

Multiclass Non-Technical Loss Detection with Reduced Synthetic Data by Two-Stage Hierarchical 1D-CNN

Katsarin Seepromting¹, Nattaya Rajitroj², Jakub Michel³, Arun Onlam⁴,
Chayada Surawanitkun^{1†}, and Apirat Siritaratiwat⁵, Non-members

ABSTRACT

Non-Technical Loss (NTL) detection is a significant issue in power distribution, particularly with conventional mechanical meters. This paper introduces a two-stage hierarchical One-Dimensional Convolutional Neural Network (1D-CNN) framework for multiclass NTL classification using data from the Provincial Electricity Authority (PEA) in Khon Kaen, Thailand. The first stage classifies data into three classes: normal, energy theft, and defective meters, achieving 100% recall for defective meters. The second stage uses the Synthetic Minority Over-sampling Technique (SMOTE) to enhance the detection of energy theft, reducing synthetic data by 50% while maintaining a 99.95% recall rate for theft. Compared to baselines like SVM and Random Forest, the proposed method streamlines operations, enabling direct repair dispatch for defective meters and investigations for theft. This approach minimizes biases and overhead, offering a practical solution for NTL management in resource-limited settings..

Keywords: Non-Technical Loss (NTL), Multiclass Classification, 1D-CNN, Imbalanced Data, SMOTE, Conventional Meter, Energy Theft, Defective Meter, Power Distribution

1. INTRODUCTION

The Non-Technical Loss (NTL) in power distribution systems refers to energy delivered but not billed due to

factors unrelated to technical inefficiencies [1], primarily energy theft and defective metering equipment. Unlike Technical Losses (TL) from transmission or equipment heating [2], NTL arises from deliberate customer actions such as meter tampering, illegal connections, bypass circuits, or hardware malfunctions that cause inaccurate readings [3]. These losses impose significant financial burdens on utilities through revenue leakage [4], distort load forecasting, and necessitate costly field inspections [5]. Globally, NTL accounts for 10-40% of total distribution losses [6-8], with developing economies experiencing the highest impact [9].

In Thailand, NTL represents a critical operational challenge for the Provincial Electricity Authority (PEA) [10], which manages distribution across 74 provinces [11]. Official reports indicate annual NTL rates exceeding 6-8% of distributed energy in several regions, translating to billions of baht in lost revenue [12]. Among PEA's service areas, Khon Kaen province records the highest NTL incidence, driven by its status as the fastest-growing urban center in the northeastern region [13]. Rapid population influx, expanding commercial activity, and mixed residential-commercial load profiles create fertile conditions for both organized theft schemes and undetected meter degradation [14]. This makes Khon Kaen an ideal and urgent case study for advanced NTL detection.

Most prior NTL detection research leverages high-frequency smart meter data with sub-hourly or daily resolution and two-way communication capabilities [15-17]. However, in Thailand, over 90% of meters remain conventional mechanical units or digital meters with one-way communication that require manual monthly readings by field operators [12]. This fundamental infrastructure gap results in a much coarser temporal resolution of 36 monthly timesteps, compared to thousands of daily or hourly points in smart meter systems; eliminates real-time monitoring and remote diagnostics; increases susceptibility to human error in recording; and produces consumption patterns shaped primarily by billing cycles rather than continuous usage [18]. These constraints render many smart-meter-centric models inapplicable and necessitate algorithms that are robust to sparse, operator-recorded time series [19].

Advanced computational intelligence and meta heuristic optimization algorithms [20] have increasingly been

Manuscript received on December 28, 2025; revised on February 4, 2026; accepted on February 16, 2026. This paper was recommended by Associate Editor Siraporn Sakphrom.

¹The authors are with Center of Multidisciplinary Innovation Network Talent (MINT Center), Faculty of Interdisciplinary Studies, Khon Kaen University, Nong Khai Campus, Nong Khai 43000, Thailand.

²The author is with Department of Electrical Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand.

³The author is with Department of Mathematical Sciences, Webb School of Engineering, High Point University, High Point, NC, USA.

⁴The author is with Merter and Transformer Section, Provincial Electricity Authority, Khon Kaen, Thailand.

⁵The author is with Department of IoT and Information Engineering, Faculty of Engineering, King Mongkut's Institute of Technology, Ladkrabang, Bangkok, 10520, Thailand.

[†]Corresponding author: chayasu@kku.ac.th

©2026 Author(s). This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License. To view a copy of this license visit: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Digital Object Identifier: 10.37936/ecti-ec.2026242.263037

deployed to solve complex operational challenges in power systems. In the specific domain of NTL detection, extensive prior research has applied machine learning and deep learning to NTL detection, including classical classifiers such as Support Vector Machine (SVM), Random Forest (RF) [21-22] and modern architectures including Long Short-Term Memory (LSTM), One-Dimensional Convolutional Neural Network (CNN), and hybrid models [23-24]. Despite this progress, severe class imbalance remains a fundamental barrier: normal consumption dominates (>95%) [25], while theft and defective cases are rare (<5%) [26]. Training on such skewed distributions leads to majority-class bias, where models achieve high overall accuracy by trivially predicting “normal,” yet fail to detect critical anomalies. To address class imbalance, many studies employ synthetic data generation, with Synthetic Minority Over-sampling Technique (SMOTE) being the dominant approach [27-29]. While effective in controlled experiments, over-reliance on synthetic samples introduces artifact bias, overfitting to generated patterns, and poor generalization when real-world anomalies deviate from interpolated templates. Moreover, the prevalent binary classification (normal vs. abnormal) in the literature fails to distinguish theft from defects, forcing manual triage and leading to misaligned operational responses: legal action for theft versus technical repair for faults.

2. PROBLEM STATEMENT

The theoretical backgrounds related to this research that should be reviewed beforehand for more understanding are the solution conductivity, temperature compensation for electrical conductivity measurement and total dissolved solids (TDS).

2.1 Solution Conductivity

This section formalizes the problem addressed in this study. We identify two interrelated limitations in current non-technical loss detection research: the common reduction of multiple operational conditions to a single abnormal class, and the widespread dependence on synthetic balancing techniques to address extreme class imbalance. These limitations degrade operational decision-making and undermine model reliability in deployment.

2.2 Binary-class issues

Much of the literature on NTL detection frames the task as a binary decision problem, (normal and abnormal) because this simplification fits standard supervised-learning pipelines and simplifies evaluation. However, this binary perspective conceals essential operational distinctions: what utilities categorize as an “abnormality” often requires qualitatively different responses. Under the PEA workflow, for example, energy theft triggers field inspections, evidence collection and potential legal action, whereas a defective meter prompts a maintenance

dispatch for repair or replacement without legal consequences. These responses differ markedly in urgency, cost, required personnel, and legal risk. By categorizing all cases simply as “abnormal,” the model requires manual triage by field teams, increases response latency, raises the likelihood of misdirected legal procedures, and undermines efficient resource allocation. To be operationally useful, NTL detection should be reframed as a three-class problem (normal, energy theft, defective meter), enabling alerts to be automatically routed to the appropriate operational unit.

2.3 Extreme imbalance and the pitfalls of balancing

Real-world NTL datasets exhibit extreme class imbalance: abnormal events, which encompass both energy theft and defective meters, typically account for only a few percent of all observations. In our data set, 309 abnormal instances are contrasted with 9,801 normal records. Common mitigation strategies in the literature include data-level balancing such as SMOTE, random oversampling and augmentation based on generative adversarial networks. Although these techniques often improve evaluation metrics in controlled experiments, they introduce substantive risks when models are deployed in production. Synthetic samples can embed artifact bias or unrealistic temporal and behavioral patterns, prompting models to learn augmentation specific noise rather than the true discriminative signals that separate theft from meter faults. This tendency to overfit to synthetic templates degrades generalization when novel anomalies deviate from the prototypes used during augmentation, producing fragile field performance. Heavy reliance on augmentation also increases practical burdens, because synthetic data must be regenerated for each retraining cycle, and computational cost grows, complicating continuous deployment. By contrast, approaches that operate directly on raw, imbalanced data preserve distributional fidelity and tend to be more robust, but they require more sophisticated modeling strategies to identify rare events without resorting to extensive synthetic inflation.

2.4 Contributions

This work addresses the semantic and reliability gaps in existing NTL research through the following contributions:

1. A multiclass detection framework is proposed that explicitly discriminates among normal, energy theft, and defective meter conditions, enabling automated routing of alerts to maintenance or investigative units and removing the operational ambiguity associated with binary labels.
2. SMOTE is applied selectively to the theft class in the second stage, reducing the volume of synthetic data by approximately 50%, lowering computational overhead, and enhancing generalization for rare-event theft detection relative to indiscriminate balancing.

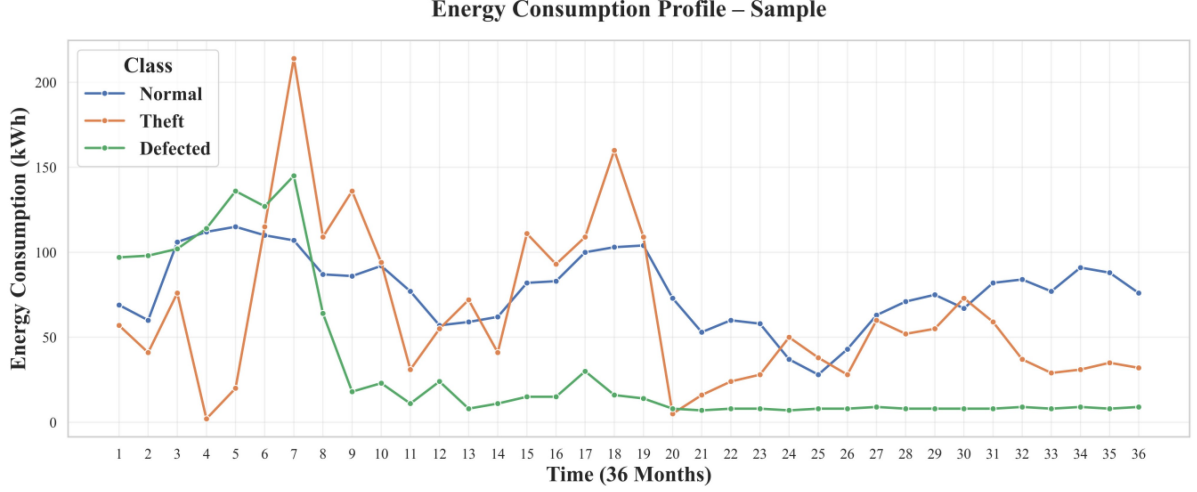


Fig. 1: Load profiles for Normal class, Theft class and Defected Class.

3. A hierarchical two-stage 1D-CNN pipeline is introduced that processes naturally imbalanced data in the first stage to isolate defective-meter cases that achieving perfect recall in the experimental evaluation thereby preserving distributional fidelity and avoiding synthetic artifacts in safety-critical defect detection.

Collectively, these contributions produce an operationally aligned, empirically validated approach to robust multiclass NTL detection suitable for real-world deployment.

3. PROPOSED METHODOLOGY

A two-stage hierarchical 1D-CNN framework is proposed to enable multiclass NTL detection while minimizing reliance on synthetic data. In the first stage, naturally imbalanced consumption sequences are used to detect defective meters; samples predicted as non-defective proceed to a second stage, where targeted synthetic oversampling is applied only to the theft class to support normal versus energy theft discrimination. Experiments are conducted to identify the best-performing model for three-class classification under minimal synthetic augmentation. The following subsections present the model architecture, training procedure, and evaluation protocol in detail.

3.1 Raw data and synthetic data

The experiments use a real-world dataset provided by the PEA, Mueang, Khon Kaen, Thailand, for nontechnical loss detection. The data were extracted from routine meter-reading systems and consist of monthly energy consumption in kilowatt-hours (kWh) for residential and selected commercial customers over 36 months (2021–2023). The dataset contains 10,110 samples corresponding to individual meters, with class counts of 9,801 normal, 120 energy theft, and 189 defective meters (approximately 97.0%, 1.2%, and 1.9% of the dataset, respectively), reflecting substantial class imbalance. All

records were anonymized to preserve privacy and confidentiality. Each sample is represented as a one-dimensional time series of 36 monthly consumption values (timesteps).

Class definitions are established based on PEA’s operational diagnostics and field validation.

- **Class [0]** – Normal: Meters exhibiting consistent, expected consumption patterns aligned with historical averages, seasonal trends, and customer load profiles. These are verified through routine inspections and billing consistency, with no detected anomalies in energy flow or meter functionality.

- **Class [1]** – Energy theft: Meters confirmed to have been tampered with or bypassed, resulting in deliberate under-reporting of consumption. Ground truth is derived from on-site forensic investigations and evidence of physical interference (e.g., shunt connections, meter reversal).

- **Class [2]** – Defective meter: Meters diagnosed with hardware or firmware faults causing inaccurate readings, either over- or under-recording consumption depending on customer behavior and load conditions. Labels are assigned following technical audits.

Representative load profiles from each class are illustrated in Fig. 1, highlighting distinct temporal patterns: stable cyclic behavior for normal (blue line), abrupt drops or flattening for energy theft (orange line), and erratic or drifting readings for defective meter (green line).

To mitigate the severe class imbalance, SMOTE is applied in experiments 2 and 4. SMOTE generates synthetic minority-class instances by interpolating between a sample x_i and one of its k nearest neighbors ($k = 5$) in feature space [30]:

$$x_{\text{syn}} = x_i + \lambda \cdot (x_{nn} - x_i), \lambda \sim \mathcal{U}(0, 1) \quad (1)$$

This creates new, plausible variations while avoiding exact duplication. Although effective in improving classifier performance on more balanced training sets, applying SMOTE to ultra-rare classes (e.g., fewer than

200 real samples) risks introducing non-physical temporal artifacts and overly generalized patterns, potentially reducing robustness when deployed on evolving real-world anomalies.

3.2 Models and performance metrics

Five classifiers are evaluated: three conventional machine learning models and two deep learning architectures. The selection of baseline models in this study is designed to cover the spectrum of algorithms most prevalent in NTL detection literature. SVM and RF act as robust benchmarks for classical machine learning. XGBoost is included to represent state-of-the-art gradient boosting techniques. LSTM serves as the standard deep learning baseline for time-series forecasting. Finally, 1D-CNN is chosen for the proposed framework due to its computational efficiency and superior capability in extracting local morphological features from short, sparse time-series data typical of monthly billing cycles. All models are trained and tested using five-fold stratified cross-validation. Representative hyperparameter settings used in the reported experiments are listed below to facilitate reproducibility.

Classical models:

- **SVM, implemented using scikit-learn:** radial basis function (RBF) kernel with $C = 10$, $\gamma = \text{scale}$, and $\text{class_weight} = \text{balanced}$.
- **RF, implemented using scikit-learn:** $n_estimators = 500$, $\text{max_depth} = \text{None}$, $\text{min_samples_split} = 2$, $\text{min_samples_leaf} = 1$, $\text{class_weight} = \text{balanced_subsample}$, $\text{random_state} = 42$, and parallelized with $n_jobs = -1$.
- **XGBoost:** $n_estimators = 500$, $\text{learning_rate} = 0.1$, $\text{max_depth} = 6$, $\text{subsample} = 0.8$, $\text{colsample_bytree} = 0.8$, $\text{objective} = \text{binary:logistic}$ for binary classification or multi:softmax for multiclass classification, $\text{eval_metric} = \text{logloss}$, $\text{random_state} = 42$, and $n_jobs = -1$.

Deep learning models:

- **1D-CNN, PyTorch:** The network comprises three convolutional blocks with 32, 64, and 128 filters, respectively. Each block consists of a 1D convolution (kernel size = 3, padding = 1), batch normalization, rectified linear unit (ReLU) activation, and max pooling (pool size = 2). The resulting feature map is flattened and passed to a classifier with a linear layer of 256 units, ReLU activation, dropout (rate = 0.5), and a final linear output layer with softmax (2 or 3 output classes). Input sequences are reshaped to (batch, 1, T). Optimization uses Adam with a learning rate of 1×10^{-3} , weight decay of 1×10^{-4} , batch size 64, and early stopping with patience set to 10 based on validation accuracy.

- **LSTM, PyTorch:** Two stacked LSTM layers with 64 and 128 hidden units, respectively, are used (unidirectional, batch-first). After the first LSTM, the final hidden state undergoes batch normalization and dropout (rate = 0.3), then is fed into the second LSTM, followed by batch

normalization and dropout (rate = 0.5). The final hidden state from the second LSTM is passed through a classifier consisting of a linear layer (256 units), ReLU activation, dropout (rate = 0.5), and a softmax output layer (2 or 3 classes). The input is reshaped to (batch, T, 1). Training uses the Adam optimizer with a learning rate of 1×10^{-3} , weight decay of 1×10^{-4} , batch size 64, and early stopping with patience = 10 based on validation accuracy.

We evaluate and compare models using five primary metrics: overall accuracy, per-class precision, per-class recall, per-class F1-score, and false negative rate (FNR). Overall accuracy assesses general correctness, while per-class precision, recall, and F1-score characterize performance on each label. F1-score is emphasized for the energy theft and defective meter classes because it balances precision and recall for these rare but operationally critical events. FNR is reported to quantify missed detections, which have important operational and legal consequences. In addition to the per-class results, we present macro-averaged precision, recall, and F1-score to provide an unweighted summary of performance across classes. The metrics are defined as follows [31]:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

$$\text{False Negative Rate (FNR)} = \frac{FN}{FN + TP} \quad (6)$$

where TP , FP , TN , and FN denote true positives, false positives, true negatives, and false negatives, respectively. Macro-averaged metrics are computed by taking the unweighted mean of the per-class metric values.

3.3 Experiment design

This study systematically evaluates the proposed two-stage hierarchical 1D-CNN framework across four experiments, with the corresponding data pipelines depicted in Fig. 2. The first two experiments benchmark performance in the conventional binary classification setting (normal vs. abnormal), while the latter two adopt the multiclass formulation (normal, energy theft, defective meter) under both imbalanced and balanced conditions, directly addressing the operational and reliability gaps identified in section 2.

Experiment 1 employs the raw imbalanced dataset without any synthetic augmentation. The binary abnormal class combines energy theft (120 samples) and defective meter (189 samples), totaling 309 abnormal instances against 9,801 normal samples. Five models (SVM, RF, XGBoost, 1D-CNN, and LSTM) are trained and evaluated under this realistic, deployment-like condition to highlight the limitations of binary classification on

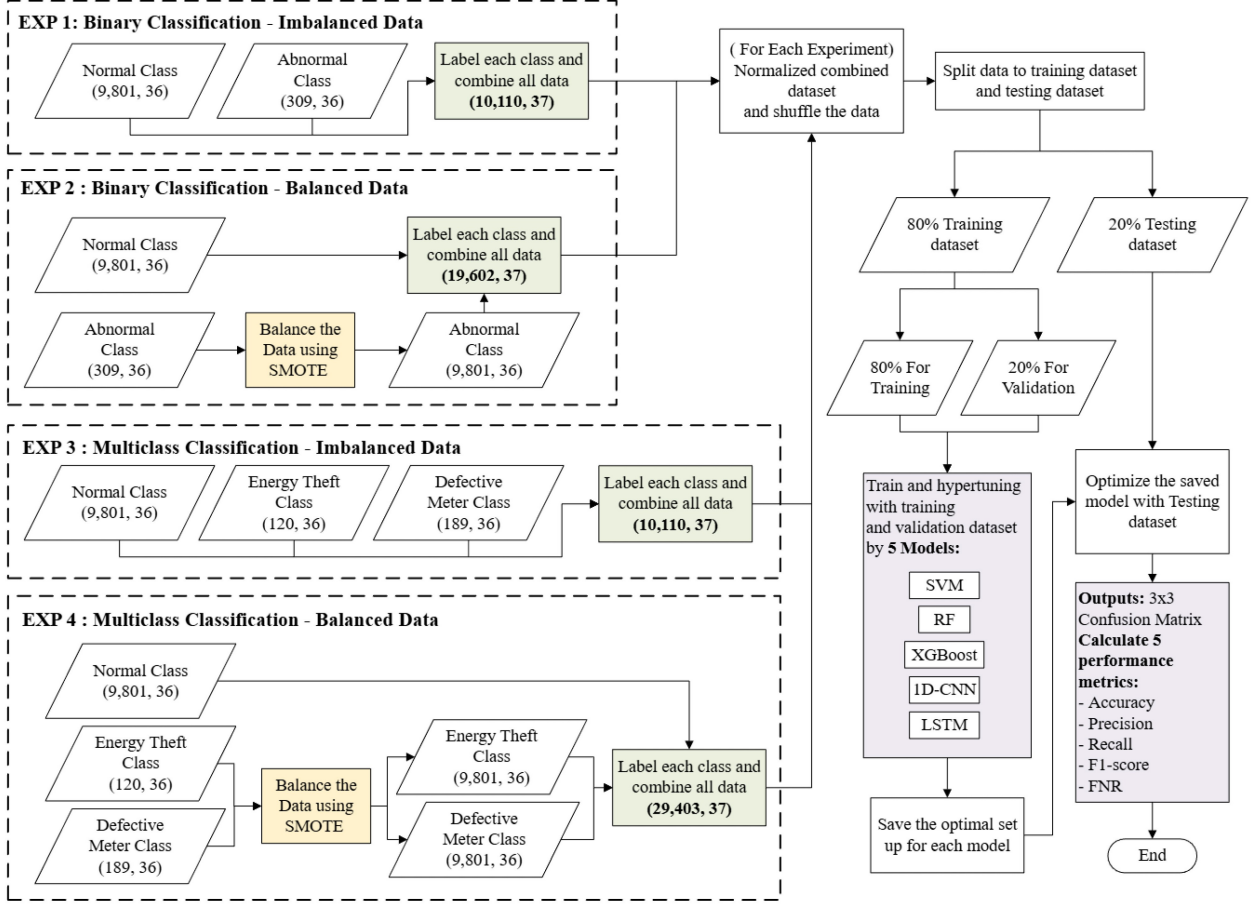


Fig. 2: Data pipeline of 4-experiment setup.

naturally imbalanced data. Experiment 2 retains the binary structure but applies SMOTE exclusively to the combined abnormal class, generating synthetic samples until both normal and abnormal classes reach 9,801 instances. This balanced configuration isolates the effect of class imbalance on model performance and establishes an upper-bound reference for binary NTL detection under full synthetic augmentation. The primary contribution is introduced in experiment 3, where the raw imbalanced dataset is reformulated as a three-class problem: normal (9,801), energy theft (120), and defective meter (189). No synthetic data are introduced, enabling direct evaluation of multiclass classification on real-world data. Experiment 4 extends the multiclass setup by applying SMOTE independently to each minority class (energy theft and defective meter) until all three classes are balanced at 9,801 samples each. This experiment defines a theoretical performance ceiling under fully balanced synthetic conditions while highlighting the risks of generating artificial patterns from very limited real samples. All datasets undergo standard preprocessing before model training. Outliers are handled via the 1st to 99th percentile clipping method, where values below the 1st percentile are set clipped to the 1st percentile and those above the 99th percentile are replaced by the 99th percentile. No imputation is performed; samples

containing not a number (NaN) values are excluded. Features are then normalized using min-max scaling [32]:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (7)$$

The preprocessed data are partitioned by unique meter IDs into 80% training and 20% testing sets. This meter-wise split ensures that the test set consists of fully isolated, unseen meters, preventing any data leakage between training and evaluation phases. The training subset is further divided into 80% training and 20% validation (i.e., 64% training, 16% validation, 20% test of the original data). The training set is used to optimize model parameters and learn patterns in load profiles that distinguish between classes. In contrast, the validation set is used for hyperparameter tuning (e.g., learning rate, filter counts, dropout) and to monitor overfitting or underfitting. Training continues until validation loss converges, as confirmed by stable learning curves, and the best-performing model is saved. The held-out test set, which is entirely unseen during training, is then used to compute final performance metrics: accuracy, precision, recall, F1-score, and FNR, both per class and macro-averaged. Results from experiments 1-4 are presented in Tables 1-4, providing the foundation for the comparative analysis in the following discussion.

Table 1: Comparison of five models for a two-class classification problem using an imbalanced dataset. (Experiment 1).

Models	Acc.	Precision			Recall			F1-Score			FNR		
		[0]	[1]	[Avg]	[0]	[1]	[Avg]	[0]	[1]	[Avg]	[0]	[1]	[Avg]
SVM	97.80	99.60	59.30	79.50	98.10	87.10	92.60	98.80	70.60	84.70	1.90	12.90	7.40
RF	99.00	99.30	89.10	94.20	99.70	79.00	89.40	99.50	83.80	91.60	0.30	21.00	10.60
XGBoost	99.00	99.30	90.70	95.00	99.70	79.00	89.40	99.50	84.50	92.00	0.30	21.00	10.60
1D-CNN	99.10	99.40	90.90	95.20	99.70	80.60	90.20	99.50	85.50	92.50	0.30	19.40	9.90
LSTM	94.91	98.38	30.39	64.39	98.39	50.00	74.19	98.38	38.27	68.33	1.61	50.00	25.81

Table 2: Comparison of five models for a two-class classification problem using a balanced dataset. (Experiment 2).

Models	Acc.	Precision			Recall			F1-Score			FNR		
		[0]	[1]	[Avg]	[0]	[1]	[Avg]	[0]	[1]	[Avg]	[0]	[1]	[Avg]
SVM	98.93	98.08	98.97	98.53	98.98	98.08	98.53	98.53	98.53	98.53	1.02	1.92	1.47
RF	99.47	99.59	98.42	99.01	98.44	99.59	99.02	99.01	99.00	99.01	1.56	0.41	0.99
XGBoost	99.42	99.44	99.74	99.59	99.74	99.44	99.59	99.59	99.59	99.59	0.26	0.56	0.41
1D-CNN	99.93	99.95	99.85	99.90	99.85	99.95	99.90	99.90	99.90	99.90	0.15	0.05	0.10
LSTM	92.66	88.49	97.44	92.96	97.71	87.30	92.51	92.86	92.12	92.49	2.29	12.70	7.50

Table 3: Comparison of five models for a three-class classification problem using an imbalanced dataset. (Experiment 3).

Models	Acc.	Precision				Recall				F1-Score				FNR			
		[0]	[1]	[2]	[Avg]	[0]	[1]	[2]	[Avg]	[0]	[1]	[2]	[Avg]	[0]	[1]	[2]	[Avg]
SVM	97.30	99.40	22.60	97.40	73.13	97.90	50.00	97.40	81.77	98.60	31.40	97.40	75.80	2.10	50.00	2.60	18.23
RF	98.90	99.70	58.30	97.40	85.13	99.70	58.30	97.40	85.13	99.70	58.30	97.40	85.13	0.30	41.70	2.60	14.87
XGBoost	99.10	99.50	82.40	95.00	92.30	99.80	58.30	100.00	86.03	99.70	68.30	97.40	88.47	0.20	41.70	0.00	13.97
1D-CNN	99.26	99.39	90.91	95.00	95.10	99.95	41.67	100.00	80.54	99.67	57.14	97.44	84.75	0.05	58.33	0.00	19.46
LSTM	94.17	98.63	17.78	31.25	49.22	95.46	33.33	65.79	64.86	97.02	23.4	42.37	54.26	4.54	66.67	34.21	35.14

Table 4: Comparison of five models for a three-class classification problem using a balanced dataset. (Experiment 4).

Models	Acc.	Precision				Recall				F1-Score				FNR			
		[0]	[1]	[2]	[Avg]	[0]	[1]	[2]	[Avg]	[0]	[1]	[2]	[Avg]	[0]	[1]	[2]	[Avg]
SVM	99.15	99.23	98.43	99.64	99.10	98.42	98.88	100.00	99.10	98.82	98.65	99.82	99.10	1.58	1.12	0.00	0.90
RF	99.23	98.68	99.18	99.85	99.24	99.44	98.52	99.75	99.24	99.06	98.85	99.80	99.24	0.56	1.48	0.25	0.76
XGBoost	99.80	99.69	99.74	99.95	99.79	99.74	99.69	99.95	99.79	99.72	99.72	99.95	99.80	0.26	0.31	0.05	0.21
1D-CNN	99.91	99.95	99.80	100.00	99.92	99.80	99.95	100.00	99.92	99.87	99.87	100.00	99.91	0.20	0.05	0.00	0.08
LSTM	99.35	99.85	98.49	99.9	99.41	98.52	99.64	99.80	99.22	99.18	99.06	99.85	99.36	1.48	0.36	0.20	0.68

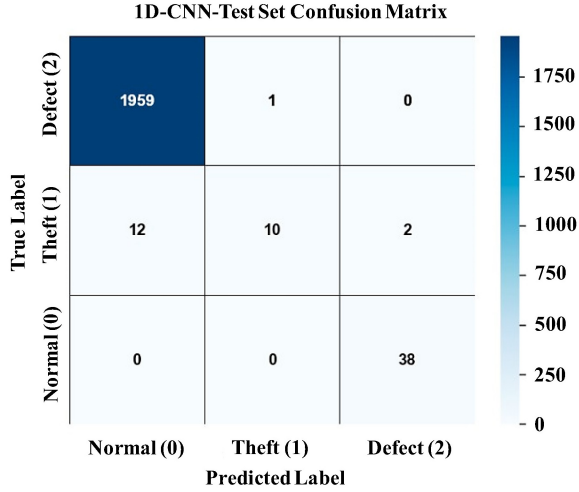


Fig. 3: Confusion Matrix with testing of Experiment 3.

3.4 RESULTS AND DISCUSSION

This section presents the evaluation metrics defined in Section 3.2 reported both per class and as macro averages for each of the four experiments, and is followed by a detailed discussion of the results.

Experiment 1: Binary classification (normal vs. abnormal) is evaluated on the raw, naturally imbalanced dataset, where the abnormal label aggregates energy theft and defective meter cases. The results of experiment 1 are summarized in Table 1, where the 1D-CNN attained the highest nominal accuracy (Acc.) of 99.10%; however, this figure is misleading because the majority class overwhelmingly dominates (9,801 normal versus 309 abnormal). Under such extreme imbalance, overall accuracy primarily reflects correct majority-class predictions and conceals bias toward normal outcomes, so it is not informative about rare-event detection performance.

To quantify the model’s true discriminative capability for the minority condition, class-wise precision, recall, and the F1-score for the abnormal class (Class [1]) were examined. The 1D-CNN achieved an F1-score of 85.50% for Class [1], outperforming all evaluated baselines (XGBoost: 84.50%, RF: 83.80%, SVM: 70.60%, LSTM: 38.27%). This pattern indicates that, despite operating on unaugmented, imbalanced data, the 1D-CNN provides substantially better balanced detection of abnormal load profiles than traditional classifiers, yielding moderate-to-strong reliability for operationally relevant rare-event identification.

Experiment 2: Applies SMOTE to balance minority classes to 9,801 samples with the results are summarized in Table 2, where a 1D-CNN achieves 99.93% accuracy, 99.95% recall, and 99.90% F1-score for the abnormal class, reducing FNR to 0.05%. XGBoost and RF also exceed 99% in key metrics, while SVM improves significantly. This confirms class imbalance as the dominant bottleneck in binary NTL detection and shows that balanced data enables near-perfect performance even with conventional

models. However, heavy reliance on synthetic data introduces critical risks: SMOTE-generated samples may embed unrealistic patterns, overfit to artifacts, and fail to generalize to real-world anomalies. The resulting models, while strong in controlled tests, remain fragile and potentially computationally costly in deployment.

Experiment 3: This experiment evaluates the proposed multiclass formulation on the raw, imbalanced dataset (normal: 9,801; energy theft: 120; defective meter: 189). The 1D-CNN results, summarized in Table 3, show that the 1D-CNN achieves 100% recall and 0% false-negative rate for defective meters (Class 2) with 95.00% precision, indicating that no defective meter is missed and that false alarms occur in approximately one out of twenty defect predictions. This capability enables fully automated, low-risk dispatch of all predicted Class 2 instances to maintenance crews, directly supporting reliable repair workflows. The confusion matrix for the test set (Fig. 3) visually confirms this outcome: all 38 true defective meters are correctly classified with no misclassifications into normal or theft. Only two false positives originate from the energy theft class, which yields the reported 95.00% precision. By contrast, energy theft (Class 1) exhibits lower sensitivity: recall for theft is 41.67% (10 of 24 test theft instances correctly identified), with 12 misclassified as normal and 2 as defective, reflecting a conservative bias toward the dominant normal class under severe imbalance and a tendency to prioritize avoidance of false alarms over minority detection.

Further analysis reveals that these misclassified theft instances often exhibited subtle, gradual declines rather than abrupt drops, mimicking natural consumption variations. Future improvements could involve integrating external features such as customer category and geographic location to enhance the model’s ability to distinguish between legitimate conservation and sophisticated tampering.

Despite the uneven performance, the 1D-CNN demonstrates better overall balance across classes, outperforming XGBoost on macro F1-score. Crucially, no defective meter is ever confused with theft in the 1D-CNN outputs, thereby preventing costly and inappropriate legal escalation; this reliability is the principal reason for selecting the 1D-CNN for further development. Thus, experiment 3 establishes that a multiclass 1D-CNN trained on raw data can achieve operationally reliable defective-meter isolation. This finding is a cornerstone of the proposed two-stage framework while simultaneously highlighting the need for targeted augmentation in stage 2 to increase theft recall without compromising the integrity of defect detection.

Experiment 4: This applies SMOTE independently to each class, upsampling all three labels to 9,801 samples. As reported in Table 4, this synthetic balancing yields near-perfect results across models: the 1D-CNN attains 99.91% accuracy, 100% recall for defective meters, and 99.95% recall for energy theft. However, these gains represent a theoretical upper bound that depends on

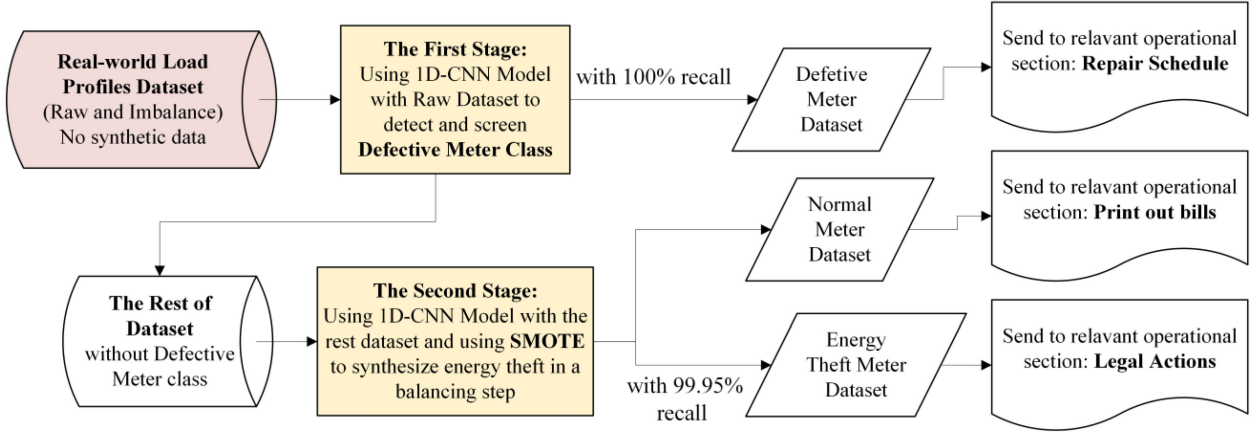


Fig. 4: Framework of the main contribution with multiclass classification for utility to take actions.

Table 5: Performance Comparison of Multiclass NTL Detection Strategies and the proposed method.

Method of 3-class classification	% Recall [1]	% Recall [2]	Amount of synthetic data
1. No synthetic data	41.67	100	0
2. Synthetic data	99.95	100	19,293
3. The two -stage 1D-CNN(Proposed)	99.95	100	9,681

synthesizing more than 97% of the data for the two minority classes, which raises substantive concerns. Heavy reliance on synthetic examples increases the risk of overfitting to augmentation artifacts, reduces robustness to novel anomaly patterns, and imposes substantial computational overhead for repeated re-training. Moreover, the marginal improvement over experiment 3 for defective-meter detection, for which recall is already 100% on raw data. This underscores that generating synthetic defective samples is unnecessary and potentially harmful when raw-data performance is already operationally sufficient. The experimental results highlight the complementary strengths of each configuration that together produce a coherent, operationally viable solution. The proposed two-stage hierarchical framework, depicted in Fig. 4, integrates these insights into a single deployable pipeline. Stage 1 leverages a 1D-CNN trained on the raw, highly imbalanced dataset (the experiment 3 configuration) to isolate defective meters. Operating without any synthetic augmentation, this stage attains 100% recall for defective-meter cases, thereby preserving the native data distribution and avoiding augmentation artifacts in safety-critical repair decisions. Cases classified as defective are routed directly to maintenance crews, eliminating unnecessary legal exposure and manual triage.

Then, samples not assigned to the defective class form a reduced subset of normal and potential theft profiles and are processed in stage 2. Here, augmentation via SMOTE is applied solely to the theft class to rebalance

the binary normal versus energy theft task. By confining synthesis to the theft domain, the approach capitalizes on the highly performance demonstrated in experiment 2 while limiting synthetic exposure: theft detection performance is balancing, and the total volume of synthetic data is reduced by approximately 50% compared with full three-class SMOTE. This hierarchical strategy therefore aligns classification outputs with operational workflows, enhances robustness by protecting defect detection from augmentation bias, and reduces computational and managerial overhead associated with extensive synthetic augmentation. The hierarchical design addresses both gaps identified in this study. It closes the operational gap by producing outputs that map directly to operational actions: the defective class (Class 2) is routed to the repair schedule, while the theft class (Class 1) is directed to investigation or legal procedures, minimizing the amount of manual review required before action is taken. It mitigates the reliability gap by avoiding synthetic defective samples and restricting augmentation to the class supported by sufficient real examples, thereby reducing overfitting risk, improving generalization to novel anomalies, and lowering computational cost. As shown in Table 5, compared with a single-stage, three-class SMOTE approach (experiment 4), the two-stage framework reduces synthetic data generation by 49.82% while achieving equivalent operational outcomes with zero missed defects and near-zero missed thefts. Overall, the experimental progression indicates that class imbalance, rather than model capacity, constrains performance, and that targeted, stage-wise augmentation combined with reliance on raw data yields a principled, utility-aligned blueprint for actionable NTL detection.

4. CONCLUSION, LIMITATIONS AND FUTURE WORK

This paper introduces a two-stage hierarchical 1D-CNN framework for multiclass NTL detection using real-world monthly consumption data from conventional mechanical meters operated by the PEA in Khon Kaen, Thailand. The method addresses both the operational

limitations of binary classification and the reliability issues associated with extensive synthetic oversampling by performing three-class classification (normal, energy theft and defective meter) on raw imbalanced data in the first stage, achieving 100% recall and 95.00% precision for defective meters with zero synthetic samples. The second stage applies SMOTE to the remaining theft class, reducing synthetic data volume by approximately 50% (from 19,293 to 9,681 samples) while maintaining 99.95% theft recall. Crucially, this proposed framework elevates the theft detection capability from a baseline of 41.67% (without synthetic data) to 99.95%. This substantial improvement means the utility can identify and recover revenue from approximately 58% more theft cases that would otherwise go undetected by traditional methods. Compared to full synthetic balancing and baselines (SVM, RF, XGBoost, LSTM), the proposed framework delivers equivalent or superior detection performance at lower computational cost, with reduced artifact risk and direct operational alignment, enabling automated routing of repair dispatch for defective meters and legal investigation for theft.

Despite these strengths, this work has several limitations. First, the model depends on real-world training samples and may not generalize to new abnormality patterns, such as novel theft techniques or meter failure modes, that are not represented in the current dataset; in such cases, it would require retraining with updated labeled data. To adapt this framework to other regions with different customer behaviors or meter types, utilities should implement periodic retraining to capture evolving patterns, employ transfer learning to fine-tune the 1D-CNN on local datasets with minimal labeling effort, and utilize regional feature normalization to ensure consumption scales remain consistent across different service areas. Second, monthly temporal resolution limits detection of short-term anomalies (e.g., intra-month theft), potentially missing dynamic events. Third, manual meter readings may introduce recording errors or delays, which propagate as noise and affect pattern consistency. However, this framework enables “semi-real-time” detection at the billing stage, allowing utilities to intercept anomalies before invoice issuance. The architecture also remains adaptable to future smart meter data for true real-time monitoring. Specifically, the 1D-CNN input layer can be resized to process high-frequency sequences (e.g., hourly or daily readings), allowing the model to capture finer-grained temporal patterns such as intra-day theft or short-term meter malfunctions without altering the core hierarchical framework.

Future work will focus on online incremental learning to adapt to evolving patterns without full retraining, integration of explainable AI (e.g., using SHAP to identify key monthly contributors to anomaly scores or Grad-CAM/Attention-based visualization to pinpoint specific months where consumption deviations indicate meter tampering or failure) to support field validation, and a large-scale pilot deployment across multiple PEA

districts to assess long-term robustness, quantify cost savings, and explore hybrid data fusion with smart meter subsets for finer-grained monitoring.

Finally, regarding post-detection, utility will dispatch field teams to verify anomalies on-site, followed by legal action for theft or meter replacement for defects.

ACKNOWLEDGMENTS

The authors wholeheartedly thank the Center of Multidisciplinary Innovation Network Talent (MINT Center), the Department of Technology and Engineering, the Faculty of Interdisciplinary Studies, Khon Kaen University, Nong Khai Campus, Nong Khai Province, Thailand. The authors also express sincere gratitude to the Provincial Electricity Authority (PEA), Khon Kaen Province, Thailand, for providing the dataset used in this study. Their combined support played a pivotal role in enabling the successful completion of the research presented in this paper.

AUTHOR CONTRIBUTIONS

The manuscript was written with contributions from all authors. They discussed the results and commented on the manuscript. All authors approved the final version of the manuscript.

REFERENCES

- [1] R. M. R. Barros, E. G. Da Costa, and J. F. Araujo, “Maximizing the Financial Return of Non-Technical Loss Management in Power Distribution Systems,” *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1634–1641, Mar. 2022.
- [2] H. O. Henriques, R. L. S. Corrêa, M. Z. Fortes, B. S. M. C. Borba, and V. H. Ferreira, “Monitoring technical losses to improve non-technical losses estimation and detection in LV distribution systems,” *Measurement*, vol. 161, Art. no. 107840, Sep. 2020.
- [3] J. B. Leite and J. R. S. Mantovani, “Detecting and locating non-technical losses in modern distribution networks,” *IEEE Trans Smart Grid*, vol. 9, no. 2, pp. 1023–1032, Mar. 2018.
- [4] D. Carr and M. Thomson, “Non-Technical Electricity Losses,” *Energies*, vol. 15, no. 6, p. 2218, Mar. 2022.
- [5] V. Gaur and E. Gupta, “The determinants of electricity theft: An empirical analysis of Indian states,” *Energy Policy*, vol. 93, pp. 127–136, Jun. 2016.
- [6] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, “Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft,” *Energy Policy*, vol. 39, no. 2, pp. 1007–1015, Feb. 2011.
- [7] L. Ventura, G. E. Felix, R. Vargas, L. T. Faria, and J. D. Melo, “Estimation of non-technical loss rates by

- regions,” *Electric Power Systems Research*, vol. 223, Art. no. 109685, Aug. 2023.
- [8] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, “Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft,” *Energy Policy*, vol. 39, no. 2, pp. 1007–1015, Feb. 2011.
- [9] S. Kumar V., J. Prasad, and R. Samikannu, “Overview, issues and prevention of energy theft in smart grids and virtual power plants in Indian context,” *Energy Policy*, vol. 110, pp. 365–374, Nov. 2017.
- [10] S. Janthong, R. Duangsoithong, and K. Chalermyanont, “Feature Extraction of Risk Group and Electricity Theft by using Electrical Profiles and Physical Data for Classification in the Power Utilities,” *ECTI Transactions on Computer and Information Technology*, vol. 18, no. 1, pp. 51–63, Jan. 2024.
- [11] S. Janthong, K. Chalermyanont, and R. Duangsoithong, “Unbalanced Data Handling Techniques for Classifying Energy Theft and Defective Meters in the Provincial Electricity Authority of Thailand,” *IEEE Access*, vol. 11, pp. 46522–46540, May. 2023.
- [12] PEA, “Performance Report According to the PEA Operational Plan for the Year 2023 Quarter 1–4,” Available from <https://www.pea.co.th/en/about-pea/annual-report>. Accessed: Nov. 08, 2025.
- [13] S. Boonjaraspinyo *et al.*, “Prevalence and Associated Risk Factors of Intestinal Parasitic Infections: A Population-Based Study in Phra Lap Sub-District, Mueang Khon Kaen District, Khon Kaen Province, Northeastern Thailand,” *Tropical Medicine and Infectious Disease*, vol. 8, no. 1, Art. no. 22, Jan. 2023.
- [14] C. Adelina and J. Y. C. Han, “Entrepreneurial urbanization and masculine identities: an exploratory study of Khon Kaen, Thailand,” *Asian Journal of Social Science*, vol. 53, no. 2, Art. no. 100197, Jun. 2025.
- [15] J. L. Gallardo, M. A. Ahmed, and N. Jara, “Clustering algorithm-based network planning for advanced metering infrastructure in smart grid,” *IEEE Access*, vol. 9, pp. 48992–49006, Mar. 2021.
- [16] J. Y. Kim, Y. M. Hwang, Y. G. Sun, I. Sim, D. I. Kim, and X. Wang, “Detection for Non-Technical Loss by Smart Energy Theft with Intermediate Monitor Meter in Smart Grid,” *IEEE Access*, vol. 7, pp. 129043–129053, Sep. 2019.
- [17] M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, “Hybrid Deep Neural Networks for Detection of Non-Technical Losses in Electricity Smart Meters,” *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1254–1263, Mar. 2020.
- [18] F. Jamil, “Electricity theft among residential consumers in Rawalpindi and Islamabad,” *Energy Policy*, vol. 123, pp. 147–154, Dec. 2018.
- [19] A. Nadeem and N. Arshad, “A data-driven approach to reduce electricity theft in developing countries,” *Util Policy*, vol. 73, Art. no. 101304, Dec. 2021.
- [20] S. Khunkitti, A. Siritaratiwat, and S. Premrudeepreechacharn, “A Many-Objective Marine Predators Algorithm for Solving Many-Objective Optimal Power Flow Problem,” *Applied Sciences*, vol. 12, no. 22, Art. no. 11829, Nov. 2022.
- [21] G. M. Messinis and N. D. Hatziaargyriou, “Review of non-technical loss detection methods,” *Electric Power Systems Research*, vol. 158, pp. 250–266, May. 2018.
- [22] Q. Cai, P. Li, and R. Wang, “Electricity theft detection based on hybrid random forest and weighted support vector data description,” *International Journal of Electrical Power and Energy Systems*, vol. 153, Art. no. 109283, Nov. 2023.
- [23] N. Rajitroj, K. Seeprompting, and R. Chatthaworn, “1D-CNN and A Weight-Based Balancing Technique for Improved Detection of Non-Technical Losses in Power Distribution Systems,” in *2025 15th International Conference on Future Environment and Energy (ICFEE 2025)*, Sapporo, Japan, 2025, Art. no. 06005.
- [24] S. K. Gunturi and D. Sarkar, “Ensemble machine learning models for the detection of energy theft,” *Electric Power Systems Research*, vol. 192, Art. no. 106904, Mar. 2021.
- [25] Z. A. Khan, M. Adil, N. Javaid, M. N. Saqib, M. Shafiq, and J. G. Choi, “Electricity theft detection using supervised learning techniques on smart meter data,” *Sustainability*, vol. 12, no. 19, Art. no. 8023, Oct. 2020.
- [26] M. A. Souza *et al.*, “Detection of Non-Technical Losses on a Smart Distribution Grid Based on Artificial Intelligence Models,” *Energies*, vol. 17, no. 7, Art. no. 1729, Apr. 2024.
- [27] M. Jené-Vinuesa, M. Aragüés-Peñalba, and A. Sumper, “Comprehensive Data-Driven Framework for Detecting and Classifying Non-Technical Distribution Losses,” *IEEE Access*, vol. 12, pp. 96537–96551, Jul. 2024.
- [28] Z. Qu, H. Li, Y. Wang, J. Zhang, A. Abu-Siada, and Y. Yao, “Detection of electricity theft behavior based on improved synthetic minority over-sampling technique and random forest classifier,” *Energies*, vol. 13, no. 8, Art. no. 2039, Apr. 2020.
- [29] M. Jene-Vinuesa, M. Aragüés-Peñalba, and A. Sumper, “Comprehensive Data-Driven Framework for Detecting and Classifying Non-Technical Distribution Losses,” *IEEE Access*, vol. 12, pp. 96537–96551, Jul. 2024.
- [30] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *The Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002.
- [31] D. M. W. Powers and Ailab, “Evaluation: from

Precision, Recall and F-measure to ROC, Informedness, markedness & correlation”, *International Journal of Machine Learning Technology*, pp. 37-63, 2011.

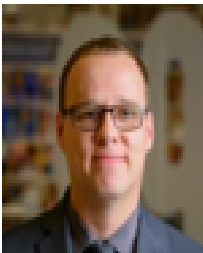
- [32] S. G. K. Patro and K. K. Sahu, “Normalization: A Preprocessing Stage.” *arXiv:1503.06462*, Mar. 2015.



Katsarin Seepromting received the B.Eng. and M.Eng. degrees in Electrical Engineering from Khon Kaen University, Khon Kaen, Thailand. She is currently pursuing the Ph.D. degree in Electrical and Computer Engineering at Khon Kaen University, Nong Khai Campus, Thailand. Her research interests include the integration of artificial intelligence in power distribution systems, advanced data analytics, and renewable resource management.



Nattaya Rajitroj received the B.Eng. degree in Mechanical Engineering from Khon Kaen University, Khon Kaen, Thailand. She is currently pursuing the Ph.D. degree in Electrical Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, Thailand. Her current research interests focus on smart grid technologies, deep learning applications, and artificial intelligence for energy management.



Jakub Michel received his PhD in Statistics from University of New Hampshire in 2023. His current position is an Assistant Professor at High Point University. His current interests are in applied statistical and machine-learning methods that support evidence-based decisions in energy, health, and other real-world systems, with an emphasis on rigor, interpretability, and impact. Dr. Michel is a member of Sigma Xi, The Scientific Research Honor Society and a MAA Project

NExT Fellow.



Arun Onlam received the M.Eng. degree in Electrical Engineering from Khon Kaen University, Khon Kaen, Thailand. He currently holds the position of Chief of the Meter and Transformer Section at the Provincial Electricity Authority (PEA), Khon Kaen Province, Thailand. With over 10 years of industrial experience in metering infrastructure and power systems, his expertise encompasses utility asset management and smart metering deployment strategies.



Chayada Surawanitkun received the B.Eng. degree in Electrical Engineering (1st class honor) and the Ph.D. degree in Electrical Engineering from Khon Kaen University, Thailand, in 2009 and 2013, respectively. She was granted the B.Eng. Scholarship from the Industrial/University Cooperative Research Center in HDD Component, supported by the Ministry of Science and Technology, and a Ph.D. scholarship from the Royal Golden Jubilee Ph.D. Program under The Thailand

Research Fund. She is Program Chair of Master's and Doctoral programs in Electrical and Computer Engineering of the Faculty of Interdisciplinary Studies, Department of Technology and Engineering, Khon Kaen University, Nong Khai Campus, Thailand. Her research interests include power systems, data storage technology, magnetic

random-access memory, thermal analysis of nanodevices, nanotechnology, renewable and sustainable energy, and energy management systems.



Apirat siritaratiwat received the B.Eng. degree in Electrical Engineering from Khon Kaen University, Thailand, in 1992. After working in industry for a few years, he joined the Department of Electrical Engineering, Khon Kaen University (KKU) in 1994 and continued his postgraduate study at the University of Manchester, U.K., between 1995 and 1999. He has done research in HDD and is one of the pioneer researchers of ESD/EOS and EMI in recording heads with the KKU-Seagate Cooperation Research Laboratory, Faculty of Engineering, Khon Kaen University. Currently, He works at Department of IoT and Information Engineering, Faculty of Engineering, King Mongkut's Institute of Technology, Ladkrabang, Bangkok. His research interests are ESD/EOS, EMI, renewable energy and energy management system.