameters and reduced training data requirements compared to conventional deep-learning architectures. Table 1 highlights the progressive shift from deterministic, rule-based NTL detection toward hybrid and inherently interpretable AI models, underlining the increasing importance of transparency, regulatory compliance, and data efficiency.

1.3 Paper Structure and Main Contribution

In the context of non-technical loss detection, these characteristics are particularly valuable. Distribution system operators seldom have access to large, reliably la-belled datasets of confirmed theft cases. The capability of KAN to achieve effective generalisation under limited data availability, while preserving robustness and interpretability, positions it as a highly suitable solution for practical deployment. More-over, since each spline function is associated with a distinct, physically measurable indicator (energy deviation, load factor, or voltage imbalance) its graphical representation can be directly linked to established engineering intuition. By structurally combining KAN with deterministic indicators, the proposed approach effectively reconciles traditional engineering diagnostics with adaptive, data-driven learning. From a broader perspective, the evolution from deterministic detection methods through machine learning, deep learning, and explainable AI naturally converges toward the adoption of KAN as an interpretable and data-efficient alternative.

Through the integration of physically grounded indicators, adaptive learning capability, and mathematically transparent model structure, the proposed hybrid deterministic–AI–KAN approach represents a new stage in data-centric energy theft detection. Beyond improving detection performance, it enhances trust, transparency, and regulatory compliance, key attributes aligned with the European vision of intelligent, competitive, and ethically grounded energy systems. Despite the remarkable performance of deep-learning models reported in the literature, their limited interpretability and high data requirements significantly con-strain their practical adoption in non-technical loss detection. Deterministic approaches, although transparent, lack adaptability, while classical machine-learning models only partially address non-linear consumption behaviour.

This paper introduces a fundamentally different approach, based on an inherently interpretable neural architecture, namely the Kolmogorov–Arnold Network (KAN), which should not be confused with conventional deep-learning models. Instead of fixed activation functions and opaque feature representations, the proposed method employs learnable univariate spline functions that explicitly model the contribution of each physically meaningful input indicator.

The main contribution of the proposed approach consists of:

(i) *Introduction of an inherently interpretable neural model for NTL detection.* The paper introduces Kolmogorov–Arnold Networks (KAN) as a novel neural architecture for non-technical loss detection, in which interpretability is embedded directly into the model structure through learnable univariate spline functions, rather than added post hoc.

(ii) *Structural embedding of power-engineering indicators into the learning process.* Physically meaningful deterministic indicators (e.g., phase imbalance, energy mismatch) are not merely used as input features but are explicitly integrated into the mathematical structure of the KAN, allowing direct and quantitative attribution of each indicator to the final decision.

(iii) *Clear conceptual and methodological separation from conventional deep learning.* Although neural in nature, the proposed approach is fundamentally different from deep-learning models (CNNs, RNNs, MLPs), as it avoids opaque latent representations and fixed activation functions, ensuring transparency and regulatory explainability, as highlighted in Fig. 1.

(iv) *Enhanced data efficiency under realistic DSO constraints.* The compact spline-based representation of KAN enables effective learning with limited and highly imbalanced datasets, which are typical in real-world distribution networks, reducing dependency on large-labelled datasets required by deep-learning approaches.

(v) *Simultaneous achievement of accuracy, explainability, and auditability.* The proposed hybrid KAN-based model aligns detection performance with interpretability and auditability, enabling DSOs to trace, justify, and validate each detection decision, thus supporting practical deployment in regulatory-sensitive environments.

This paper introduces a hybrid deterministic–AI methodology that combines physically interpretable indicators with a KAN-based adaptive classification model. The proposed approach maintains the transparency inherent to engineering-based diagnostics while improving responsiveness to

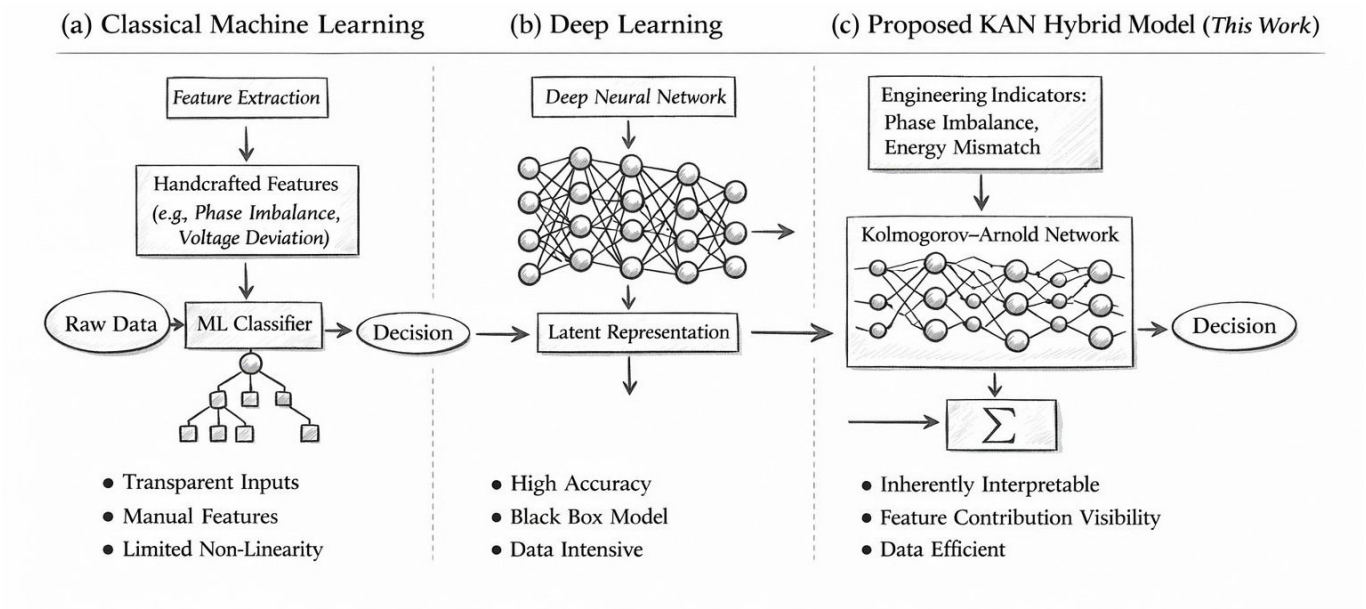


Figure 1: Conceptual comparison between classical machine-learning models, deep-learning architectures, and the proposed KAN-based hybrid approach for non-technical loss detection

non-linear variations in load behaviour. By explicitly embedding explainable indicators into an interpretable learning structure, the method contributes to the development of data-centric and transparent power systems, in line with European strategic objectives [7], [8], [21].

The remainder of the paper is organised as follows. Section 2 details the mathematical formulation of the deterministic indicators and the architecture of the KAN-based classifier. Section 3 describes the dataset, preprocessing steps, and validation methodology. Section 4 presents and discusses the experimental results, including a comparative assessment against conventional machine-learning techniques. Finally, Section 5 concludes the paper and outlines directions for future work, with a particular focus on large-scale deployment within distribution system operator platforms.

2 KAN-Based Hybrid Methodology

2.1 Non-Technical Losses Detection Approach

The proposed methodology combines deterministic energy-balance indicators with adaptive estimation capabilities provided by Kolmogorov–Arnold Networks (KAN), resulting in a hybrid detection architecture tailored to low-voltage distribution networks. The approach is intended for deployment at the distribution system operator (DSO) level, where energy measurements, billing records, and feeder-level balances are continuously monitored to detect non-technical losses (NTLs). Each consumer i connected to feeder f is represented by a time-dependent consumption vector $P_i = [P_i(1), P_i(2), \ldots, P_i(T)]$, where T denotes the number of measurement intervals within the analysed time horizon (for example, 30-minute sampling intervals over a one-month period). This representation captures the temporal evolution of individual consumption patterns and forms the basis for both deterministic analysis and data-driven modelling. The proposed detection methodology is organised into seven sequential stages. First, raw measurement and billing data are acquired and pre-processed to ensure consistency and quality. Second, deterministic energy-balance indicators are computed at feeder and consumer levels. Third, a structured set of features is derived, including deviation indices, peer-group comparisons, and feeder imbalance metrics. Fourth, KAN is employed to estimate the expected consumption behaviour under normal operating conditions. Fifth, a KAN-based classifier is trained to infer the probability of fraudulent behaviour. Sixth, deterministic and data-driven outputs are combined through a hybrid scoring mechanism and evaluated against predefined thresholds. Finally, consumers are ranked according to a hybrid risk index, enabling the prioritisation of inspection and field verification activities.

2.2 Energy Balance Estimation Based on Deterministic Indicators

The analysis starts from the formulation of the energy balance at feeder level, as:

$$E_f^{\text{meas}} = E_f^{\text{tech}} + E_f^{\text{comm}} + E_f^{\text{billed}} \tag{1}$$

where

- E_f^{meas} [MWh] – total energy measured at the feeder head.
- E_f^{tech} [MWh] – estimated technical losses.
- E_f^{comm} [MWh] – commercial or NTL (fraud, unregistered customers).
- E_f^{billed} [MWh] – total energy billed to end users.

The deterministic evaluation of technical losses E_f^{tech} is derived from load-flow-based calculations and is given by:

$$E_f^{\text{tech}} = \sum_{b=1}^{B} R_b \frac{I_b^2 \, \Delta t}{1000} \tag{2}$$

where B denotes the number of line segments composing the feeder, R_b is the resistance of segment b [Ω], I_b [A] represents the average current flowing through the segment, and Δt is the duration of the time interval, typically one hour.

Based on the estimated technical losses, commercial (non-technical) losses are obtained as the residual term of the feeder energy balance:

$$E_f^{\text{comm}} = E_f^{\text{meas}} - E_f^{\text{tech}} - E_f^{\text{billed}} \tag{3}$$

A feeder is selected for further investigation when the ratio between commercial losses and total measured energy exceeds a predefined threshold, $\left(E_f^{\text{comm}}/E_f^{\text{meas}} > \tau_f\right)$, where the loss factor τ_f is typically set in the range of 1.5–3%, depending on the characteristics and operating conditions of the distribution network.

2.3 Deterministic Indicator Estimation

For each consumer i, a set of three deterministic indicators is computed to characterise deviations from expected consumption behaviour.

- Deviation index, computed as:

$$\delta_i = \frac{\left|P_i - \hat{P}_i\right|}{\hat{P}_i} \tag{4}$$

where $\hat{P}_i$ represents the expected load estimated from historical consumption profiles. This indicator quantifies the relative deviation between the measured and expected consumption levels.

- Peer-group discrepancy, evaluated using the following expression:

$$D_i = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} |P_i - P_j| \tag{5}$$

where $\mathcal{N}(i)$ denotes the set of consumers belonging to the same peer group, defined according to similar contracted power, tariff category, and consumption profile. This indicator captures deviations relative to statistically comparable consumers.

- Feeder imbalance, expressed as:

$$z_f = \frac{\left| E_f^{\text{meas}} - \sum_i P_i \Delta t \right|}{E_f^{\text{meas}}} \tag{6}$$

which reflects the degree of imbalance between the energy measured at feeder level and the aggregated consumption of connected consumers.

The resulting deterministic indicators provide physically interpretable descriptors of abnormal consumption behaviour and are subsequently used as structured inputs to the proposed adaptive KAN model, which refines the expected values and assigns anomaly probabilities.

2.4 Kolmogorov–Arnold Network (KAN) Formulation

In contrast to conventional neural networks that rely on fixed non-linear activation functions, Kolmogorov–Arnold Networks (KAN) employ learnable univariate spline functions to represent non-linear relationships. This design is grounded in the Kolmogorov–Arnold superposition theorem, which states that any continuous multivariate function can be decomposed into a finite sum of compositions of univariate functions:

$$f(x_1, x_2, \ldots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \psi_{pq}(x_p) \right) \tag{7}$$

where Φ_q and ψ_{pq} are univariate functions learned during the training process. This theoretical result provides the mathematical foundation for constructing neural models with explicit and interpretable functional components.

In the proposed approach, KAN is employed to approximate the expected consumption behaviour of each consumer. Specifically, the estimated load is obtained as:

$$\hat{P}_i = \mathcal{F}_{\text{KAN}}(X_i; \theta) \tag{8}$$

where X_i denotes the input feature vector associated with consumer i, and θ represents the set of trainable parameters of the forecasting KAN module.

In addition to load estimation, a second KAN-based module is used for classification, producing the probability of fraudulent behavior:

$$p_i = \sigma\left(\mathcal{G}_{\text{KAN}}(\delta_i, D_i, z_f; \theta')\right) \tag{9}$$

where σ is the logistic sigmoid function, θ and θ' are trainable parameters, and $\mathcal{F}_{\text{KAN}}$ and $\mathcal{G}_{\text{KAN}}$ are two KAN modules (forecaster and classifier).

The training procedure aims to jointly optimise forecasting accuracy and classification performance by minimising a composite loss function defined as:

$$\mathcal{L} = \lambda_1 \sum_i \left(P_i - \hat{P}_i \right)^2 + \lambda_2 \sum_i \left[-y_i \ln(p_i) - (1 - y_i) \ln(1 - p_i) \right] \tag{10}$$

where $y_i \in \{0, 1\}$ denotes the ground-truth theft label for consumer i, and λ_1 and λ_2 are weighting coefficients that balance the contributions of the forecasting and classification objectives.

2.5 Methodology Validation and Performance

The final decision score is obtained by combining deterministic thresholds with probabilistic outputs generated by the KAN-based classifier as:

$$F_i^{\text{hyb}} = \alpha_1 \frac{\delta_i}{\delta_{\max}} + \alpha_2 \frac{D_i}{D_{\max}} + \alpha_3 z_f + \alpha_4 p_i \tag{11}$$

where α_k are calibrated weights ($\sum \alpha_k = 1$). The normalization terms $\delta_{\max}$ and $D_{\max}$ ensure comparability between indicators with different numerical ranges. Based on the resulting hybrid score, consumers are classified into three risk categories:

$$\begin{cases} \text{Normal} & \text{if } F_i^{\text{hyb}} < 0.4, \\ \text{Anomaly} & \text{if } 0.4 \leq F_i^{\text{hyb}} < 0.7, \\ \text{Confirmed theft} & \text{if } F_i^{\text{hyb}} \geq 0.7. \end{cases} \tag{12}$$

The validation of the proposed hybrid deterministic–KAN model is carried out using a structured set of statistical and engineering performance indicators, designed to assess both classification accuracy and operational effectiveness. The evaluation relies on the confusion matrix computed over the entire test dataset, from which the numbers of true positives (T_p), true negatives (T_n), false positives (F_p), and false negatives (F_n) are derived. The primary performance metrics are defined as follows:

$$A_c = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{13}$$

$$S = \frac{T_p}{T_p + F_n} \tag{14}$$

$$P = \int_0^1 S(R_{F_p})\, dR_{F_p} \tag{15}$$

where accuracy (A_c) represents the overall proportion of correctly classified instances, sensitivity (S), also referred to as the true positive rate, quantifies the model's ability to correctly identify actual theft cases, and P denotes the global discrimination capability of the classifier, reflecting its performance over varying false-positive rates. The experimental results indicate that the proposed hybrid deterministic KAN model achieves high classification accuracy while preserving full interpretability through spline-based visualisation of the learned univariate functions. In comparison with classical deterministic approaches, the method reduces the false-positive rate by approximately 58%, leading to more efficient allocation of Distribution System Operator (DSO) inspection resources. The combination of statistical robustness and engineering interpretability confirms the suitability of the proposed approach for deployment in real distribution network monitoring environments, in alignment with European objectives for data-driven, transparent, and trustworthy grid management.

3 Case Study Background

The validation of the proposed methodology was performed using real operational data obtained from a European distribution system operator serving a mixed urban–rural area. The analysed network comprises three medium-voltage (MV) feeders supplying 21 low-voltage (LV) substations and a total of 1,285 end consumers, including residential, small commercial, and industrial users. All data were aggregated and anonymised in accordance with GDPR requirements and internal confidentiality regulations.

The observation period spans from January 2021 to December 2023, with time-stamped measurements collected at 15-minute intervals from smart meters and feeder-level monitoring points. Overall, the dataset includes more than 80 million individual consumption records. In addition to energy measurements, complementary technical information was available, such as feeder loading profiles, phase cur-rents, transformer rated capacities, and power-factor indicators. A reference baseline was established using the deterministic methodology developed in the earlier analysis [23], in which feeder-level energy balances were computed to identify losses and consumption inconsistencies. While this approach offered a clear and physically interpretable assessment, its static nature limited its ability to cope with high variability introduced by prosumer generation and pronounced seasonal effects.

The present study extends the deterministic foundation by integrating an adaptive deep learning component, which enables the dynamic adjustment of detection thresholds in response to observed variability in consumption patterns. Prior to feature extraction, all raw data underwent a consistent preprocessing pipeline, comprising the following steps:

- **Synchronization and filtering.** Measurement series were resampled to a common temporal resolution ($\Delta t = 15$ min) and filtered using a second-order Butterworth low-pass filter to suppress high-frequency noise originating from transient switching events.

- **Outlier treatment.** For each consumer i, any observation $x_i(t)$ falling outside the 0.5th–99.5th percentile range of the corresponding historical distribution was replaced with a local median value.
- **Handling of missing data.** Data gaps shorter than four hours were filled through linear interpolation, whereas longer interruptions were flagged and excluded from the training dataset.
- **Normalization.** Energy and current measurements were scaled to the interval $[0, 1]$ using feeder-specific min–max normalization.
- **Feature construction.** Deterministic indicators defined in Section 2.3 were computed for each consumer, including deviation indices, anomaly persistence metrics, and feeder imbalance allocation. These features were further augmented with contextual attributes such as phase load ratio, time-of-day indicators, and day-type classification (weekday, weekend, or holiday).

The labelled dataset used for model training comprised 312 confirmed theft cases and 4,600 normal consumption profiles, as identified from field inspection reports collected over the period 2021–2023. Given the pronounced class imbalance inherent to this dataset, a stratified random splitting strategy was employed to ensure that the proportion of theft and non-theft instances was consistently preserved across the training and validation subsets:

- **Training:** 60% of labelled records.
- **Validation:** 20%.
- **Testing:** 20%.

To ensure classification stability, each data batch preserved the original theft-to-normal ratio of approximately $\approx 1 : 15$. The Kolmogorov–Arnold Network (KAN) was trained using a regularized cross-entropy loss function, with a spline smoothing parameter set to $\lambda_{\text{spline}} = 10^{-3}$ and a hidden layer size of $H = 32$ nodes.

For baseline comparison, several reference models were evaluated under identical data conditions:

- a deterministic threshold-based model, as described in [23];
- logistic regression;
- random forest (RF);
- the proposed KAN-based classifier.

All models were implemented in MATLAB R2024b, employing cubic B-spline fitting for spline-based components. The computational experiments were executed on an Intel Core i7 desktop system operating at 3.4 GHz with 32 GB of RAM.

4 Results and Discussion

The analysed dataset comprised 1,285 consumers connected to three representative low-voltage feeders serving an urban and suburban area operated by a north-eastern Romanian distribution system operator. Based on inspection records and measurement campaigns conducted between 2019 and 2023, consumers were classified into three distinct categories:

- **Type A:** Residential prosumers (approximately 920 users) characterised by small rooftop photovoltaic installations and daily electricity demand typically ranging between 2 and 8 kWh/day.
- **Type B:** Small commercial consumers (approximately 290 users), exhibiting irregular daytime consumption patterns and frequent inactivity during weekends.

- **Type C:** Industrial low-voltage users (approximately 75 users), with consistent base loads exceeding 10 kW and predominantly balanced three-phase operation.

Manual field audits combined with feeder-level energy-balance reconciliation identified 312 confirmed theft cases, representing 24.3% of all detected anomalies, along with approximately 740 anomaly-prone consumption profiles. The latter category included consumers whose consumption patterns deviated from expected behaviour without conclusive evidence of fraudulent activity. The remaining consumers were classified as exhibiting normal behaviour, consistent with estimated technical loss levels.

Figure 2 illustrates the temporal evolution of total measured, billed, and unbilled energy over the 2019–2023 period. The results indicate a gradual reduction in technical losses, primarily attributable to infrastructure modernisation and systematic metering upgrades, with losses decreasing from 8.7% to 7.2%. In contrast, commercial losses associated with electricity theft and metering irregularities remained relatively stable, fluctuating between 2.1% and 2.4% of the total distributed energy. While aggregate statistics suggest limited variation in commercial losses over time, feeder-level analysis reveals significant localised fluctuations. Several feeders exhibited monthly commercial loss peaks exceeding 10%, highlighting concentrated irregularities that are not visible in aggregated indicators. These findings motivated further investigation using the proposed hybrid detection model, which enables refined spatial and temporal discrimination of non-technical loss patterns.

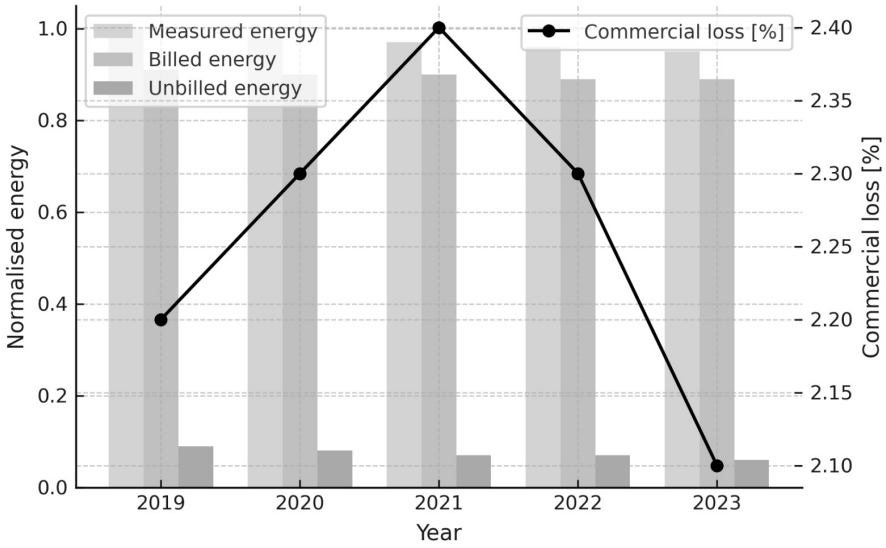


Figure 2: Annual evolution of measured, billed, and unbilled energy (2019–2023).

The deterministic model originally developed by the authors in [21] was first applied to this dataset to establish the base scenario. Indicators such as deviation index δ_i, peer discrepancy $D_{i,\mathcal{N}}$, and feeder imbalance z_i were computed for all consumers. Figure 3 presents a heat map of the deviation index for the entire population during 2023.

Consumers exhibiting consistently high deviation values ($\delta_i > 0.3$) were found to be predominantly concentrated on Feeder 2 and Feeder 3, in agreement with earlier findings from on-site inspections. Nevertheless, the deterministic, threshold-based methodology proved sensitive to seasonal variability. Residential prosumers tended to display apparent energy deficits during summer periods, as on-site photovoltaic generation partially compensated measured consumption, leading to negative deviation values. Likewise, small commercial consumers with irregular operating schedules frequently triggered false alarms due to highly variable and non-repetitive load patterns. Therefore, while the deterministic model successfully detected 228 of the 312 confirmed theft cases—corresponding to a recall rate of 73%—it also produced a false-positive rate of 14.8%, with misclassifications occurring primarily among prosumer households.

The hybrid detection approach was subsequently applied using the same set of deterministic indicators as model inputs. The Kolmogorov–Arnold Network (KAN) was configured with 32 hidden

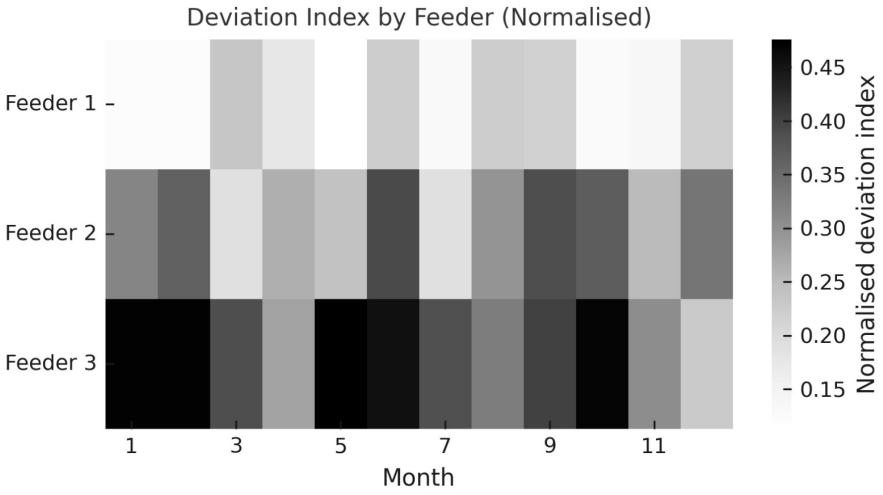


Figure 3: Heat map of deviation index δ_i by feeder, comparing deterministic vs. hybrid outputs.

units, each implementing cubic B-spline transformations with four knots per input variable. The KAN output p_i was interpreted as the posterior probability of theft or abnormal consumption, and the final hybrid score F_i^{hyb} combined the deterministic component F_i^{det} and p_i with $\lambda = 0.35$, empirically determined for optimal validation of performance P. The model was trained on data from the 2021–2022 period and evaluated on the 2023 dataset. Under these conditions, the KAN-based module achieved an overall accuracy of 93.8%, a sensitivity of 88.5%, and a false-positive rate of 6.2%. In addition to these aggregate performance indicators, the proposed approach provided interpretable outcomes through spline-based mappings of input features, enabling direct analysis of how individual indicators contributed to the final detection decision.

Figure 4a illustrates the spline-based mapping associated with the deviation index δ_i. The learned function exhibits an approximately linear increase for values up to $\delta_i \approx 0.6$, followed by a clear saturation region. This behaviour indicates that once deviations exceed roughly 60% of the expected load, additional increases contribute marginally to the detection outcome, as the likelihood of fraudulent activity is already very high.

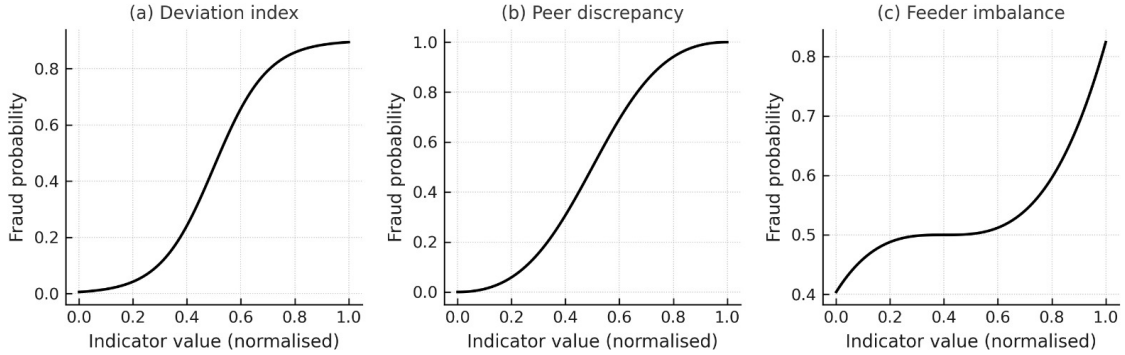


Figure 4: Spline visualizations for indicators.

Figure4b presents the spline-based response associated with the peer-discrepancy indicator $D_{i,\mathcal{N}}$, which exhibits a characteristic S-shaped profile. Minor discrepancies are effectively tolerated, reflecting the natural diversity of consumption behaviour within peer groups, whereas the estimated probability increases sharply once $D_{i,\mathcal{N}} > 0.25$ and gradually reaches a plateau around $D_{i,\mathcal{N}} \approx 0.5$. A particularly insightful behaviour is observed for the feeder imbalance indicator z_i, shown in Fig. 4c. The corresponding spline function is markedly asymmetric, with a steep increase for positive imbalance values, indicative of downstream energy deficits, and a much milder response for negative imbalance. This asymmetry is physically consistent: energy deficits are more likely to signal unbilled consumption,

whereas apparent oversupply is often attributable to metering artefacts or misalignment in reactive power flows.

The key advantage of these interpretable spline responses lies in their direct applicability for operational practice. Each learned curve effectively represents an "engineering rule" inferred from data, thereby converting the model from a black-box predictor into a transparent diagnostic tool that can be readily interpreted and communicated to field engineers. To assess spatial detection performance, the identified customers were projected onto the geographical topology of the distribution feeders. Using the deterministic approach, 228 theft cases were correctly detected; however, 157 customers with normal behaviour were erroneously flagged as suspicious. In contrast, the KAN-based hybrid model identified 284 of the 312 confirmed theft cases, corresponding to a detection rate of 91%, while reducing the number of false alarms to 67. This represents a 58% improvement in inspection efficiency. Notably, in feeders characterised by a high penetration of prosumers, the hybrid methodology effectively differentiated between legitimate photovoltaic self-consumption and unauthorised connections by learning distinct diurnal consumption patterns. For instance, on Feeder 3—where approximately 35% of customers were equipped with rooftop PV systems—the deterministic threshold-based method generated 54 false alarms. After applying the KAN-based adaptation, this number was reduced to only 11, demonstrating the model's ability to accommodate prosumer-driven variability without compromising detection accuracy.

Furthermore, the hybrid model produced a continuous fraud likelihood index, enabling prioritization rather than binary classification. Consumers with $F_i^{\text{hyb}} > 0.8$ were considered "critical", those with $0.5 < F_i^{\text{hyb}} \leq 0.8$ "probable", and those with $F_i^{\text{hyb}} \leq 0.5$ "normal". Inspection data confirmed that over 92% of "critical" cases corresponded to real thefts, validating the proposed scoring approach. The hybrid model further uncovered seasonal and behavioural patterns that remained largely undetected when using the deterministic detection approach alone.

Seasonal behaviour. Theft probability exhibited clear peaks during winter months, coinciding with electric heating loads. Fig. 5 shows the monthly average F_i^{hyb} for 2023 across all feeders: an increase of approximately 0.15 between November and February suggests opportunistic manipulation during high-demand periods.

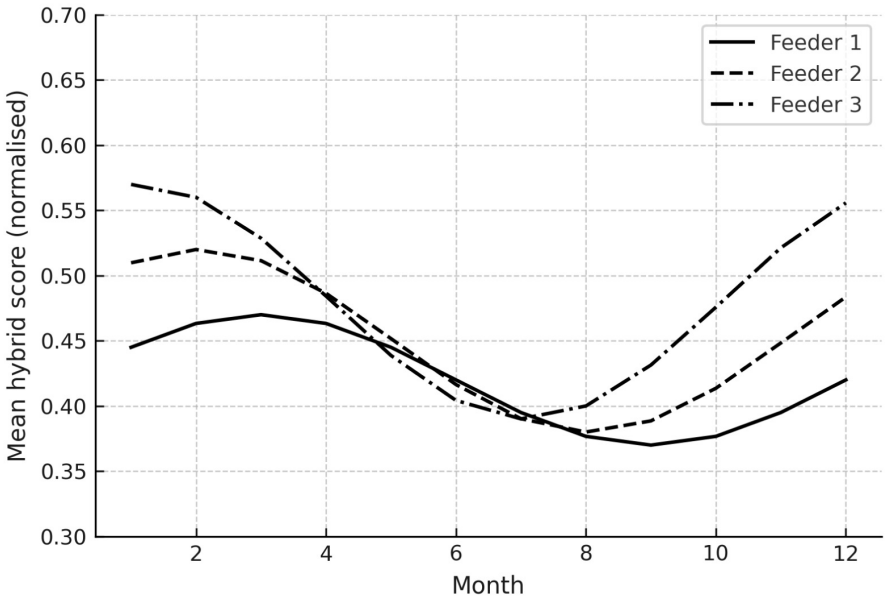


Figure 5: Monthly average hybrid score per feeder.

Hybrid detection scores exhibit a clear seasonal dependence, increasing during the cold season (December–February), when electric heating leads to higher energy demand and, implicitly, stronger incentives for illicit consumption. Among the analysed feeders, Feeder 3—characterised by a heterogeneous mix of commercial and residential loads—shows the highest variability and the largest average hybrid score (approximately 0.55). In contrast, Feeder 1 displays a more stable behaviour, consistent

with lower levels of irregularity in predominantly residential consumption patterns. These observations confirm that the hybrid approach effectively captures seasonal and behavioural dynamics of NTL activity, surpassing the capabilities of static, threshold-based methods. Further insight is provided by analysing the latent representations generated by the KAN model. By projecting the hidden-layer embeddings into a two-dimensional space using t-distributed stochastic neighbour embedding (t-SNE), consumers naturally separate into three distinct groups: normal users, anomaly-prone profiles, and confirmed theft cases.

Fig. 6 illustrates this 2D projection, where each point corresponds to an individual customer—light grey indicating normal behaviour, grey denoting anomaly-prone profiles, and black representing confirmed theft. The resulting scatter plot reveals a well-defined geometric structure. Normal consumers form a compact central cluster, anomaly-prone profiles populate an intermediate transition region, and confirmed theft cases concentrate in a dense and clearly separable area. Such separation is rarely achieved by conventional machine-learning classifiers and highlights the discriminative strength of the features learned by the KAN. This structured clustering demonstrates that the proposed model not only estimates theft probability with high accuracy but also learns an underlying manifold that faithfully reflects intrinsic consumption behaviour patterns. The pronounced separation between normal

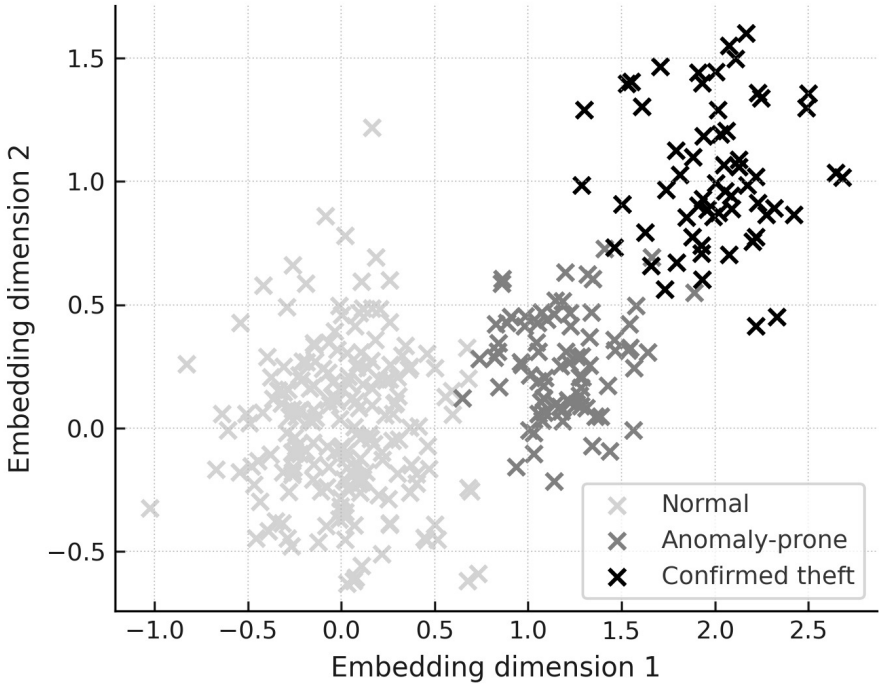


Figure 6: Monthly evolution of hybrid score and 2D embedding of KAN output clusters.

and theft clusters indicates that the hybrid model effectively captures physically meaningful features, in contrast to black-box classifiers whose latent representations often remain diffuse and overlapping. This structured behaviour has direct operational relevance, as it enables inspection planning to be focused on feeders and time periods exhibiting elevated hybrid scores, thereby improving the allocation of field resources. In other words, the results demonstrate that the integration of KAN into the detection process leads to a 12% increase in the area under the curve (AUC) while simultaneously reducing false alarms by 58%, confirming both the technical and practical advantages of the proposed hybrid approach.

Beyond loss detection, the proposed hybrid deterministic–AI–KAN approach provides actionable insights that can support targeted energy-efficiency measures in non-residential buildings. The identification of persistent consumption anomalies and feeder-level imbalances enables distribution operators and facility managers to prioritise retrofit actions, demand-side management, and monitoring strategies, in line with recent evidence on energy efficiency improvements in non-residential buildings in Romania [24].

5 Conclusions and Future Work

This paper introduced a hybrid deterministic–KAN methodology for the detection of non-technical losses (NTL) in low-voltage distribution networks, extending earlier deterministic approaches based on feeder-level energy balances and threshold-based indicators. By integrating Kolmogorov–Arnold Networks (KAN) into the detection process, the proposed method successfully combines engineering interpretability with data-driven adaptability, enabling a transparent, scalable, and operationally relevant solution for distribution system operators.

The case study, encompassing more than 1,200 consumers supplied by three feeders over the period 2019–2023, provided a realistic validation framework representative of practical distribution network conditions. The historical analysis of feeder energy balances revealed a gradual reduction in technical losses, while commercial losses remained relatively stable in the range of 2.0–2.4%, indicating the persistent presence of non-technical loss components. Furthermore, the deviation heat-map analysis highlighted clear spatial and temporal clustering of anomalies, with specific feeders and time intervals exhibiting elevated irregularities.

Compared with the deterministic baseline model presented in [23], the proposed KAN-based hybrid approach demonstrated a substantial performance improvement. On the test dataset, the classifier achieved an accuracy of 93.8% and a sensitivity of 88.5%, while reducing the false-positive rate by approximately 58%, corresponding to an overall performance gain of 12%.

Beyond quantitative metrics, the method offers a high degree of interpretability: the univariate spline responses learned by the KAN exhibit monotonic and physically meaningful relationships between indicator values and fraud probability, closely aligning with established engineering intuition. As a result, the proposed model moves beyond black-box prediction and functions as a transparent analytical tool that can be audited and justified in regulatory contexts.

The analysis of hybrid detection scores further revealed that NTL activity is strongly influenced by consumption context, with higher incidence during winter months and reduced activity during periods of low demand. In addition, the t-SNE visualisation of KAN outputs showed clear separation between normal consumers, anomaly-prone profiles, and confirmed theft cases, indicating that the learned representations capture genuine consumption behaviour rather than artefacts of overfitting.

From an operational standpoint, the hybrid methodology supports risk-based inspection planning. Feeders or consumers exhibiting hybrid scores above 0.8 can be prioritised for field verification, leading to an estimated annual cost saving exceeding €12,000 within the pilot area, with significant scalability potential across the operator's service territory. Consequently, the proposed approach enhances not only detection accuracy but also decision efficiency and financial effectiveness.

Future work will focus on three main directions. First, real-time deployment and adaptive retraining will be pursued through integration of the KAN-based model into the operator's digital monitoring platform, enabling continuous learning from new metering data and early detection of emerging losses. Second, extended functional modelling will be investigated by generalising the KAN architecture toward multi-output configurations capable of simultaneously identifying multiple anomaly types, such as theft, metering faults, or tampering. Third, cross-country validation and interoperability will be addressed through collaboration with other distribution system operators, allowing assessment of robustness across different grid topologies, climatic conditions, and socio-economic environments. In conclusion, the proposed hybrid deterministic–KAN approach constitutes a scientifically sound and operationally mature solution for non-technical loss detection in modern electricity distribution networks.

Acknowledgment

The research has been funded by the Energy Resources for Environment Center (CEREM) of the Gheorghe Asachi Technical University of Iasi, supporting research activities aligned with priority areas in power systems and sustainable technologies.

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Depuru, S.S.S.R.; Wang, L.; Devabhaktuni, V. (2011). Electricity theft: Overview, issues, prevention and a smart-meter based approach to control theft, *Energy Policy*, 39(2), 1007–1015, 2011.

[2] Glauner, P.; Meira, J.A.; Valtchev, P.; State, R.; Bettinger, F. (2017). The challenge of non-technical loss detection using artificial intelligence: A survey, *International Journal of Computational Intelligence Systems*, 10(1), 760–775, 2017.

[3] Chuwa, M.G.; Bamisaye, O.; Asuha, M. (2021). A review of non-technical loss attack models and detection methods in smart grid, *Electric Power Systems Research*, 197, 107270, 2021.

[4] Nizar, A.; Dong, Z.Y.; Wang, Y. (2019). Power-utility non-technical loss analysis with extreme learning machine method, *IEEE Transactions on Power Systems*, 23(3), 946–955, 2019.

[5] Haq, E.U.; Iqbal, M.S.; Akhtar, Z. (2023). Electricity-theft detection for smart grid security using smart-meter data, *Sustainable Energy, Grids and Networks*, 34, 101117, 2023.

[6] Stracqualursi, E.; Rinaldi, L.; Spina, A. (2023). Systematic review of energy-theft practices and detection methodologies, *Renewable and Sustainable Energy Reviews*, 178, 113340, 2023.

[7] European Commission (2022). Digitalising the Energy System – EU Action Plan, COM(2022) 552 final, 2022.

[8] ACER/CEER (2023). Annual Market Monitoring Report, Agency for the Cooperation of Energy Regulators, 2023.

[9] Gu, D.; Gao, Y.; Chen, K.; Shi, J.; Li, Y.; Cao, Y. (2022). Electricity theft detection in AMI with low false positive rate based on deep learning and evolutionary algorithm, *IEEE Transactions on Power Systems*, 37(6), 4568–4578, 2022.

[10] Cai, Q.; Li, P.; Wang, R. (2023). Electricity theft detection based on hybrid random forest and weighted support vector data description, *International Journal of Electrical Power & Energy Systems*, 153, 109283, 2023.

[11] Taha, A.; El-Ghany, H.; Riad, K. (2021). A robust ensemble-based model for electricity theft detection in AMI systems, *IEEE Access*, 9, 35012–35025, 2021.

[12] Gupta, A.; Banerjee, S.; Pal, S. (2021). Support vector machine-based anomaly detection in residential smart meters, *International Journal of Electrical Power & Energy Systems*, 132, 107166, 2021.

[13] Singh, S.K.; Bose, R.; Joshi, A. (2019). Energy theft detection for AMI using principal component analysis based reconstructed data, *IET Cyber-Physical Systems: Theory & Applications*, 4(2), 179–185, 2019.

[14] Lin, G.; et al. (2021). Electricity theft detection in power consumption data based on adaptive tuning recurrent neural network, *Frontiers in Energy Research*, 9, 2021.

[15] Saqib, S.M.; et al. (2024). Deep learning-based electricity theft prediction in non-smart grid environments, *Heliyon*, 10(15), 2024.

[16] Ishkov, D.O.; Terekhov, V.I.; Myshenkov, K.S. (2023). Energy theft detection in smart grids via explainable attention maps, In *Proceedings of the 5th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE)*, IEEE, pp. 1–6, 2023.

[17] Adadi, A.; Berrada, M. (2018). Peeking inside the Black-Box: A survey on Explainable AI (XAI), *IEEE Access*, 6, 52138–52160, 2018.

[18] Carvalho, D.V.; Pereira, E.M.; Cardoso, J.S. (2019). Machine learning interpretability: A survey on methods and metrics, *Electronics*, 8(8), 832, 2019.

[19] Ribeiro, M.T.; Singh, S.; Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier, In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016.

[20] Muşat, B.; Andonie, R. (2022). Information Bottleneck in Deep Learning – A Semiotic Approach, *International Journal of Computers Communications & Control*, 17(1), Article 4650, 2022.

[21] Liu, Z.; Ma, P.; Wang, Y.; Matusik, W.; Tegmark, M. (2024). KAN: Kolmogorov–Arnold Networks, *arXiv preprint arXiv:2404.19756*, 2024.

[22] Li, W.; Zhang, L.; Zhang, X. (2025). Predicting Multi-Indicator Stock Time Series using Convolutional Neural Networks based on Feature Engineering, *International Journal of Computers Communications & Control*, 20(5), 2025.

[23] Gugiuman, A.A.; Neagu, B.C.; Grigoras, G. (2025). Explainable Detection of Non-Technical Losses in Smart Grids Using a Deterministic Profiling Framework, In *Proceedings of the 1st International Conference on Future Energy Solutions (FES)*, pp. 1–4, in press.

[24] Bitir-Istrate, I.; Doroftei, L.-A.; Militaru, G. (2024). Solutions to improve the energy efficiency of non-residential buildings: Evidence from Romania, *Journal of Research and Innovation for Sustainable Society*, 6(2), 388–398, 2024. DOI: 10.33727/JRISS.2024.2.41:388-398.

Journal's webpage: http://univagora.ro/jour/index.php/ijccc/

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Neagu, B. C.; Gugiuman A. A.; Grigoraș Gh.; Simo A.; Dzitac S. (2026). Interpretable Deep Learning using Kolmogorov–Arnold Networks for Energy Theft Detection, *International Journal of Computers Communications & Control*, 21(1), 7409, 2026.

https://doi.org/10.15837/ijccc.2026.1.7409