National College of Ireland

Smart Meter Electricity Theft Detection Using Machine and Deep Learning Approaches

MSc Research Project
MSc in Data Analytics

Serap BATIR
Student ID: X22205179

School of Computing
National College of Ireland

Supervisor: Furqan Rustam

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Serap Batır
Student ID:	X22205179
Programme:	Msc in Data Analytics
Year:	2025
Module:	MSc Research Project
Supervisor:	Furqan Rustam
Submission Due Date:	10th August 2025
Project Title:	Smart Meter Electricity Theft Detection Using Machine and Deep Learning Approaches
Word Count:	8425
Page Count:	30

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Serap BATIR
Date:	10th August 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Smart Meter Electricity Theft Detection Using Machine and Deep Learning Approaches

SERAP BATIR
X22205179

Abstract

Electricity theft is a growing problem for utility companies all over the world. This problem has been shown to cause significant loss of money and inefficiency in operations. This study analysed the use of machine learning and deep learning methods in identifying fraudulent consumption behaviour in smart meter data. The study employed a scenario-based approach, which creates flat reduction, selective tampering and consumption mean manipulation scenarios and used them in the training and evaluation of detection problems. Comparative experimentation showed that Long Short-Term Memory and Gated Recurred Unit (GRU) models perform better (94% accuracy) in the identification of temporal patterns, whereas Random Forest classifiers were highly effective (94% accuracy). The research reveals major drawbacks in the conventional methods, as basic K-nearest Neighbors were less efficient in dealing with dependencies over long periods and Support Vector Machines had poor scaling. XGBoost incurred the highest training time with only moderate accuracy gains. To summarize, this paper brings value to the field because it proposes an evaluation framework to be applied to theft detection algorithms, which has been tested with realistic consumption data. The study also sets down workable criteria on model selection based on the needs of an operation, given the balance between accuracy and resources. This contribution shows which methods work well in detecting electricity theft and highlights some of main challenges in implementing them.

Keywords: Electricity theft, deep learning, smart meter, machine learning, anomaly detection

1 Introduction

The theft of electricity is a significant problem worldwide, causing utility companies to incur huge losses and face operational challenges. Electricity theft is the most prevalent cause of Non-Technical Losses (NTLs), resulting in over 96 billion dollars' worth of damages per year (Lepolesa et al.,2022). Such theft of power can be as high as 50% of what is generated, especially in some places in sub-Saharan Africa (Louv&Bokoro,2019). Theft of electricity in India is a significant factor contributing to power loss of up to 25%, amounting to 16.2 billion dollars in 2015 (Chen et al.,2023). Similarly, other nations such as Brazil, Russia, and South Africa are also highly vulnerable to revenue loss, as it is

estimated that Eskom of South Africa loses nearly 1.31 billion a year through theft (Lepolesa et al.,2022).

The practice has not only undermined the financial strength of utility providers but also skewed the accuracy of data about energy utilization. While this happens, electric theft is becoming increasingly complex to solve using traditional methods, such as manual checks and straightforward anomaly detection rules (Fährmann et al.,2024). Current methods of electricity theft detection through manual inspection and rule-based systems show limits because they are inefficient and challenging to scale. Smart meter data shows dynamic characteristics which makes rule-based systems unreliable and manual inspections take more time. As such, a practical methodology will be required to deal with the problem of electricity theft. This paper investigates the potential of a deep learning (DL) and machine learning (ML) framework for tracking theft using smart meter data, offering a novel solution to a common issue.

This paper examined the performance of machine learning models, such as SVM, XGBoost, KNN and Random Forest, in comparison to deep learning models like LSTM, LSTM (a few shot) and GRU. The research mainly compared the accuracy, precision, recall, and F1-score of these models as regards detecting electricity theft patterns. A labelled dataset based on scenario-based synthesis was created, making it possible to train a supervised learning model efficiently. Also, the research sought to reveal how the mentioned advanced methods have the potential to enhance the identification of anormal consumption patterns and assist utility companies in mitigating the loss of revenues due to electricity theft.

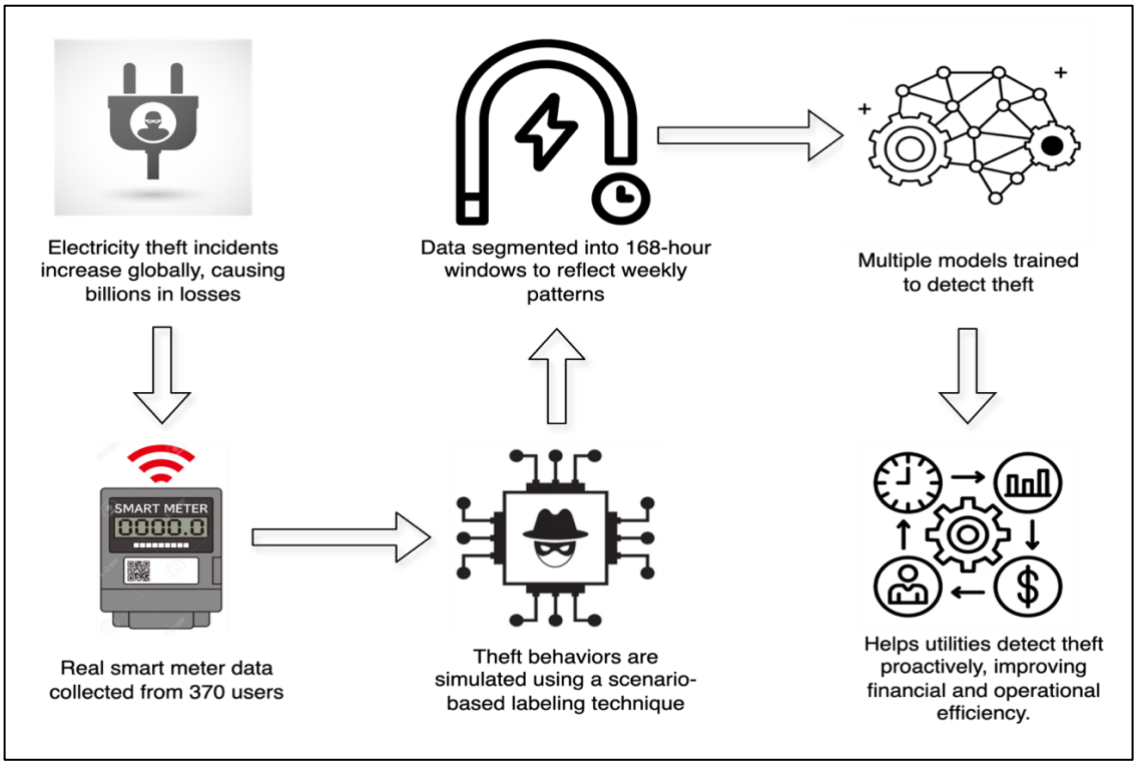


Figure 1 Conceptual overview of the study

The addressed framework (Figure 1) integrates real smart meter data with theft scenario simulations to replicate realistic theft patterns. The approach relies on the need to capture temporal consumption trends to detect theft within 168- hour fixed segments of data. This design focuses of the ML and DL techniques with smart meters consumption data to

provide theft detection evaluation. The study directly targeting the research question: How can ML and DL techniques be effectively applied to detect electricity theft using consumption data from smart meters?

This research contributions of this paper are as follows:

- Examined the smart meter consumption data to assess the feasibility of applying ML and DL methodologies for the detection of electricity theft.
- Addressed limitations of existing detection methodologies through scenario-oriented labelling to create simulations of realistic theft behaviors.
- Implemented and evaluated multiple ML (SVM, KNN, RF, XGBoost) and DL models (LSTM, LSTM (a few shot), GRU) on a realistic dataset.
- Conducted a comparative performance analysis using standard evaluation metrics) to identify the most effective approach for operational use.

This paper has seven topics. Section 1 presents the issue of electricity theft and highlights the problems faced by utility companies. Section 2 provides a literature review of electricity theft detection, including ML, DL, hybrid and other technologies. Also, this section shows the gaps in existing research. Section 3 describes the methodology, in particular, data gathering, preprocessing, and the use of synthetic scenario-oriented labeling. It also includes the employed machine learning and deep learning models evaluation metrics. Section 4 shows the description of the system implementation and model design, with an indication of the proposed methods and algorithms. Section 5 is about the implementation process, the tools and languages applied, the techniques used, the outcomes generated, the model predictions, and the performance. Section 6 looks at the experimental results by comparing DL models (LSTM, GRU, LSTM a few shot) to ML models (SVM, KNN, Random Forest, XGBoost), and it analyzes the results using confusion matrix and key performance measures. Lastly, the paper is closed in Section 7 with a summary of findings, a discussion of implications, and directions for future research on electricity theft detection.

2 Literature Review

Electricity theft has received primary concern since its effect on utility companies is both economically and operationally challenging across the world. Both ML methods like SVM and Random Forest or DL models like LSTM and GRU have been explored for their capability to identify theft in modern smart meters. This review of the literature focuses on the techniques used for detecting and addressing the problem of electricity theft.

2.1 Traditional Approaches

According to Depuru, Wang, and Devabhaktuni (2011), electricity theft detection has traditionally used methods; manual meter reading, billing audits, rule-based threshold checks, and seal inspections. These methods, however, are increasingly out of touch with modern grid systems which are characterized by more complex and variable consumption patterns. Smith (2004) conducted a cross-country comparison and highlighted the manual

inspection, field visit, and sealing practices as cost inefficient and with widely varying effectiveness. Amin and Wollenberg (2005) emphasized rule-based anomaly detection systems struggle to adapt to evolving power delivery conditions. In their work, highlighting the need for more adaptive and scalable solutions.

2.2 Machine Learning

In Indonesia, Taruna, Arisona, Irwanto, Bestari, and Juniawan (2025) applied machine learning techniques to construct a framework for identifying electricity theft among postpaid residential users of the State Electricity Company (PLN). The research analyzed consumption data from subsidized 450 VA households monthly and tested multiple classifiers. RF showed the most stable performance with 0.89 accuracy, although KNN had higher recall performance. The authors proposed a sequential detection pipeline which synergistically improved capturing targeted detection success across multiple customers with varying histories of electricity theft and exhibited robustness to class imbalance and missing data, maintaining high accuracy with 25% of records missing. The inability to identify short-term theft patterns due to the use of monthly aggregated data and absence of sequence-based deep learning methods, which limit the ability to capture evolving consumption behaviors, are gaps addressed in the present study.

In the article published in 2023, Zidi et al. presented the benchmark dataset ETD2022 meant for the detection of electricity theft for smart meters with several classes. They derived the dataset from the Open Energy Data Initiative (OEDI) which consists of 16 types of consumers and includes six different scenarios of theft which are generated by a custom theft generator. The authors analyzed five machine learning classifiers: KNN, Decision Tree, RF, Bagging, and ANN. As highlighted in the paper, RF was the best performing algorithm with over 94% accuracy in the six-class scenario, demonstrating the model's efficacy even without consumer type details. The study pointed out as a drawback the lower accuracy for detection of ‘theft 6’ due to its similarity to normal consumption indicator.

2.3 Deep Learning Based Approaches

The work of Kocaman and Tümen (2020) applies deep learning to electricity theft detection using an LSTM network to capture temporal consumption patterns. The study aimed to capture sequential dependencies using hourly consumption data from 3,000 customers with smart meters over a year, a capability that most machine learning algorithms lack. The LSTM model not only achieved an accuracy of 92% but also outperformed DT and SVM in identifying both continuous and intermittent theft patterns. One of the key strengths of this work is the ability of the LSTM to learn from raw time series data without requiring extensive or manual feature engineering.

Using a Deep Neural Network (DNN) architecture, Naresh and Krishnachaitantya (2022) created a smart grid-based electricity theft detection framework employing deep learning techniques. The proposed system utilized smart meter data to train the model to detect electricity theft. The deep neural network outperformed traditional ML models like DT, SVM, and KNN with better accuracy suggesting it captures complex and nonlinear

relationships underlying the consumption data. The study also noted that the multi-layered structure of the DNN made it possible to automate the extraction of important features from the input data, thus decreasing the amount of manual preprocessing required.

2.4 Hybrid or Other Relevant Approaches

Haq et al. (2021) developed a hybrid framework for detecting electricity theft that incorporates feature extraction with deep learning techniques, followed by classification using traditional machine learning. The approach utilizes a CNN to automate feature extraction from electricity consumption time series, followed by classification with an SVM. The hybrid CNN–SVM model using smart meter readings from residential customers data. With this technique, the study achieved superior detection accuracy as compared to standalone CNN or SVM models. The study also verified the enhanced balanced classification across normal and theft cases, as shown by the improvements on both precision and recall. Nonetheless, the model evaluation was limited to a small dataset. The model's applicability to larger and more diverse populations of consumers was not examined.

Khan et al. (2021) designed a hybrid framework of big data analytics for electricity theft detection while incorporating advanced data preprocessing techniques alongside an SVM configured through Bayesian Optimization. The preprocessing pipeline included handling missing values, removing outliers, standardizing features, and addressing class imbalance using Synthetic Minority Oversampling Technique (SMOTE). Bayesian Optimization was also utilized for hyperparameter tuning during the SVM training for better performance during classification. Their evaluation on a broad smart meter dataset demonstrated that access to the framework outperformed baseline SVM rival the performance and detection rates of other classical and SVM-classifiers.

2.5 Synopsis of the Literature Review

Approaches	Authors	Year	Objective	Method	Evaluation Metrics	Findings
Traditional Approaches	Depuru et al.	(2011)	Review traditional electricity theft detection methods	Manual reading	-	Identified common methods; noted limits in scalability and adaptability
	Smith	(2004)	Compare anti-theft practices across countries	Manual inspections, field visits, sealing	-	Effectiveness varies; often costly and inconsistently applied
	Amin& Wollenberg	(2005)	Highlight limits of existing grid protection systems	Centralised control, rule-based systems	-	Existing systems slow and rigid; need adaptive, scalable solutions

Machine Learning	Hussain et al.	(2021)	Address class imbalance in theft detection	CatBoost+ SMOTE Tomek	Accuracy, Precision, Recall, F1-score	Accuracy 93%, Improved SVM accuracy via feature engineering
	Gunturi & Sarkar	(2021)	Compare ensemble methods for theft detection	RF, Extra Trees, AdaBoost, CatBoost, XGBoost	AUC, Accuracy, Precision, Recall, F1-score	Outperformed AdaBoost and decision trees
	Kawoosa et al.	(2023)	Detect region-specific theft patterns	RF, XGBoost, Logistic Regression, Decision Tree, Naïve Bayes	Accuracy, Precision, Recall, F1-score, DR	High precision for localized theft
	Appiah et al.	(2023)	Optimise theft detection with hyperparameter tuning.	ExtraTrees, SMOTE Tomek, GridSearchCV	Matthew's Correlation Coefficient (MCC)	High accuracy with minimal overfitting
	Taruna et al.	(2025)	Address data imbalance in traditional meters	RF, KNN, Logistic Regression, Decision Tree, Naïve Bayes	Accuracy, Precision, Recall, F1-score	Random Forest showed best performance for postpaid residential customers, while KNN underperformed in certain cases
Deep Learning	Takiddin et al.	(2022)	Detect anomalies in electricity theft cyberattacks.	Deep Autoencoder + Isolation Forest	ROC-AUC, Accuracy, Precision, Recall, F1-score	Achieved high recall and AUC, showing strong sensitivity to adversarial attacks
	Pamir et al.	(2022)	Reduce false positives in theft detection	Autoencoder+BiGRU	AUC, Accuracy, Precision, Recall, F1-score	FPR 10,2%, Focused on theft-specific anomalies
	Naeem et al.	(2023)	Enhance temporal feature extraction	DenseNet-FCN + GRU+ LightGBM+ ROBC	AUC-ROC, Accuracy, Precision, Recall, F1-score,PR-AUC	Achieved AUC-ROC of 0.92 and PR-AUC of 0.90, demonstrating high precision in dynamic environments and improved generalization on imbalanced datasets
	Eddin et al.	(2022)	Optimise RNN for theft attacks in the generation domain	Fine-tuned RNN	Accuracy, Precision, Recall, F1-score, DR	Robust against zero-day attacks
	Duarte Soares et al.	(2022)	Classify theft using a hybrid GRU-CNN	BiGRU-CNN	AUC, Accuracy, Precision, Recall, F1-score	Resistant to unbalanced data

	El-Toukhy et al.	(2023)	Adapt to dynamic theft patterns.	Deep Reinforcement Learning (DQN/DDQN, CNN+GRU)	FAR,FNR	Effective against evolving attacks
	Munawar et al.	(2022)	Address temporal patterns in theft data	BiGRU + BiLSTM	AUC, Accuracy, Precision, Recall, F1-score	Achieved high accuracy and AUC (0.93) with reduced false positive rate compared to existing DL models
	Lepolesa et al	(2022)	Detect electricity theft using smart meter data	DNN, PCA, mRMR, and hyperparameter optimisation	AUC, Accuracy, Precision, Recall, F1-score	Achieved 97% AUC and 91.8% accuracy; combined features outperformed time-domain alone
	Chen et al	(2023)	Detect electricity theft using smart meter data	Focal Loss-based 1D Densely Connected CNN (FLB-DCNNs) with feature-engineered indicators	Precision, Recall	Demonstrating high accuracy and robustness for real-world theft detection
	Fährmann et al.	(2024)	Review anomaly detection in smart environments	CNN, RNN, Autoencoder	-	DL methods promising for dynamic smart meter data; scalability & explainability challenges
Hybrid/Other	Bahrami et al.	(2024)	Combine rule-based and ML for interpretability	Rule-based ML (RML)	Precision, Recall, F1-score	Explainable but inflexible to new patterns
	Zheng et al.	(2018)	Detect theft without labelled data	MIC + CFSFDP clustering	AUC, MAP@20	Combined correlation-based and clustering-based methods for improved theft detection in various tampering scenarios
	Nazmul et al.	(2019)	Incorporate hardware features for theft detection	CNN-LSTM	Accuracy, Precision, Recall, F1-score	Effective for imbalanced datasets
	Haq et al.	(2021)	Hybrid theft detection using deep learning and SVM	CNN + SVM	ROC-AUC, Accuracy, Precision, Recall, F1-score, MCC	Robust performance with noisy data
	Ullah et al.	(2021)	Combine deep learning with smart meter antennas	Hybrid CNN-PSO-GRU	AUC, Accuracy, Precision, Recall, F1-score	Improved detection using hardware features
	Louv & Bokoro	(2019)	Detect and mitigate electricity theft in	Combination of smart meter data analysis	-	A hybrid detection and mitigation

			South Africa	and alternative mitigation strategies		technique tailored in South Africa

Table 1 Synopsis of the literature review

2.6 Limitations & Research Gaps

Electricity theft detection systems in prior share similar common limitations. Some studies have used on consumption data gathered during long time intervals. Identifying minute-scale or short-term theft is more challenging for utility providers. Taruna et al. (2025) used aggregated (monthly) consumption data in their study. This is limiting the ability to detect theft that happens in shorter time frames. In contrast, this study uses hourly smart meter data into weekly windows. With this way ,easier the detection of more subtle and short-term theft patterns.

Another relevant work is by Kocaman and Tümen (2020), who applied LSTM-based deep learning models. While effective in capturing temporal dependencies, the absence of real theft labels was a major limitation. In this study, synthetic labelling was also applied because of the absence of real theft data. Their approach focused mainly on detecting theft as a binary classification problem, without covering diverse theft scenarios.

A benchmark dataset that includes consumption of electricity, gas, and other facility-level sensors like fans, cooling, heating, and interior lights was introduced by Zidi et al., (2023). This type multimodal data are rarely available to utility providers.

In summary, there is still a considerable gap in the existing literature. The limited use of high-resolution datasets constrains the identification of short-term theft behaviours. In addition, many previous approaches use on extra features which are not readily available to electric utility companies. This study addresses these issues by using hourly smart meter data with weekly windows and using only consumption data to guarantee useful feasibility in real-world utility scenarios. Also, with constructing a balanced dataset, the research ensures a just environment where both machine and deep learning approaches can be evaluated with fairness.

3 Methodology

The overall workflow of the proposed methodology is illustrated in Figure 2. It summarizes the main steps of the study. It including data preprocessing, theft scenario generation, dataset structuring, model training, and evaluation. This diagram provides an overview of how raw smart meter data is transformed and analyzed to detect electricity theft.

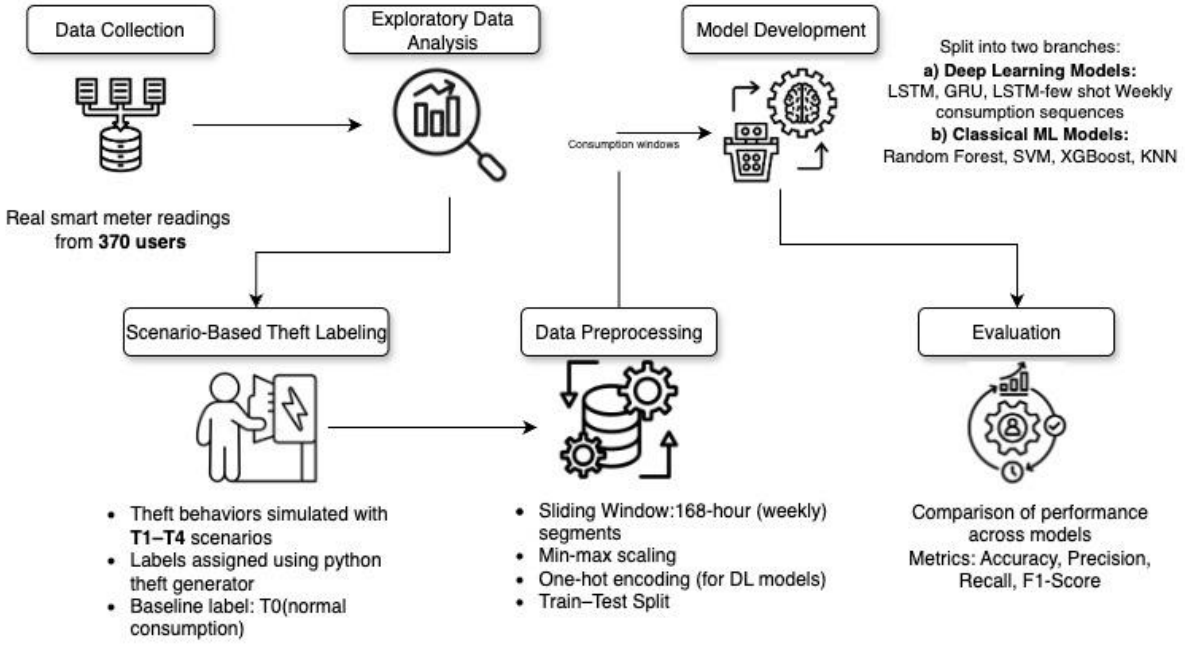


Figure 2 Research Methodology

This study utilized systematic procedures to examine ML and DL models in the context of detecting energy theft through smart meter data analysis. The research followed a systematic method which it comprised of preparing the data, generating synthetic scenarios, model building, and performance analysis. The study process was designed to facilitate a direct comparison between classical and deep learning approaches, ensuring the research was reproducible and scientifically sound.

The data of this research was sourced from the Electricity Load Diagrams 2011-2014 dataset of the UCI Machine Learning Repository (Trindade, 2015). This contained active power consumption data of 370 customers per 15 minutes over four years. The study was founded on 4 years of data, sufficient to analyse in terms of time to increase the efficiency in the computations, as summarized in Table 2.

Item	Value
Total number of raw instances	12,264,000 (15-min intervals, 2011–2014)
Number of users (after cleaning)	350
Number of features	3 (timestamp, customer id, consumption)
Window size	168 hours (weekly)
Total number of windows	508,000
Train/Test split	80/20 (101,780 test windows)

Table 2 Summary of Dataset Statistics

To support supervised learning, the research applied a synthetic labelling strategy to imitate four different electricity theft scenarios. These scenarios were based on the patterns recorded concerning meter tampering and unauthorized consumption. The manipulations were synthetic and included uniform consumption reduction, selective period reduction (only weekends), random zeroing of measurements, and the flattening of the whole curve. The detailed description and calculation logic of these scenarios are summarized in Table 3. All the manipulations were done with python generator so that they could be consistent and reproducible. The non-manipulated data windows were maintained as the foundation slate of no theft. In this method, a well-balanced dataset emerged with distinct classes to train and test the model.

Scenario	Description	Label
T0	Original consumption pattern (no theft)	0
T1	Flat reduction across all measurements by a random factor between 30% and 80%	1
T2	Reduction during weekends only by a random factor between 30% and 80%	2
T3	Randomly set between 5 and 20 hourly values to zero within the window	3
T4	Mean flattening, all values set to the mean consumption of the window	4

Table 3 Theft Scenarios and Labels

To ensure equity in model training and evaluation, the synthetic scenarios were distributed in a balanced manner. It can be seen in Figure 3 that the dataset is constructed in a way that the majority class corresponds to the normal consumption, which is T0, while each theft scenario labelled T1 to T4, is equally represented at 12.5% each. This balanced distribution was designed to prevent bias towards any theft type and to enable consistent comparative analysis across models.

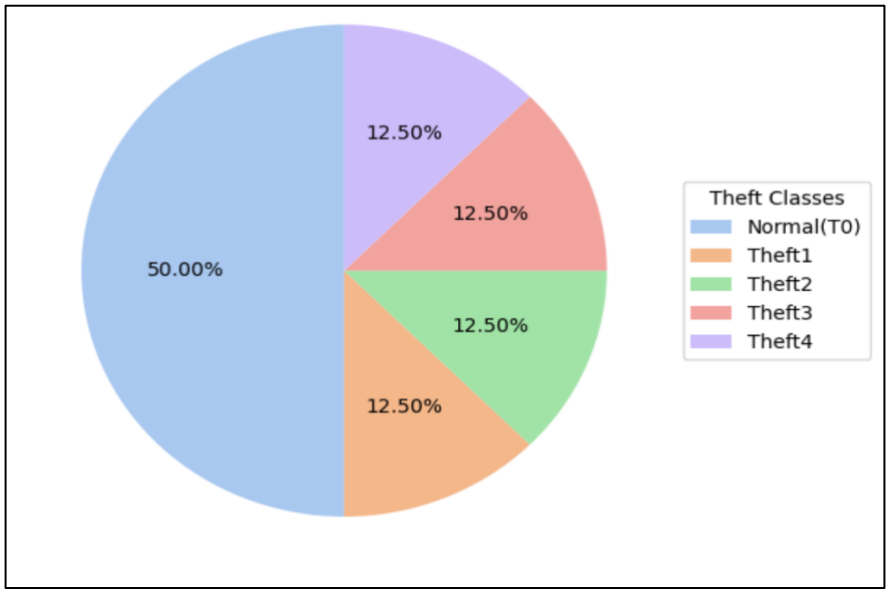


Figure 3 Distribution of Theft Scenarios

It involved several essential steps in the preprocessing and transformation stages. To begin with, the customers with incomplete records removed, which results in the set of 350 users. The raw dataset, originally in wide format, was converted into a long format with explicit identifiers for each customer. To reduce variability across different consumption ranges, normalization was applied: deep learning models used min-max scaling, while machine learning models used standard scaling. For categorical labels (theft scenarios), one hot encoding was applied. To prepare the data for temporal learning, a sliding window approach was used. Each window contained 168 hourly values (7 days) and a 24-hour stride between the subsequent windows which yields a total of 508,000 weekly segments. This transformation provided adequate temporal coverage and generated structured sequences for both ML and DL approaches.

The modeling step incorporated both DL and ML approaches to assess electricity theft detection in all five scenarios. For the DL branch, three models were implemented: Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), and few-shot LSTM. LSTM networks are widely recognized for their capability to capture long-range temporal dependencies in sequential energy data (Kocaman & Tümen, 2020). GRU offers a computationally efficient alternative to LSTM, and recent work has demonstrated its strong capability in enhancing temporal feature extraction and generalization in dynamic and imbalanced environments (Naeem et al., 2023). Also, few-shot LSTM was built to understand how well temporal models perform under data scarcity, a common situation in the market where theft indicators are infrequently available. As for all DL models, the recurrent layers were followed by batch normalization, dropout regularization, and dense layers for multi-class classification.

In the ML branch, the Random Forest, Support Vector Machines (SVM), Extreme Gradient Boosting (XGBoost), and K-Nearest Neighbors (KNN) models were all applied. SVM was selected due to its proven efficiency in margin-based classification and its prior use in hybrid electricity theft detection frameworks (Haq et al., 2021). The inclusion of Random Forest and XGBoost was because of their ability to perform exceptionally strong with structured consumption datasets. In particular, Zidi et al. (2023) reported that Random Forest, which reported high overall accuracy, struggled with capturing subtle

variations in the theft cases. Lastly, KNN was implemented as a baseline distance-based classifier in this study, providing a simple reference point for evaluating the performance of more advanced models.

In this step, ML and DL models were trained on the same weekly consumption windows. Each DL model consisted of a recurrent layer, followed by batch normalization and dropout regularization, and then dense layers for classifying the final output. It provides to ensure a fair comparison under identical input conditions. Every model was evaluated based on Recall, Accuracy, F1-score, and confusion matrices. Confusion matrices were used not only for overall accuracy but also to analyse misclassifications across theft types. This approach enabled a systematic comparison. Also proving that sequence-based DL models capture temporal patterns and that classical ML techniques are efficient for structured consumption data.

4 Design Specification

A comprehensive model for detecting electricity thefts, combining ML and DL techniques built. The architecture a workflow that systematically addressed data preparation, model development, and evaluation. It utilized the adopted practices from past studies and introduced a new feature: generating synthetic data based on scenarios to improve model training. To obtain the dataset used, the module of data acquisition accessed Electricity Load Diagrams 2011-2014 data on the UCI Machine Learning Repository.

Module	Function	Key Operations
Data Acquisition	Source raw consumption data	Dataset retrieval, format conversion
Preprocessing	Prepare data for analysis	Cleaning, min max normalization, Standard Scaler, windowing
Scenario Generation	Create synthetic theft cases	Pattern manipulation, labelling
Model Training	Develop detection algorithms	Parameter optimization, validation
Evaluation	Assess model performance	Metric calculation

Table 4 Project Architecture Components

The preprocessing phase performed data cleaning, normalization, and windowing. The raw data, which was initially stored in a wide format with one column per customer. The data transformed into a long format containing three main features. Converting to long format ensured consistency across customers and simplified preprocessing. Also, it allowed the models to directly utilize customer-level information. The features are the timestamp, customer id and consumption value. Customers were filtered based on their readings in which the 15-minute readings were aggregated into hourly intervals. Then, additional users with inadequate or absent records were filtered out for 370 users. With this process yielding a total of 350 users. Normalization was applied to reduce variability across customers: min–max scaling for deep learning models and standard scaling for machine learning models. The categorical label of the theft scenarios was one-hot

encoded to enable multi-class classification. Ultimately, data was transformed into structured sequences, where the data was sliced into weekly segments of 168 hours with a stride of 24 hours. This resulted in about 508,000 structured sequences for further analysis. Time-series data was divided into 168 measurements. Within these windows, took a stride of 24 hours between consecutive windows. Each window was normalized in min-max so that the scaling stayed the same between various customers consumption patterns.

The scenario generation python program coded 5 different types of theft patterns by code manipulation of data. These patterns were uniform reduction of consumption by a random factor between 30% and 80% (T1), weekend-specific reduction by a random factor between 30% and 80% (T2), random insertion of zeros in 5 to 20 hourly slots (T3) and flattening of the entire curve to the mean consumption value (T4). All the manipulations carried out systematically to come up with labeled examples for supervised learning. The reference window used was the non-theft window, which was not altered (T0). This way, it was possible to test the detection abilities of the system in a controlled manner across various theft behaviours.

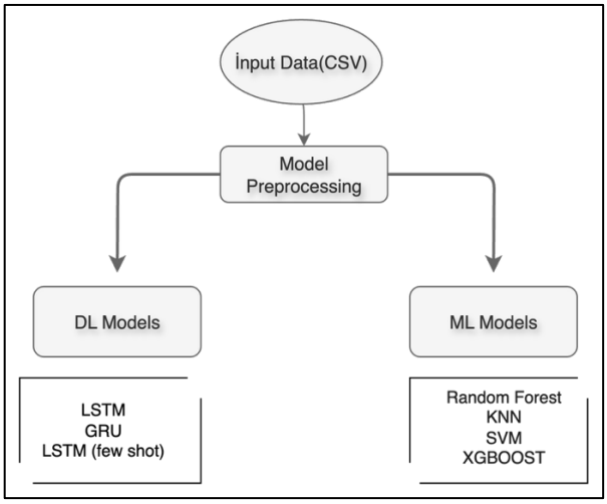


Figure 4 Research Models Overview

The modules of the model development included the ML and DL architectures. In the case of ML, the study utilized RF, SVM, KNN and XGBoost algorithms. The input features in the models were assumed to be flattened window vectors. The DL technique utilized recurrent neural network variations, including LSTM and GRU networks. A consistent structure of DL models was encompassed by having recurrent layers, dropout regularization, and dense layers of classification.

The design incorporated several technical challenges in detecting electricity theft. The sliding window methodology was able to capture the time-based consumption patterns and, at the same time, was computationally efficient. The modular structure enabled the assessment of different detection strategies in isolation, in a comparable way with the same data and assessment metrics. Each design choice recorded so that reproducibility and further extension of the framework could be considered.

5 Implementation

The presentation stage involved the use of machine and deep learning to develop an electricity theft detection system based on smart meter data. It was based on the Python language, and development was performed with the help of Jupyter Notebooks using the Anaconda. The preprocessing of data in terms of implementation was initiated by formatting the raw consumption records in the form of a time series. Min-max normalization applied to scale the values of consumption within the range of 0-1 by customer to cover baseline usage variability. A sliding window was used to generate input samples, forming 168-hour samples with 24-hour steps between successive windows. In this procedure, the subset of 2014 data was used to generate 12,540 total samples. The training data were generated using the programmatically implemented synthetic theft scenarios and transformed normalized windows.

Initially, the dataset included 370 customers and 140,256 timestamps which were 15-minute intervals readings 4 years. After the cleaning process, the number of customers was reduced to 350. After the dataset was transformed to a long format, it had around 49 million entries with three features: timestamp, customer id, and consumption. After implementing a sliding window approach with 168-hour windows and 24-hour strides, a total of 508,000 sequences were obtained. Each window was encoded as [168, 1] for deep learning models and flattened to [168] for machine learning models. The final train/test split yielded 406,220 training windows and 101,780 test windows.

After preparing the dataset and generating the structured sequences, each model was implemented with its own set of hyperparameters. The DL branch comprised an LSTM, a GRU, and a Few-shot LSTM network, whereas the ML branch contained KNN, Random Forest, SVM, and XGBoost. The key parameters and training times of all models are summarized in Table 5. With this table, providing a clear overview of the design specifications and computational costs for each method.

Approach	**Model**	**Key Parameters / Hyperparameters**	**Training Time(min)**
Deep Learning	LSTM	LSTM units=64, LSTM units=32, Dropout=0.3, Dense=64 (ReLU), Output=5 (Softmax), Optimizer=Adam	117.58
	GRU	Units=64, Dropout=0.3, Dense=32 (ReLU), Output=5 (Softmax), Optimizer=Adam	28.46
	LSTM (few shot)	Same as LSTM, trained with few-shot learning approach	1.51
Machine Learning	KNN	n_neighbors=5, metric='minkowski', weights='uniform'	0. 60

	RF	n_estimators=100	17.39
	SVM	C=1.0, gamma='scale', Kernel='rbf',	695.39
	XGBoost	-	122.30

Table 5 Models and Key Hyperparameters with Training Times

As highlighted in Table 5, DL models required longer training periods. Especially, the LSTM being the most computationally demanding. This is attributed to LSTM's more sequential architecture, which is more computationally expensive. GRU showed shorter training times compared to LSTM, reflecting its simplified gating structure. he few-shot LSTM variant was the fastest among DL models as it used a reduced training setup. Among the ML models, RF and XGBoost trained in a reasonable timeframe while SVM's computational expense was more pronounced because of employing kernel-based optimization. KNN was the overall fastest because of its simple, distance-based approach.

The module used evaluation of performance, measuring it according to standard classification metrics. Each of the model-scenario pairs was assessed with accuracy, Weighted F1, and Macro F1. The entire system has been coded in Anaconda Jupyter Lab 3.12, with the deep learning aspects being coded in TensorFlow 2.8 and classical machine learning aspects using scikit-learn 1.2.

The overall data flow of the proposed system, previously illustrated in Figure 2, begins with data collection from smart meters. Then followed by the generation of theft scenarios to label the dataset. In next, preprocessing steps (normalization, windowing etc.) are applied. In the last steps, ML and DL models were employed for training and evaluation. This integrated workflow ensures that every step of the pipeline is connected, from data acquisition to model assessment.

6 Results & Evaluation

The review was performed to compare seven models based on their capability to detect synthetic electricity theft scenarios. The results of the performance indicators showed, there were significant differences between the DL and classical ML methods.

Figure 5 provides a detailed overview of the performance of all models. Their results are reported with three measures: Accuracy, Macro F1, and Weighted F1. LSTM and GRU models performed best among the deep learning techniques, achieving over 93% across all metrics. A few-shot LSTM lower accuracy due to data scarcity simulation. On the ML side, RF and XGBoost also performed competitively, achieving over 93% accuracy. In the other hand, SVM and KNN demonstrated relatively lower performance. This consolidated figure, along with the rest of the figures, provides a direct comparison across all models which serves to demonstrate the effectiveness of ML and DL techniques in the detection of electricity theft.

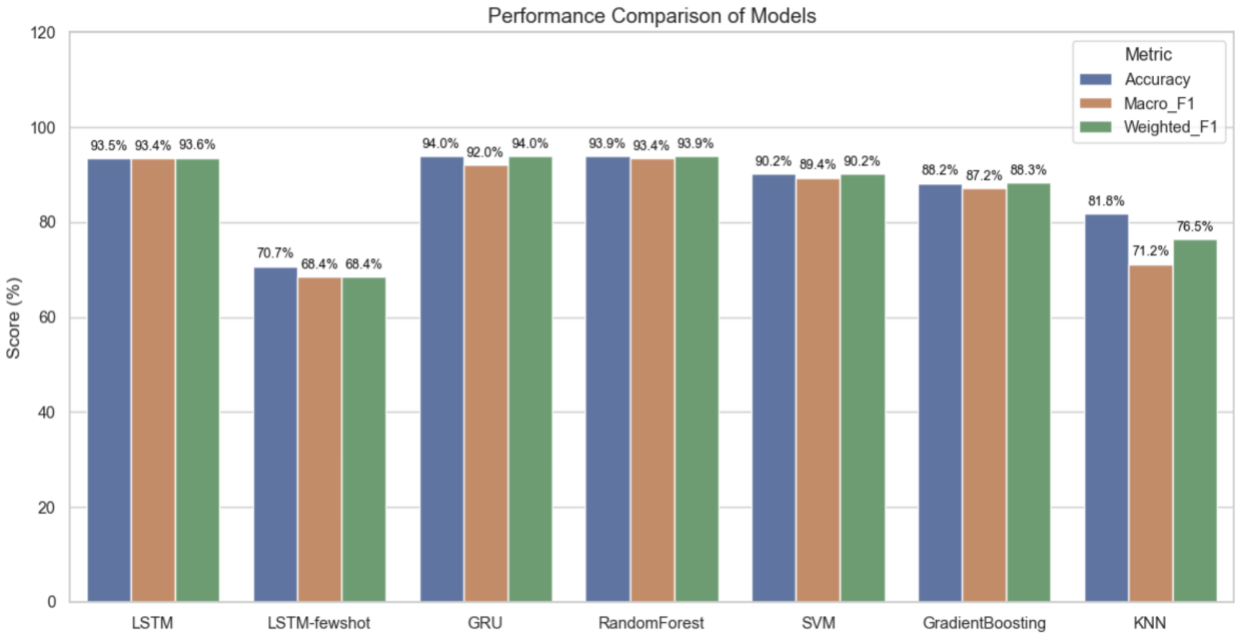


Figure 5 Overall performance comparison of models

According to the experimental results, the different ML and DL approaches have been shown to have different performances in the detection of electricity theft. The accuracy of the LSTM model 94%, which also proves the earlier results regarding the use of this model in the field of temporal pattern recognition. The model, however, computationally demanding and required over 6 times more training than Random Forest. This is a limitation about complex neural network deployment issues.

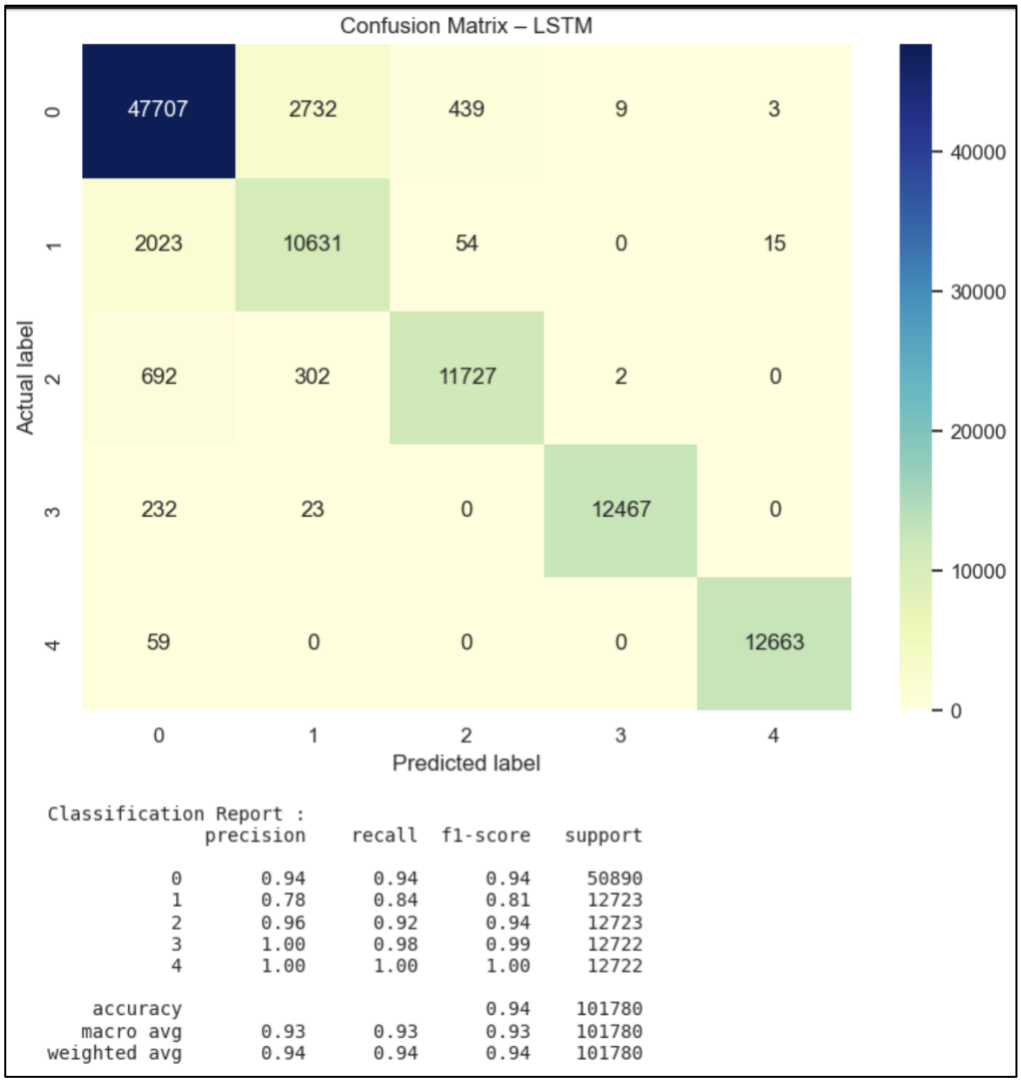
Random Forest is efficient in its computation achieving 94% accuracy, which is on par with LSTM given equal information inputs. Its ensemble architecture enabled effective capturing of the varying consumption behaviours and was more efficient relative to deep learning models like LSTM. While RF showed limited performance in T3 recall as compared to LSTM, indicating insensitivity to irregular patterns, this suggests RF works best in more controlled environments with clear and consistent consumption outliers. it may require feature engineering or integration with temporal models to improve detection of subtle, irregular theft patterns. The results of the GRU model 94% accuracy confirmed the adoption of it as an alternative lightweight variant of LSTM.

ML models yielded a maximum accuracy achievement of 94% using the Random Forest algorithm. This model demonstrated a high level of generalization for most of the theft instances as well as cost-effective computation. Another correctly performing model with an accuracy of 88% XGBoost. This model is noted for its high performance with structured data, thus its good performance on the consumption data for the ML model expected. The performance of SVM and KNN models was 90% and 82% respectively, which is considered a lack of performance when compared to the other models. Both models are noted to perform lower on irregular, high-dimensional data, which suggests the need for feature engineering and/or dimensionality reduction. The above results can be interpreted as ensemble ML methods noted to perform well on a variety of theft detection problems. KNN and SVM, which are known as distance and margin-based classifiers, need further refining to appropriately adapt to complex situation designs.

The following subsections present results and observations for each model.

6.1 Experiment 1- LSTM

The LSTM model achieved an accuracy of 94%, yielding an excellent result. This was particularly evident in the weekend-specific theft (T2: 92% recall) and random zeroing condition (T3: 100% precision). The confusion matrix revealed that minimal misclassification happened across different categories of theft, confirming its ability to detect temporal patterns in consumption. Regularization with dropout (0.3) also effective in avoiding overfitting, as there was less than 1% divergence between the training and validation loss.



Classification Report :

	precision	recall	f1-score	support
0	0.94	0.94	0.94	50890
1	0.78	0.84	0.81	12723
2	0.96	0.92	0.94	12723
3	1.00	0.98	0.99	12722
4	1.00	1.00	1.00	12722
accuracy			0.94	101780
macro avg	0.93	0.93	0.93	101780
weighted avg	0.94	0.94	0.94	101780

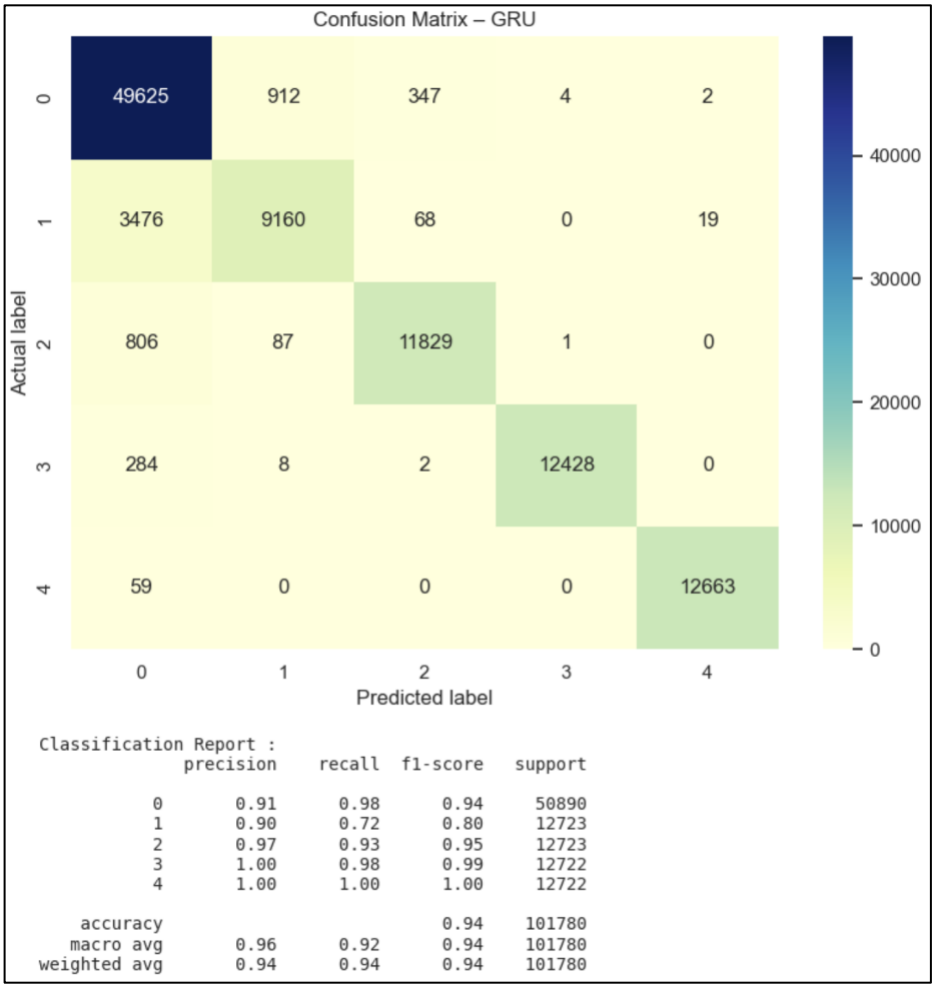
Figure 6 LSTM Result

The confusion matrix shows, the LSTM model shows a consistent level of performance throughout all theft classes with the highest precision for T3 and perfect recall for T4. The observed misclassification is primarily between the non-theft (T0) and the weekend-specific theft (T1) classes. These classes share some similarities in terms of consumption

behaviours. Overall, the distribution of correct predictions across the diagonal confirms the model's strong capability in distinguishing temporal patterns in smart meter data.

6.2 Experiment 2-GRU

Although the GRU model achieved an accuracy of 94%, which is matching the performance of LSTM, it had the advantage of a shorter training time. Strong performance was observed in the flat reduction case, T1 where precision was 0.90 and recall was 0.72, as well as the random zeroing case T3 where precision was 1.00 and recall was 0.98. The complete flattening scenario, T4 where precision and recall were 1.00, also demonstrated high GRU precision and recall. These results suggest that the GRU model can be used to detect most theft patterns with a high degree of reliability, making it appropriate for real-time systems where fast inferences are critical.



Classification Report :

	precision	recall	f1-score	support
0	0.91	0.98	0.94	50890
1	0.90	0.72	0.80	12723
2	0.97	0.93	0.95	12723
3	1.00	0.98	0.99	12722
4	1.00	1.00	1.00	12722
accuracy			0.94	101780
macro avg	0.96	0.92	0.94	101780
weighted avg	0.94	0.94	0.94	101780

Figure 7 GRU Result

6.3 Experiment 3-LSTM (a few shot)

The few-shot LSTM model was designed for evaluation under the circumstances where training samples are limited, simulating situations where labeled theft data is scarce. To

mitigate overfitting, an underfitting model was constructed comprising two LSTM layers with 32 and 16 units and a dropout regularization of 0.3. The model achieved an overall accuracy of 75% with class-dependent variance in performance. For the T4 theft detection performance, the precision and recall reached maximum levels 1.00 each, while detection of the original consumption pattern T0 remained difficult 0.74 precision and 0.57 recall. These results highlight the difficulty of achieving consistent detection across all theft scenarios under limited training data conditions, even though performance improved compared to the previous low-data setup.

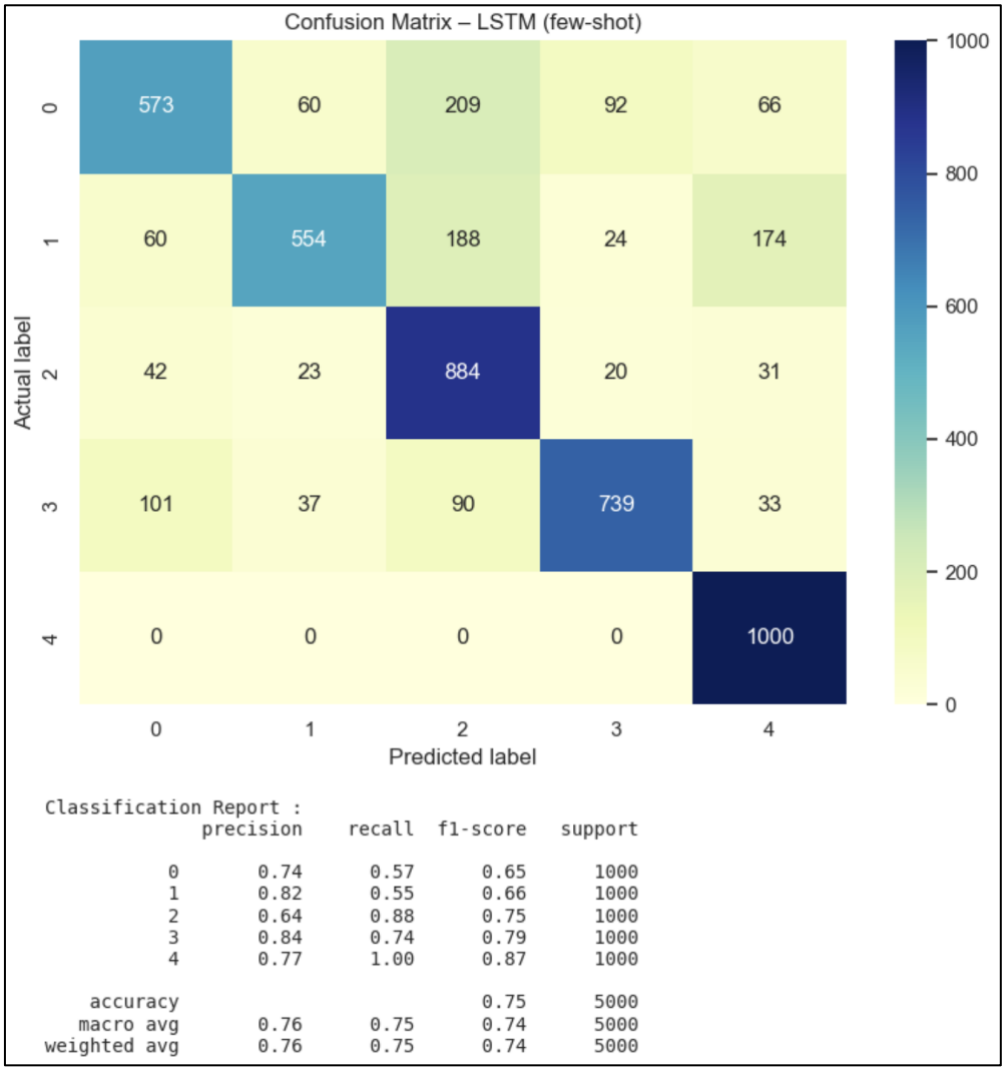
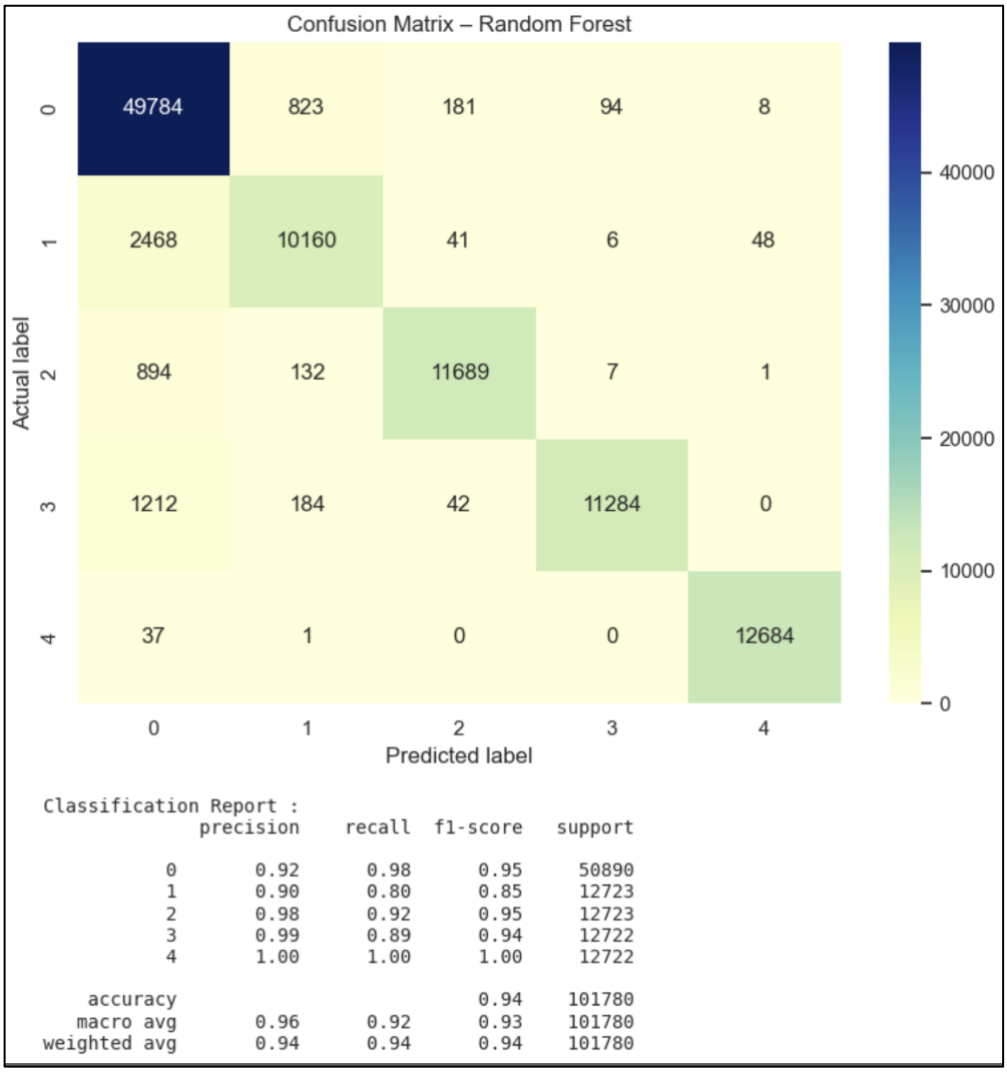


Figure 8 LSTM(few shot) Results

6.4 Experiment 4- Random Forest

Random Forest demonstrated 94.0% accuracy, which is the same level as LSTM but with less-intensive computation. In every situation, the recall value of the model was more than 90%, and the generalization demonstrated by the model also robust. The model showed strong precision across all theft types, reaching 1.00 for T4 and 0.99 for T3, while maintaining high recall for most classes.



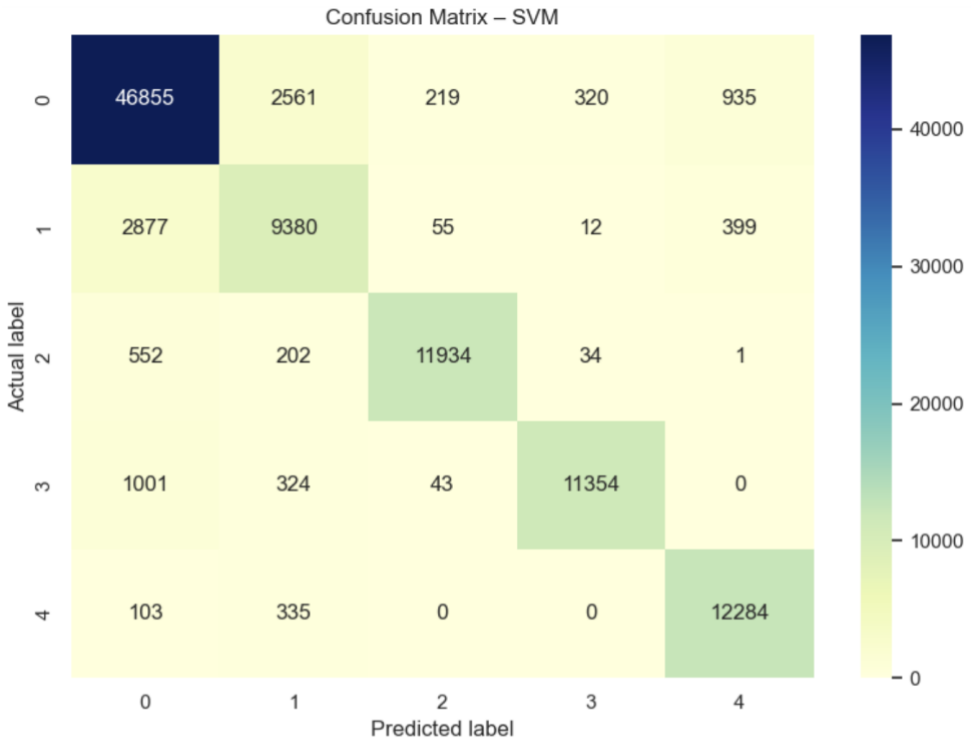
Classification Report :

	precision	recall	f1-score	support
0	0.92	0.98	0.95	50890
1	0.90	0.80	0.85	12723
2	0.98	0.92	0.95	12723
3	0.99	0.89	0.94	12722
4	1.00	1.00	1.00	12722
accuracy			0.94	101780
macro avg	0.96	0.92	0.93	101780
weighted avg	0.94	0.94	0.94	101780

Figure 9 Random Forest Results

6.5 Experiment 5- SVM

The SVM model achieved a 90% accuracy rate which is quite high. Additionally, the model showcased adequate precision and recall for most theft types. It mostly excelled at identifying the weekend-specific reductions “T2” Precision: 0.97, Recall: 0.94 and the random zeroing behaviours “T3” Precision: 0.97, Recall: 0.89. However, the model was much weaker at capturing the flat reduction behaviours T1 which had precision and recall of 0.73 and 0.74 respectively. Conclusions based on these findings indicate that while SVM is quite adept at differentiating theft patterns, it is rather oblivious to uniform changes in consumption over a period.

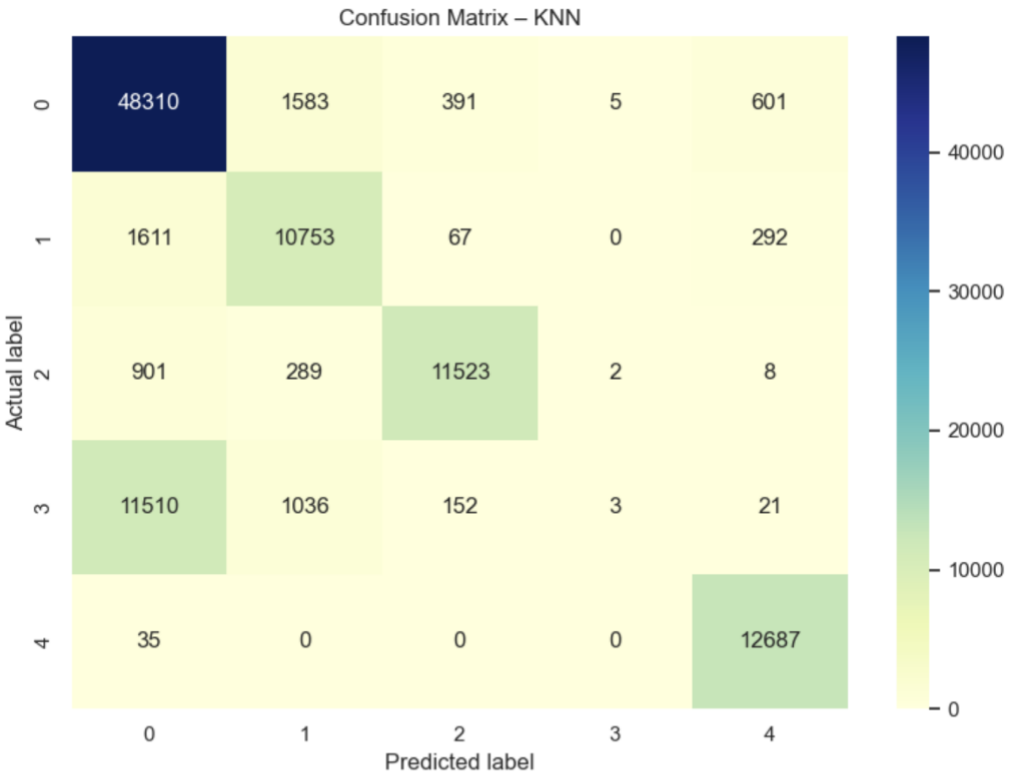



```
Classification Report :
              precision    recall  f1-score   support

           0       0.91      0.92      0.92     50890
           1       0.73      0.74      0.73     12723
           2       0.97      0.94      0.96     12723
           3       0.97      0.89      0.93     12722
           4       0.90      0.97      0.93     12722

    accuracy                           0.90    101780
   macro avg       0.90      0.89      0.89    101780
weighted avg       0.90      0.90      0.90    101780
```

Figure 10 SVM Results

6.6 Experiment 6-KNN

The KNN model achieved an accuracy level of 82%. The performance varied significantly across different theft scenarios. The model showed high recall in identifying non-theft activities T0: 0.95 and mean flattening patterns T4: 1.00. Unfortunately, showed limited performance on random zeroing T3, where recall dropped to 0.00 and precision to 0.30. This suggests that the KNN model has better recall to the more pronounced and distinctive patterns but is quite limited in the detection of theft behaviours that closely align with ordinary spending patterns, highlighting a key limitation in its applicability to subtle anomalies.



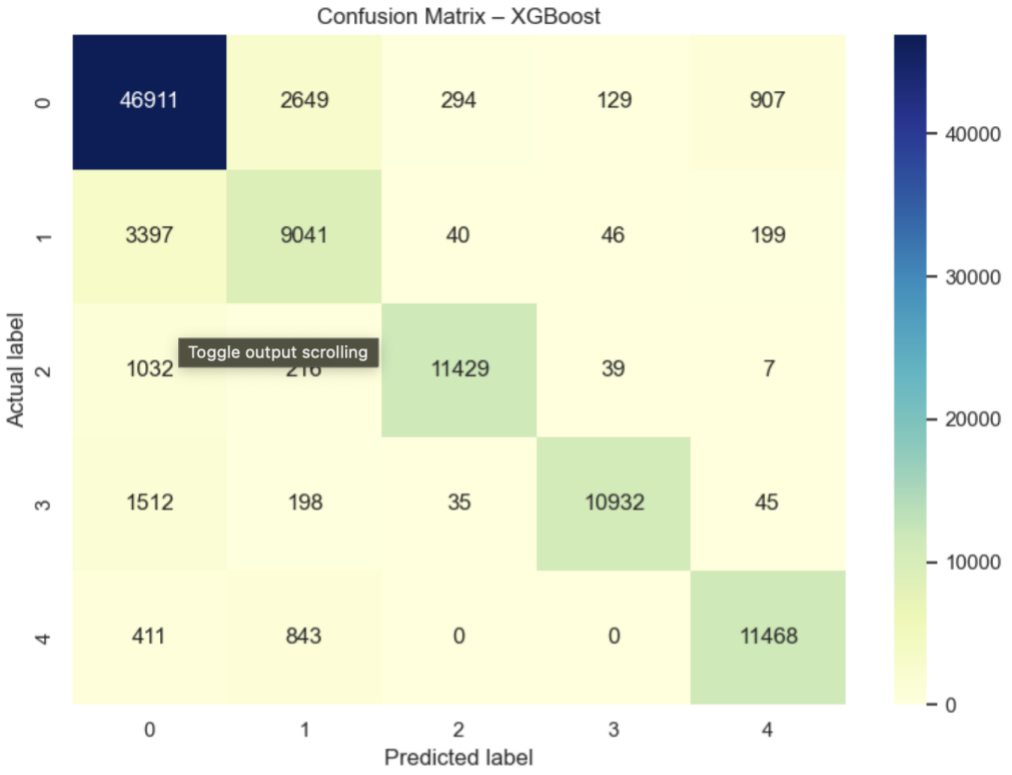
Classification Report :

	precision	recall	f1-score	support
0	0.77	0.95	0.85	50890
1	0.79	0.85	0.82	12723
2	0.95	0.91	0.93	12723
3	0.30	0.00	0.00	12722
4	0.93	1.00	0.96	12722
accuracy			0.82	101780
macro avg	0.75	0.74	0.71	101780
weighted avg	0.76	0.82	0.76	101780

Figure 11 KNN Results

6.7 Experiment 7-XGBoost

The XGBoost model attained an accuracy rate of 88%. The model did well in identifying mean flattening T4: recall 0.90 and weekend only reduction patterns T2: precision 0.97, although it showed limited performance in flat reduction T1: precision 0.70, recall 0.71. Overall, the model performed well in most categories; however, the lower macro recall (0.86) indicates that some theft patterns characterized by more subtle changes in consumption may still be difficult to detect.




```
Classification Report :
              precision    recall  f1-score   support

           0       0.88      0.92      0.90     50890
           1       0.70      0.71      0.70     12723
           2       0.97      0.90      0.93     12723
           3       0.98      0.86      0.92     12722
           4       0.91      0.90      0.90     12722

    accuracy                           0.88    101780
   macro avg       0.89      0.86      0.87    101780
weighted avg       0.88      0.88      0.88    101780
```

Figure 12 XGBoost Results

6.8 Discussion

This section discusses the findings of the study, highlighting strengths and limitations in detecting electricity theft. The purpose here is not only to present results. Also to critically compare the approaches and drawing insights beyond raw accuracy numbers. The analysis shows that GRU reached performance comparable to LSTM while requiring substantially less training time. This is making GRU a practical choice for real-time applications. Random Forest matched LSTM in accuracy but exhibited weaker recall for irregular theft scenarios (random zeroing etc.). It suggesting that it is more suited for environments with clear and consistent anomalies. XGBoost also performed competitively but lacked the ability to capture temporal dependencies as effectively as recurrent networks. In contrast, SVM and KNN lagged, reflecting the difficulty of distance- and margin-based classifiers in handling irregular, high-dimensional consumption data. These comparisons underline the trade-offs between DL and ML methods. In particular, DL excels in pattern-heavy in-depth analysis and ML models being more attractive for cost-efficient deployment.

The handling of class imbalance is an area that needs improvement. Although synthetic scenarios led to balanced data, it would be necessary to use superior methods in the real

world. The corresponding evaluation metrics put much emphasis on precision without adequate attention to false favourable rates, which are of paramount importance in the case of utility companies. The measurements of the computational efficiency indicated the practical aspects of deployment. The inference speed of RF is also lower, and this allows its use in real-time systems. In contrast, LSTM has a more elaborate pattern recognition and is more applicable to in-depth analysis.

There were several design limitations in the experiments. The method of synthetic data generation, though required because of an absence of real theft cases, will not be able to take into consideration all the actual tampering methods. The sliding window (168 hours) was not optimized, which could have resulted in the short-term theft patterns being overlooked. The concept drift was not considered, as it is characteristic of real smart meter data.

Experiments have shown that there is no single solution applicable to all theft detection needs. The decision whether to use DL or ML is subject to certain operational limitations, which are reminiscent of the trade-offs outlined (Gopalan et al., 2025). Future research should address the limitations raised and maintain the high-quality evaluation standard established in this study.

7 Conclusion and Future Work

This research successfully answered the central question of how ML and DL techniques can be effectively used to detect electricity theft on smart meter data. The study has proven that using modern techniques, it is possible to reach a high detection accuracy with the proper setup using systematic experimentation with both classical and neural network techniques. The three key points achieved by the work were the extensive comparison of the models performance, the analysis of the potential capabilities of detection in various theft scenarios, and the determination of the practical aspects of real-life implementation.

The conducted experiment provided several significant findings that contribute to the development of the area of electricity theft detection. The LSTM and GRU proved to be the most accurate ones with a 94 % detection rate, especially at detecting the complex patterns in the consumption data. The above performance was nevertheless equalled by Random Forest classifiers, which were also far more computationally efficient, a fact that might render them better suited to large-scale implementation. The paper also highlighted the drawbacks of more basic models. For instance, KNN models showed limited performance when scaled to multi-class detection problems. The results can be helpful for utilities considering the installation of automated theft detection systems.

This work produced several applicable concepts for both practical and theoretical research. Reliably high results from temporal deep learning models suggest that they have the potential to be applied to analyse time-series data from smart meters. Moreover, the effectiveness of ensemble methods demonstrates that hybrid models provide an optimal balance between accuracy and efficiency. Ensemble approaches, when properly configured, yield the best performance. Still, the study has important and specific limitations that need to be solve. Particularly the use of synthetic data, which does not

reflect the real-world behaviour of thefts. Fixation on the windowing method applied during the preprocessing stage can miss the detection of short-term anomaly theft signals.

Three key improvements could enhance future work:

1. Hybrid Approach: Combining Random Forest's efficiency with LSTM's temporal awareness could yield better results, as hinted by Munawar et al. (2021).

2. Adaptive Windowing: Implementing variable window sizes based on consumption patterns might improve the detection of short-term theft.

3. Real Data Validation: Testing models on verified theft cases would provide more reliable performance estimates.

As a summary, to begin with, a hybrid model can be a solution to better performance, that is, a combination of the advantages of various approaches. For instance, the LSTM model as a feature extractor and the Random Forest as a classifier. Second, ensuring the practical relevance of findings by involving utility companies in the validation of findings on real-world cases of theft would enhance the findings. Lastly, exploring adaptive methods to manage concept drift in consumption patterns would be beneficial in making the model more resilient to long-term changes. The commercial possibilities include developing lightweight edge computing systems for real-time analysis or creating a theft risk scoring system that prioritizes field inspections. The paper makes meaningful contributions to electricity theft detection, primarily by identifying areas that require further research. application in future studies.

References

Appiah, S. Y., Akowuah, E. K., Ikpo, V. C., & Dede, A. (2023). Extremely randomised trees machine learning model for electricity theft detection. *Machine Learning with Applications*, *12*, 100458. https://doi.org/10.1016/j.mlwa.2023.100458

Bahrami, S., Yumuk, E., Kerem, A., Topçu, B., & Kaya, A. (2024). Electricity Theft Detection Using Rule-Based Machine Leaning (rML) Approach. *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji*, *12*(2), 438–456. https://doi.org/10.29109/gujsc.1443371

Chen, J., Nanehkaran, Y. A., Chen, W., Liu, Y., & Zhang, D. (2023). Data-driven intelligent method for detection of electricity theft. *International Journal of Electrical Power & Energy Systems*, *148*, 108948. https://doi.org/10.1016/J.IJEPES.2023.108948

Depuru, S. S. S. R., Wang, L., & Devabhaktuni, V. (2011). Smart meters for power grid: Challenges, issues, advantages and status. *Renewable and Sustainable Energy Reviews*, *15*(6), 2736–2742. https://doi.org/10.1016/J.RSER.2011.02.039

Eddin, M. E., Albaseer, A., Abdallah, M., Bayhan, S., Qaraqe, M. K., Al-Kuwari, S., & Abu-Rub, H. (2022). Fine-Tuned RNN-Based Detector for Electricity Theft Attacks in Smart Grid Generation Domain. *IEEE Open Journal of the Industrial Electronics Society*, *3*, 733–750. https://doi.org/10.1109/OJIES.2022.3224784

El-Toukhy, A. T., Badr, M. M., Mahmoud, M. M. E. A., Srivastava, G., Fouda, M. M., & Alsabaan, M. (2023). Electricity Theft Detection Using Deep Reinforcement Learning in Smart Power Grids. *IEEE Access*, *11*, 59558–59574. https://doi.org/10.1109/ACCESS.2023.3284681

Fährmann, D., Martín, L., Sánchez, L., & Damer, N. (n.d.). *Anomaly Detection in Smart Environments: A Comprehensive Survey*. https://doi.org/10.1109/ACCESS.2023.0322000

Gunturi, S. K., & Sarkar, D. (2021). Ensemble machine learning models for the detection of energy theft. *Electric Power Systems Research*, *192*, 106904. https://doi.org/10.1016/J.EPSR.2020.106904

Haq, E. U., Huang, J., Xu, H., Li, K., & Ahmad, F. (2021). A hybrid approach based on deep learning and support vector machine for the detection of electricity theft in power grids. *Energy Reports*, *7*, 349–356. https://doi.org/10.1016/j.egyr.2021.08.038

Hussain, S., Mustafa, M. W., Jumani, T. A., Baloch, S. K., Alotaibi, H., Khan, I., & Khan, A. (2021). A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection. *Energy Reports*, *7*, 4425–4436. https://doi.org/10.1016/j.egyr.2021.07.008

Kawoosa, A. I., Prashar, D., Faheem, M., Jha, N., & Khan, A. A. (2023). Using machine learning ensemble method for detection of energy theft in smart meters. *IET Generation, Transmission and Distribution*, *17*(21), 4794–4809. https://doi.org/10.1049/gtd2.12997

Khan, I. U., Javaid, N., Taylor, C. J., Gamage, K. A. A., & Ma, X. (2021, June 28). Big Data Analytics for Electricity Theft Detection in Smart Grids. *2021 IEEE Madrid PowerTech, PowerTech 2021 - Conference Proceedings*. https://doi.org/10.1109/PowerTech46648.2021.9495000

Kocaman, B., & Tümen, V. (2020). Detection of electricity theft using data processing and LSTM method in distribution systems. *Sadhana - Academy Proceedings in Engineering Sciences*, *45*(1). https://doi.org/10.1007/s12046-020-01512-0

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

Louw, Q., & Bokoro, P. (2019). An alternative technique for the detection and mitigation of electricity theft in south africa. *SAIEE Africa Research Journal*, *110*(4), 209–216. https://doi.org/10.23919/SAIEE.2019.8864147

Munawar, S., Asif, M., Kabir, B., Pamir, Ullah, A., & Javaid, N. (2021). Electricity Theft Detection in Smart Meters Using a Hybrid Bi-directional GRU Bi-directional LSTM Model. *Lecture Notes in Networks and Systems*, *278*, 297–308. https://doi.org/10.1007/978-3-030-79725-6_29

Naeem, A., Aslam, Z., Shloul, T. al, Naz, A., Nadeem, M. I., Hmoud Al-Adhaileh, M., Ghadi, Y., & Mohamed, H. G. (2023). *A Novel Combined DenseNet and Gated Recurrent Unit Approach to Detect Energy Thefts in Smart Grids*. https://doi.org/10.1109/ACCESS.2020.DOI

Naresh, K., & Krishnachaitantya, C. (2025). Eletricity Theft Detection In Smart Grids Based On Deep Neural Network. *International Journal of Research Publication and Reviews*, *6*(5), 11700–11704. https://doi.org/10.55248/gengpi.6.0525.18118

Nazmul Hasan, M., Toma, R. N., Nahid, A. al, Manjurul Islam, M. M., & Kim, J. M. (2019). Electricity theft detection in smart grid systems: A CNN-LSTM based approach. *Energies*, *12*(17). https://doi.org/10.3390/en12173310

Pamir, Javaid, N., Qasim, U., Yahaya, A. S., Alkhammash, E. H., & Hadjouni, M. (2022). Non-Technical Losses Detection Using Autoencoder and Bidirectional Gated Recurrent Unit to Secure Smart Grids. *IEEE Access*, *10*, 56863–56875. https://doi.org/10.1109/ACCESS.2022.3171229

Ranjith Gopalan, Dileesh Onniyil, Ganesh Viswanathan, & Gaurav Samdani. (2025). Hybrid models combining explainable AI and traditional machine learning: A review of methods and applications. *World Journal of Advanced Engineering Technology and Sciences*, *15*(2), 1388–1402. https://doi.org/10.30574/wjaets.2025.15.2.0635

S. Massoud Amin, & B.F. Wollenberg. (2005). Toward_a_smart_grid_power_delivery_for_the_21st_century. *IEEE Power and Energy Magazine*, *3*(5). https://doi.org/DOI:10.1109/MPAE.2005.1507024

Smith, T. B. (2004). Electricity theft: a comparative analysis. *Energy Policy*, *32*(18), 2067–2076. https://doi.org/10.1016/S0301-4215(03)00182-4

Soares, L. D., Queiroz, A. de S., López, G. P., Carreño-Franco, E. M., López-Lezama, J. M., & Muñoz-Galeano, N. (2022). BiGRU-CNN Neural Network Applied to Electric Energy Theft Detection. *Electronics (Switzerland)*, *11*(5). https://doi.org/10.3390/electronics11050693

Takiddin, A., Ismail, M., Zafar, U., & Serpedin, E. (2022). Deep Autoencoder-Based Anomaly Detection of Electricity Theft Cyberattacks in Smart Grids. *IEEE Systems Journal*, *16*(3), 4106–4117. https://doi.org/10.1109/JSYST.2021.3136683

Taruna, A. P., Arisona, G., Irwanto, D., Bestari, A. B., & Juniawan, W. (2025). Electricity Theft Detection Using Machine Learning in Traditional Meter Postpaid Residential Customers: A Case Study on State Electricity Company (PLN) Indonesia. *IEEE Access*, *13*, 7167–7191. https://doi.org/10.1109/ACCESS.2025.3526764

Trindade, A. (2015). *ElectricityLoadDiagrams20112014 [Dataset]*. UCI Machine Learning Repository.

Ullah, A., Javaid, N., Yahaya, A. S., Sultana, T., Al-Zahrani, F. A., & Zaman, F. (2021). A Hybrid Deep Neural Network for Electricity Theft Detection Using Intelligent Antenna-Based Smart Meters. *Wireless Communications and Mobile Computing*, *2021*. https://doi.org/10.1155/2021/9933111

Zheng, K., Chen, Q., Wang, Y., Kang, C., & Xia, Q. (2019). A Novel Combined Data-Driven Approach for Electricity Theft Detection. *IEEE Transactions on Industrial Informatics*, *15*(3), 1809–1819. https://doi.org/10.1109/TII.2018.2873814

Zidi, S., Mihoub, A., Mian Qaisar, S., Krichen, M., & Abu Al-Haija, Q. (2023). Theft detection dataset for benchmarking and machine learning based classification in a smart grid environment. *Journal of King Saud University - Computer and Information Sciences*, *35*(1), 13–25. https://doi.org/10.1016/j.jksuci.2022.05.007