International Journal of Advances in Applied Sciences (IJAAS)
Vol. 15, No. 3, September 2026, pp. 1098~1111
ISSN: 2252-8814, DOI: 10.11591/ijaas.v15.i3.pp1098-1111

A novel combined deep learning-based technique for electricity theft detection in a smart grid

Etienne François Mouckomey[1], Jacques Bikai[2], Camille Franklin Mbey[1], Felix Gislain Yem Souhe[1], Alexandre Teplaira Boum[1], Vinny Junior Foba Kakeu[1]
[1]Department of Electrical Engineering, ENSET, University of Douala, Douala, Cameroon
[2]Department of Electrical Engineering, University of Ngaoundere, Ngaoundere, Cameroon

Article Info

Article history:

Received Mar 12, 2025
Revised Aug 8, 2026
Accepted Aug 13, 2026

Keywords:

ADAM
Bi-LADAM-SVM-GRU model
Deep learning
Electricity theft
Long short-term memory

ABSTRACT

Energy theft represents a major challenge for electricity distribution companies worldwide and constitutes a key component of non-technical losses (NTLs), alongside faulty meters and billing inaccuracies. Despite the rise of artificial intelligence techniques for electricity theft detection, existing approaches exhibit limitations in classification and data analysis when applied to large-scale datasets. These shortcomings can reduce the effectiveness and precision of the theft detection systems. Therefore, this article proposes a bidirectional long short-term memory (Bi-LSTM) with the lookahead adaptive moment estimation (LADAM) optimizer and a support vector machine (SVM) with a bidirectional gated recurrent unit (Bi-GRU) for the detection of energy theft in a smart electricity grid to assess energy supplier companies. In this work, the proposed model is implemented on the State Grid Corporation of China (SGCC) dataset. In the proposed framework, the Bi-LSTM is used for feature extraction, while the LADAM optimizer is employed for feature selection and normalization. The refined feature set is subsequently fed into the SVM classifier to distinguish between theft and non-theft instances. The obtained findings highlight the superior efficiency of the proposed strategy compared with those in the literature, highlighting its potential for practical deployment in real-world smart grid systems.

Corresponding Author:

Etienne François Mouckomey
Department of Electrical Engineering, ENSET, University of Douala
Douala, Cameroon
Email: emoucko@gmail.com

1. INTRODUCTION

The global demand for electrical energy has been increasing steadily due to population growth, industrial expansion, and the electrification of new sectors of activity. This growing demand requires power grid operators to ensure a continuous, stable, and reliable electricity supply to consumers [1]. However, maintaining such reliability remains a challenge due to various losses occurring within the power grid, which directly impact the efficiency and profitability of electricity distribution systems. Power losses in electrical grids can be broadly divided into two groups: technical losses and non-technical losses (NTLs) [2]. Technical losses are inherent to the physical characteristics of electrical components, such as resistive dissipation in transmission lines, transformer inefficiencies, and energy leakage in poorly maintained infrastructures. In contrast, NTLs represent the difference between the total energy supplied through transmission lines and the energy actually billed to end users. These losses are typically caused by human actions such as meter tampering, illegal connections, and falsified readings [3]. NTLs represent a major economic and operational

challenge, costing power utilities millions of dollars annually [4]. Beyond the financial impact, NTLs cause network overloading, voltage instability, and unfair tariff adjustments that penalize compliant consumers. Among these losses, electricity theft is the most significant and widespread issue. It refers to unauthorized consumption of electrical energy that is either unmetered or unbilled, leading to substantial revenue deficits for electricity companies [5]. For instance, electricity losses exceed 4.8 billion rupees annually in India [6], about 0.89 billion rupees in Pakistan [7], and between 8 and 12 billion rand in South Africa [8], [9]. Usually, electricity theft can be detected through manual field verifications, which are costly and time-consuming. Hardware-based approaches requiring additional sensors or monitoring equipment can improve detection accuracy but remain expensive to deploy at large scale [10], [11]. The advent of smart grids and smart meters has opened new opportunities for automated and data-driven theft detection. These devices generate large volumes of time-series consumption data, enabling the application of machine learning and deep learning techniques to identify abnormal consumption patterns and detect fraudulent behavior more efficiently [12].

Over the past decade, several research efforts have focused on developing intelligent systems to reduce NTLs, including electricity theft and false data injection based on generalized robust metrics, multi-classification support vector machine (SVM) [13], quantum machine learning [14], hybrid models such as convolutional neural network (CNN)-long short-term memory (LSTM) and bidirectional gated recurrent unit (Bi-GRU)-bidirectional long short-term memory (Bi-LSTM) [15], and CNN-adaptive boosting (AdaBoost) [16]. For instance, Gunduz and Das [17] proposed an effective hybrid CNN-based approach for detecting energy theft using consumption patterns on six different datasets. Bai *et al.* [18] introduced a Catch22-Conv-Transformer method for detection of electricity theft on a dataset of the State Grid Corporation of China (SGCC), while Nirmal *et al.* [19] suggested a hybrid CNN-AdaBoost model for electricity theft detection in the smart grid on the SGCC dataset. Similarly, Iqbal *et al.* [20] presented CNN-LSTM and Bi-GRU-BiLSTM hybrid models for identification of electricity theft on temporal sequence data of SGCC. Moreover, Algarni *et al.* [21] presented CNN-k-nearest neighbors (KNN) and CNN-LSTM hybrid models for electricity theft detection on the SGCC dataset. Deep learning methods, such as gradient-weighted class activation mapping (GradCAM) [22], random forest (RF) and KNN [23], particle swarm optimization (PSO)-based adaptive generative adversarial networks (AGAN) [24], and hybrid CNN-multilayer perceptron (MLP) model [25] have also been successfully applied to detect electricity theft in smart grid systems.

Despite these advances, challenges remain regarding detection precision and adaptability to irregular consumption behaviors. Many existing models struggle to effectively capture sudden variations in user load profiles and to maintain robustness across heterogeneous datasets. To address these limitations, this study proposes a novel hybrid deep learning approach, named bidirectional lookahead adaptive moment estimation (Bi-LADAM)-SVM-gated recurrent unit (GRU), which combines Bi-LADAM, SVM, and GRU, designed to enhance the accuracy and efficiency of electricity theft detection. This model integrates the strengths of several architectures: a Bi-LSTM network to capture long-term temporal dependencies, an adaptive moment estimation (ADAM) algorithm for optimal hyperparameter tuning, an SVM classifier for robust pattern discrimination, and a Bi-GRU for modeling short-term dynamics.

The major contributions of this work are given as follows:

- First present a comprehensive evaluation of deep learning models commonly used for electricity theft detection, including LSTM, CNN, Bi-LSTM, Bi-GRU, extreme gradient boosting (XGBoost), and decision tree.
- Then, proposed a new hybrid model (Bi-LADAM-SVM-GRU) combining Bi-LSTM, ADAM optimizer, SVM, and Bi-GRU to enhance temporal feature extraction and improve classification accuracy.
- Subsequently, this study implemented an experimental validation of the proposed model using a real consumption dataset from the SGCC to assess its practical effectiveness.
- In addition, this study evaluated the performance of the proposed hybrid model through multiple statistical indicators such as mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), accuracy rate, F1-score, recall, area under the curve (AUC), R-squared (R^2), and adjusted R^2 and compared with baseline strategies such as CNN, artificial neural network (ANN), MLP, deep neural network (DNN), and linear regression.

The rest of this study is organized as follows. Section 2 describes the materials. Section 3 describes the methodologies employed. Section 4 presents and discusses the experimental results. Finally, section 5 concludes the work and outlines perspectives for future research.

2. MATERIAL

2.1. Dataset description

The dataset given in Table 1 was recorded from smart meters or electricity consumer sensors over a period of 34 months from SGCC. SGCC offers a dataset consisting of 42,372 users over 1,035 days from

1 January to 31 October 2016. There are 38,757 normal users marked as "0" and 3,615 electricity theft users marked as "1"; the electricity theft users account for 8.5% of the global dataset.

Table 1. Detailed description of the dataset

Type of data	Energy consumption
Resolution	1 day
Number of consumers	42,372
Size	1,035 days
Normal users	38,757
Electricity theft	3,615
Preprocessing data	70%
Testing data	30%
Imbalanced data	20%

2.2. Python

Python is a programming language in the field of machine learning and deep learning. Within the framework of this study, it served as the foundation for developing the proposed deep learning model. Using its key libraries, including Matplotlib, Seaborn, NumPy, Pandas, and scikit-learn.

2.3. MATLAB

MATLAB is a specialized programming environment used in engineering and scientific applications. In this work, training and evaluation of the proposed model were made possible using the MATLAB 2021a version. Evaluation of the proposed model was also conducted using the same environment.

2.4. Computer

All experiments and computations for this study were performed on a DELL computer equipped with a Core i7 processor (2.20 GHz), 8 GB of RAM, and Windows 10 operating system. This configuration facilitated the execution of all implemented deep learning algorithms, including LSTM, GRU, Bi-LSTM networks, Bi-GRU networks, SVM, CNNs, XGBoost, and decision trees. All algorithms were executed with excellent computational time.

3. METHOD

3.1. Long short-term memory

LSTM is a specialized variant of the recurrent neural network (RNN) architecture, designed to more effectively model long-range temporal dependencies. A key advantage of LSTM is its exceptional capability in preprocessing sequential data, such as time series. There are interactions between the various components; h_{t-1} and h_t indicate the output of the preceding and current cell, respectively. The current unit input is symbolized as x_t, while the neural network layer is represented by a box.

3.2. Support vector machine

The SVM algorithm was developed in the 1980s by Vladimir Vapnik using mathematical and optimization evolutionary tools, considering high-dimensional datasets and good practical results. Practically, the SVM is made of two groups that are separated by a hyperplane. However, there is an optimal space among the separation planes, where the distance between the closest hyperplane and the training element is optimal. The problem is to find the optimal separation frontier of the samples between the two parts. Solving these problems involves constructing a hyperplane h with weights ω and ω_0 as presented in (1).

$$h(x) = \omega^P x + \omega_0 \quad (1)$$

3.3. Gated recurrent unit network

The GRU network is a kind of RNN that has garnered considerable interest for its capacity to effectively capture and model long-term relationships. GRU networks distinguish themselves from LSTM networks by integrating the cell state and hidden state into a unified variable, denoted as h_t, whereas LSTM networks maintain them as distinct entities. GRU networks employ the following mathematical equations to carry out this merging. In (2) and (3), inputs are collected from the previous hidden state and the current data point, which are then processed with respective weights.

$$z_t = \sigma(\omega_z[h_{t-1}, x_t] + b_z) \quad (2)$$

$$r_t = \sigma(\omega_r[h_{t-1}, x_t] + b_r) \quad (3)$$

3.4. Convolutional neural networks

CNNs are a specific category of ANNs that have garnered considerable interest because of their outstanding capabilities in tasks such as image identification, object detection, and medical image analysis. The structure of CNNs generally comprises convolutional layers, pooling layers, and fully connected layers. Convolutional layers facilitate the network in effectively capturing spatial hierarchies and patterns in the input data, hence making CNNs very successful for tasks that include visual information. Moreover, CNNs have proven to be effective in a wide range of time series applications, such as classifying data, detecting anomalies, and predicting outcomes.

3.5. XGBoost decision tree

XGBoost decision tree is employed in supervised machine learning applications for classification, regression, and ranking. It works by iteratively training decision trees on the residual errors of the previous tree. The system employs gradient descent optimization to reduce a loss function, which quantifies the disparity between the anticipated and actual values. The algorithm is used for supervised learning problems by predicting the target label y with training data x. Training loss and regularization terms comprise the objective function of XGBoost as in (4).

$$obj(f) = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \Omega(f) \quad (4)$$

3.6. Performance metrics

The performance of the proposed model is evaluated using the following metrics.

i) Accuracy: which describes the overall performance of a model across all classes. The calculation involves determining the proportion of accurate predictions relative to the overall number of forecasts made, as shown in (5).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

ii) Precision: which evaluates the ability of the model to correctly classify a sample, as defined in (6).

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

iii) Recall: which classifies the positive samples, as given in (7).

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

iv) F1-score: which gives the relationship between precision and sensitivity, as presented in (8).

$$F1 - score = 2 \times \left(\frac{TP}{2 \times TP+FP+FN}\right) \quad (8)$$

With true positive (TP), false positive (FP), true negative (TN), and false negative (FN). These metrics represent the possibilities of the consumers being either a true or a false consumer.

v) MSE: represents mean of square difference between the predicted values and real values, as defined in (9).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\tilde{y}_i - y_i)^2 \quad (9)$$

vi) Root mean square error (RMSE): represents the standard deviation between predicted values and real values, as given in (10).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\tilde{y}_i - y_i)^2} \quad (10)$$

vii) MAE: represents the absolute value of the difference between the predicted values and the real values, as presented in (11).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\tilde{y}_i - y_i| \quad (11)$$

viii) MAPE: represents the mean of absolute percentage prediction errors, as formulated in (12).

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\tilde{y}_i - y_i}{y_i}\right| \times 100\% \quad (12)$$

ix) R^2: represents the proportion of variance in the dependent variable that is predictable from the independent variables, as expressed in (13). With y_i the real value, $\tilde{y}_i$ the predicted value, and $\bar{y}$ the mean value of the recorded features.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\tilde{y}_i - y_i)^2}{\sum_{i=1}^{N}(\tilde{y}_i - \bar{y})} \quad (13)$$

The highest accuracy means that the fraudulent consumer has been detected. When the recall and F1-score are high, it means that a consumer is detected as fraudulent or normal.

3.7. Proposed models

3.7.1. Lookahead adaptive moment estimation model

The LADAM model is a hybrid algorithm that combines the LSTM with the ADAM algorithm for optimizing neural network hyperparameters. ADAM is one of the most robust optimizers that improves training speeds in DNN to achieve convergence quickly. It combines the best elements of momentum and processing. Data are preprocessed to verify the absence of null and missing values and to detect any outliers. Data scaling normalizes the range of variables. The data are separated into training and testing sets using a training and test split, ensuring that the temporal order is preserved. The prepared data can be further utilized for machine learning purposes by employing deep LSTM models optimized by the ADAM algorithm, as depicted in Figure 1. The models were trained using a realistic electricity consumption database that was made available by the SGCC.

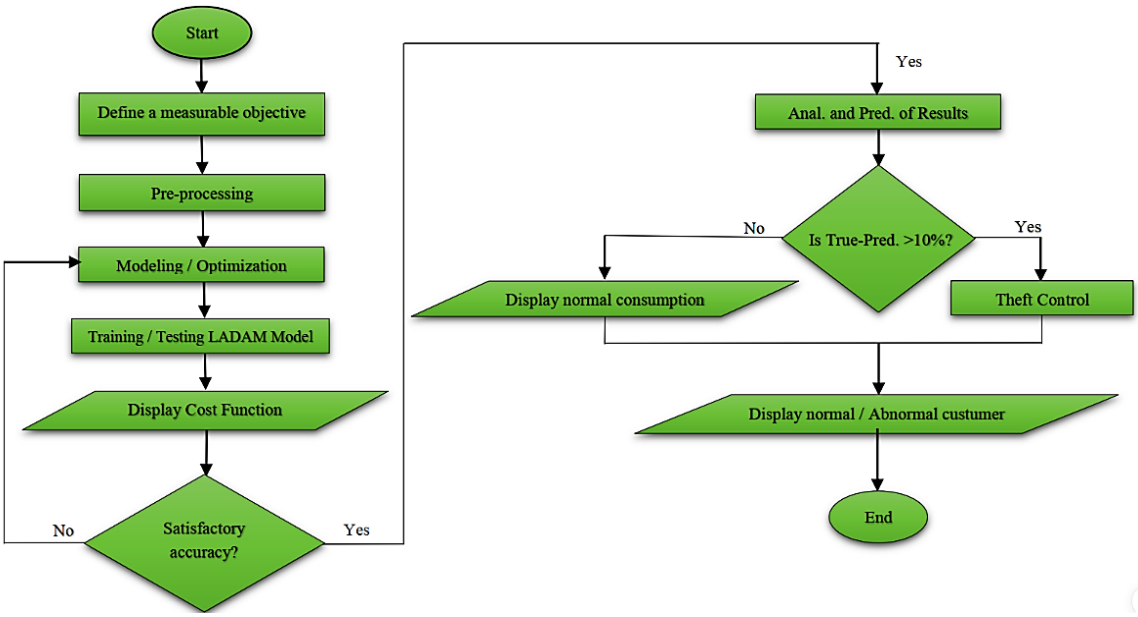

Figure 1. Proposed flowchart for the LADAM hybrid algorithm

3.7.2. Proposed methodology framework for electricity theft detection

In this study, a combined scheme based on LADAM, Bi-LSTM, Bi-GRU, and SVM is implemented for the detection of electricity theft in a smart power grid. The LADAM model, made of LSTM and ADAM optimizer, is developed for meaningful feature extraction; the Bi-LSTM is used for the extraction, while LADAM serves as feature selection. The SVM allows to classify the extracted features as normal or fraudulent, while GRU is used for feature prediction. The flowchart of the proposed methodology framework for electricity theft detection is given in Figure 2. This flowchart outlines the key stages, including data collection, data analysis and preprocessing, feature extraction and selection, feature classification and prediction, normalization, application of deep learning-based techniques, hyperparameter optimization, and performance evaluation metrics. This structured framework ensures clarity in the steps followed for model

development and validation. The parameters of the models are described as follows: 4 convolution layers, a softmax LSTM transfer function, 7 minimum batch size, 60 maximum epoch, 100 iterations, a learning rate of 0.0001, an ADAM optimizer, a 30% dropout rate, and a sigmoid activation function. The methodology for data splitting, cross-validation, and hyperparameter optimization is described as follows:

i) For the training-validation splitting method, the dataset is organized by characteristics and data nature, ensuring that each group contains a time series. This grouping strategy enables controlled overlap handling, which is essential for sequential learning models. The dataset is partitioned as follows: 70% for training, 15% for validation, and 15% for testing.

ii) To enhance generalization, data segments from all smart meters are aggregated to form comprehensive global sets. This aggregation allows for assessing temporal degradation and for classifying new, previously unseen customers. Additionally, sample weighting, focal loss, and controlled temporal oversampling (synthetic minority over-sampling technique (SMOTE)) are employed to balance the dataset, particularly by augmenting or duplicating fraudulent time windows.

iii) The dataset used in this study comprises 42,372 smart meters installed at customer premises, providing daily observations over a 10-month period.

iv) For the cross-validation strategy, this study adopts a time-aware grouped k-fold approach designed to evaluate temporal robustness while preserving group integrity by meter ID. The total observation period is divided into K contiguous time blocks. For each fold i, the model is trained on blocks 1 to i and validated on block i + 1. Consequently, the training set expands with each iteration, emulating the process of real-world deployment and enabling the model to predict future consumption without data leakage from private customer information.

v) Regarding hyperparameter tuning, this work employs a Bayesian optimization strategy within a nested search framework to mitigate over-optimism and enhance generalization. The complete modeling pipeline includes three stages: data preprocessing, feature engineering, and model augmentation. The defined hyperparameter search space includes the following: learning rate of 0.002, batch size of 32, 4 hidden layers, 64 hidden units, dropout rate of 0.3, weight decay of 0.02, sequence length of 12, and latent dimension of 16 (with the ADAM optimizer).

The Bayesian search iteratively explores the parameter space while terminating unpromising trials early. For each outer iteration, the best-performing set of hyperparameters identified internally is selected to retrain the model on the entire input dataset. The model's performance may be underestimated if fraudulent behavior is strongly correlated with metadata.

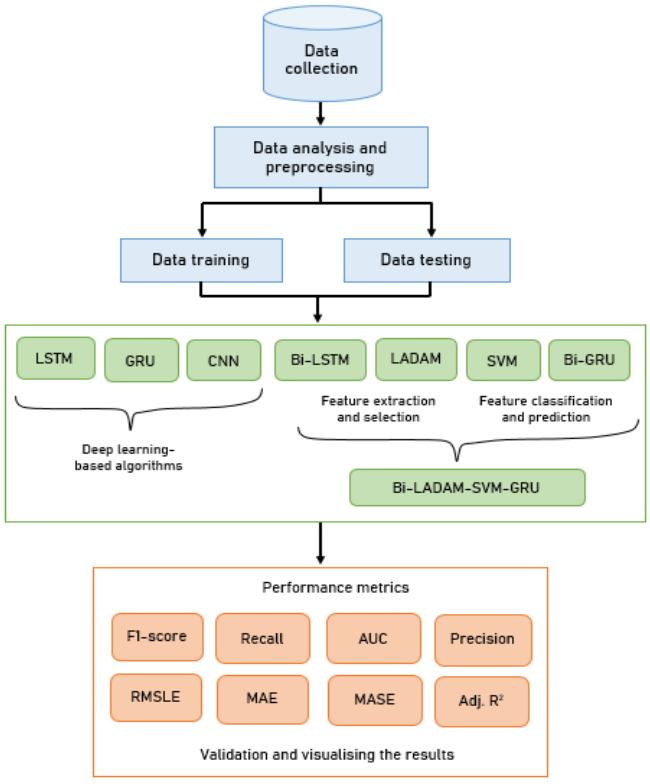

Figure 2. Proposed methodology framework for electricity theft detection

The proposed framework for electricity theft detection is implemented as follows:

i) Step 1: data collection. In this step, the data are collected every day from smart meters installed to consumers. In this work, 42,372 consumptions have been recorded during 1,035 days.
ii) Step 2: data analysis and preprocessing. The collected data from electrical consumers are analyzed and preprocessed. This step allows to avoid missing data with an interpolation method using data cleaning and data normalization. The data preprocessing step improves the data accuracy, which directly enhances the efficacy of the model.
iii) Step 3: data training and testing. In this model, the training step allows to train the dataset while the testing data are used to validate the model's generalization. The overall dataset is divided into training data of 70% and testing data of 30%.
iv) Step 4: feature extraction and selection. In this step, the Bi-LSTM allows feature extraction, while the LADAM algorithm is used to select and normalize the features obtained after the training stage.
v) Step 5: feature classification and prediction. In this stage, the SVM model allows to classify the normalized features with the aim of determining if the user is stealing energy. In this work, the SVM examines the dataset and divides it into two groups: normal consumers and theft consumers. SVM works on the phenomena of a hyperplane and divides the extracted features obtained from the LADAM model into theft and non-theft users. Then, the GRU is employed for the prediction of the classified features and finally identifies the potential fraudsters.
vi) Step 6: validation of the model. This step consists of verifying the efficiency of the proposed model with the implementation of performance metrics such as MAE, MSE, MAPE, R^2, F1-score, precision, AUC, and recall. For the theft consumers, the MAE, MAPE, and RMSE should be low, while F1-score, accuracy, recall, and AUC should be highest as possible. For the non-theft consumers, the MAE, MAPE, and RMSE should be high, while F1-score, accuracy, recall, and AUC should be lowest as possible. When the suspicious consumers are identified using the proposed Bi-LADAM-SVM-GRU model, the energy supplier company can therefore send a penalty to the fraudulent consumer to compensate for the illegally consumed energy. Then, if the consumer doesn't pay the penalty, the company can send technicians to shut down the consumer.

4. RESULTS AND DISCUSSION

4.1. Analysis of customer consumption

The collection of data allows analysis of electricity consumption behavior of users with demographic, social, and financial trends. Data are collected over 1,035 days for 42,372 consumers. In this study, consumers have installed smart meters at their homes to collect real-time electricity consumption. In this case, this study was interested in the consumption of customer 35. The analysis of its consumption was carried out over the three years for which our consumption records were taken.

4.1.1. For the year 2014

Figure 3 demonstrates customer 35 throughout 2014, with high energy consumption between January and March, mid-May and September, and mid-November and December. However, it declines between April and mid-May and then between September and November. It can be observed that the consumption is depending to the periods of the year, whether the users are at home or at vacancy.

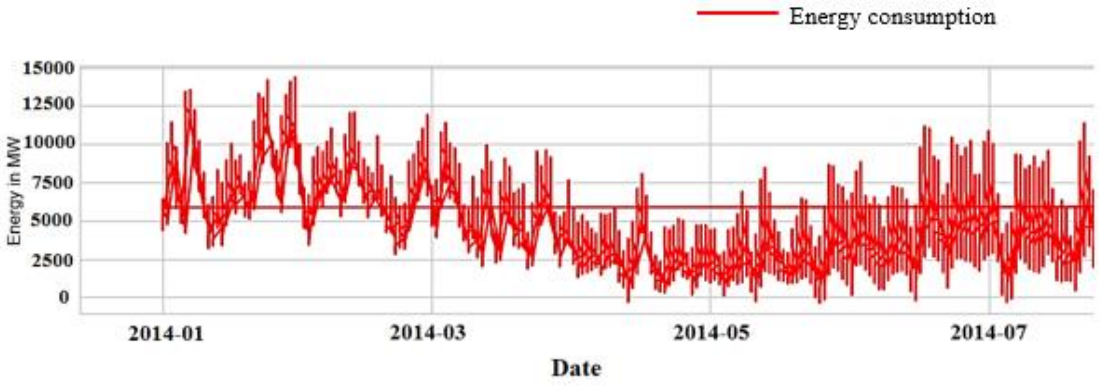

Figure 3. Energy consumption by customer 35 in 2014

4.1.2. For the year 2015

Figure 4 demonstrates the changes in consumption for customer 35 throughout 2015, with high energy consumption between January and the first half of March and then between June and September. It can be observed a high variability of consumption during this year. This can be explained with a regular enhancement of people's lifestyles during this year.

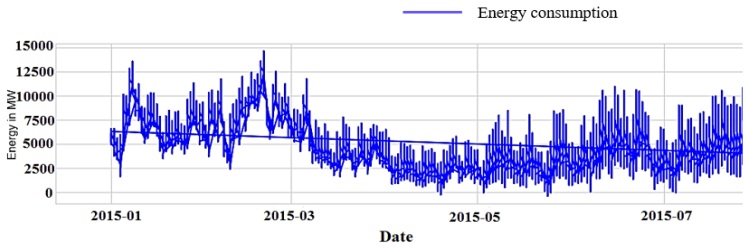

Figure 4. Energy consumption by customer 35 in 2015

4.1.3. For the year 2016

Voltage rises are recorded between January and March, June and October, and the second half of September, as illustrated in Figure 5. It can be observed a great consumption of around 10,000 MW between July and September, and a low consumption of around 7,000 MW between January and May. Moreover, a minimal consumption is recorded at around 2,000 MW between October and December.

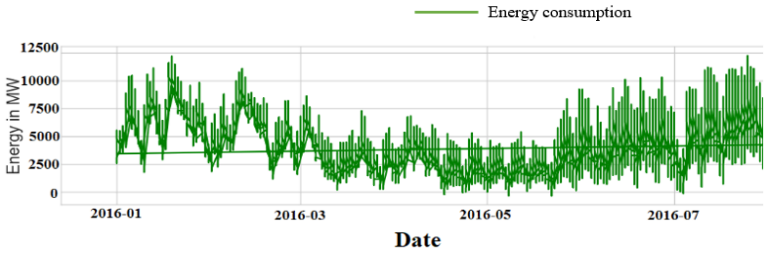

Figure 5. Energy consumed by customer 35 in 2016

For the dataset analysis, this study use LADAM model to extract important features by eliminating noisy data. From the obtained data, the SVM allows to classify the consumers as normal consumers or abnormal consumers. Therefore, sudden changes in the consumption of suspicious consumers allow the SVM to identify fraud that may occur. From the obtained results, the LADAM-SVM model allows for identifying around 462 fraudulent consumers, which represents around 1.09% of the overall consumers. The identification of these fraudsters allows for avoiding the loss of millions of dollars during this period for the electricity supplier companies.

4.2. Comparative study of optimization techniques

The choice of the model led us to carry out a comparative study with the state-of-the-art optimizers. These are different forms of gradient descent employed in machine learning to optimize the cost function by iteratively modifying the parameters towards the negative gradient direction to discover the optimal parameters for a model. The aim of gradient descent is to find the set of parameters that minimizes this deviation and improves the model's performance. The cost function, in this case the MSE, quantifies the disparity between predicted consumption and actual consumption by the model. The comparison of implemented models is given in Table 2.

The momentum and root mean square propagation (RMSProp) algorithms have been associated with the creation of the ADAM optimization algorithm. The cost function for this specific case is MSE = $6.45e^{-09}$. The Nesterov algorithm, an extension of the classic gradient descent algorithm, considers previous gradient updates. Its cost function for this specific case is MSE = 3.72 e^{-05}.

The momentum gradient descent algorithm is an optimization algorithm that follows the convergence of gradient descent by optimizing the updating of the weights. The impulse term is based on the previous update of the weights and helps the algorithm gain momentum as it moves down the loss function.

The cost function for this specific case is MSE = 8.75e^{-05}. The adaptive gradient (AdaGrad) algorithm is a gradient-based optimization algorithm. Its learning rate is adapted to the hyper parameters by knowledge incorporation of past observations. Its cost function for this specific case is MSE = 1.82e^{-04}. The RMSProp algorithm is a gradient-based method in which its cost function for this specific case is MSE = 5.79e^{-04}.

The mini-batch gradient descent algorithm is a variant of the gradient descent algorithm frequently used in training deep learning models. The basic principle of this algorithm is to separate the training data to process them sequentially. Its cost function for this specific case is MSE = 8.16e^{-04}. The gradient descent with decaying learning rate algorithm, a modern neural network training technique, starts by training the network with a high learning rate and then slowly reducing it until a local minimum is obtained. This shows that an initially high learning rate prevents the network from noisy data memorization, while a decreasing learning rate improves the learning of complex models. The cost function for this specific case is MSE = 0.27. It can also be observed that Table 2 summarizes the cost function optimized respectively with ADAM, Nesterov accelerated gradient descent, momentum-based gradient descent, AdaGrad, RMSProp, mini-batch gradient descent, and gradient descent with decaying learning rate. This shows that the LADAM-SVM optimizer, with a cost function (MSE) equal to 6.45e^{-09}, outperforms the other six optimizers. The reduction of MSE with the LADAM model proves its outperformance compared to those in the literature.

Table 2. Comparison of implemented models

Optimizers	Epochs	Learning rate	Cost function (MSE)
Proposed model (LADAM-SVM)	500	0.01	6.45e^{-09}
Nesterov accelerated gradient descent	500	0.01	3.72e^{-05}
Momentum-based gradient descent	500	0.01	8.75e^{-05}
AdaGrad	500	0.01	1.84e^{-04}
RMSProp	500	0.01	5.79e^{-04}
Mini-batch gradient descent	500	0.01	8.16e^{-04}
Gradient descent with decaying learning rate	500	0.01	0.27

4.3. Predicting customer consumption

This study set out the consumption of customer 35 from January 1, 2016 to January 1, 2017. Figure 6 shows the electricity consumption measured after each hour. This paper can observe that the consumption is growing around 8,000 KW during November 2016 and December 2016. Moreover, the consumption is reduced to around 3,000 KW during October 2016, which is caused by the reduction of living costs.

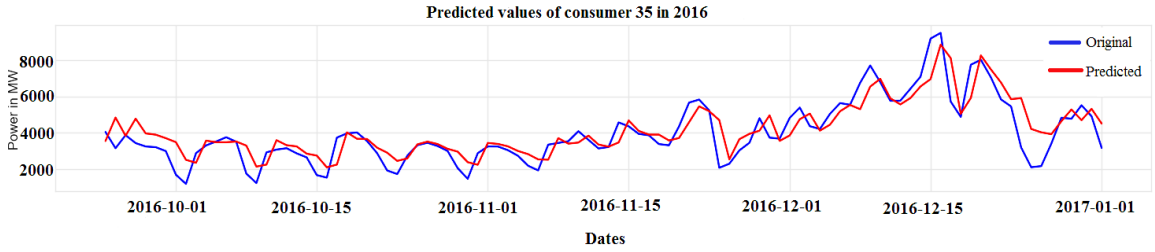

Figure 6. Prediction of customer 35's consumption in 2016

4.4. Detecting energy theft

The energy theft detection model presented in this study comprises multiple stages, with the initial stage being data pre-processing. Subsequently, the sequential diagram is developed, as detailed in the forthcoming subsection. The subsequent stage involves constructing the energy theft prediction model utilizing the LSTM approach, with its hyperparameters fine-tuned through the utilization of the ADAM algorithm. It is deemed suspicious and necessitates monitoring by the energy distribution company if the energy consumption is less than 10% of the anticipated value during a span of six months. The utilization of forecast model empowered us to potentially detect days when the customer's usage was questionable and days when it was typical. The date suspected of energy theft is the one that senses an increase in consumption of more than 10% of its predicted consumption. The consumption gap (CG) is defined as (14).

$$CG(\%) = 100 \times \frac{PC-RC}{PC} \quad (14)$$

If CG(%) < 10%, the customer's consumption is considered normal, whereas if CG(%) ≥ 10%, the consumption is classified as suspect. Here, PC denotes the predicted consumption and RC the real consumption. Furthermore, if a customer's consumption remains under 10% of the predicted value for six consecutive months, an inspection team will be dispatched. Table 3 summarizes the performance metrics obtained for electricity theft detection using the implemented deep learning strategies. The corresponding confusion matrices of the implemented deep learning models are depicted in Figure 7.

Table 3. Performance metrics for electricity theft detection

Models	F1-score	Recall	Accuracy	Precision	ROC-AUC
LSTM	0.83	0.80	0.81	0.91	0.81
GRU	0.87	0.83	0.86	0.93	0.83
LSTM-SVM	0.90	0.93	0.92	0.95	0.87
Bi-LSTM	0.91	0.95	0.94	0.95	0.91
Bi-GRU	0.91	0.96	0.95	0.97	0.01
Bi-LADAM-SVM-GRU	0.92	0.97	0.96	0.98	0.92

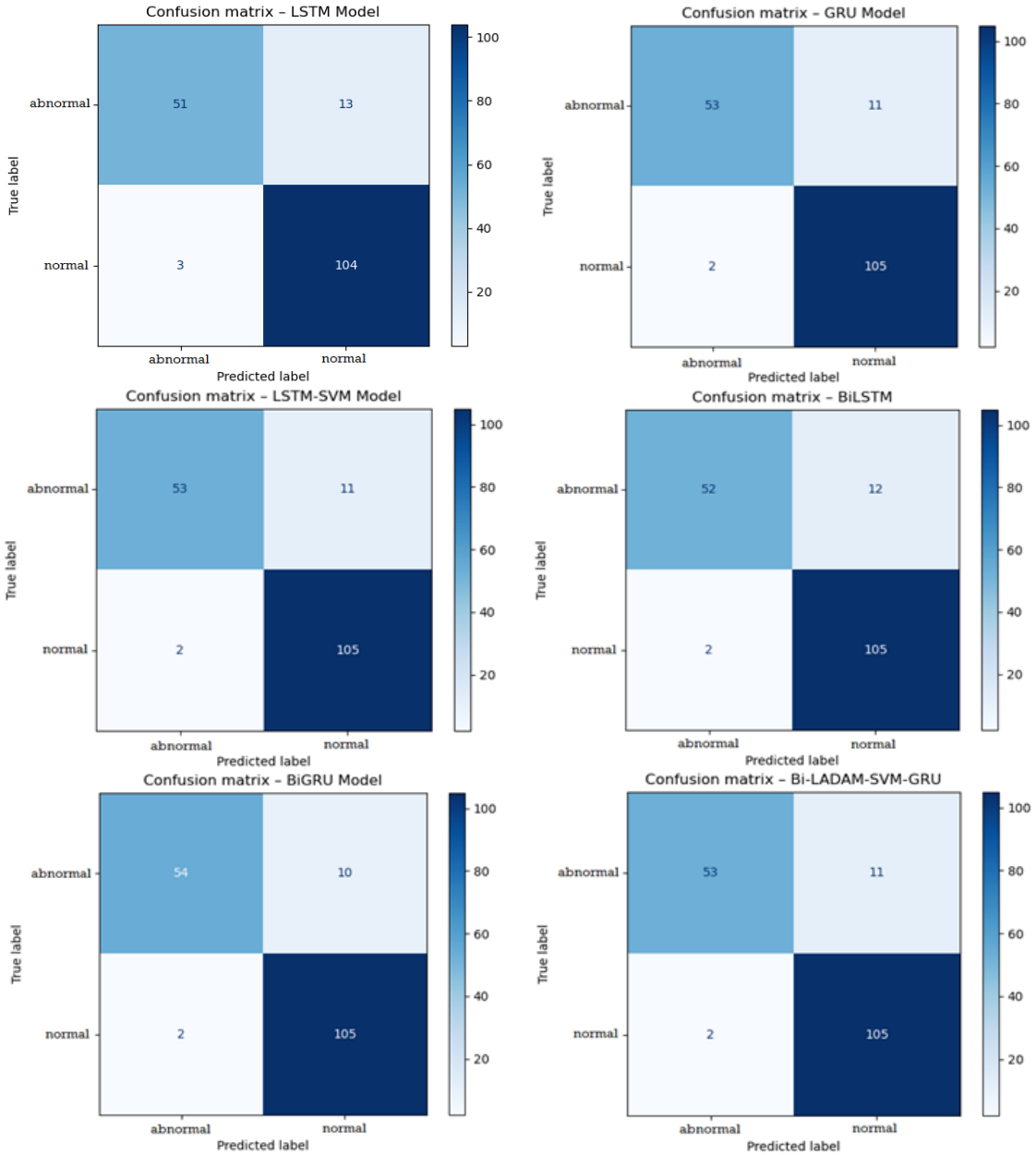

Figure 7. Confusion matrices of implemented deep learning models

As shown in confusion matrices, for the proposed deep learning model, the TP is 53, which represents the number of attack instances detected daily. The TN is 2, so the model incorrectly identified 2 abnormal instances as attacks for a day. The FP is 105, so the model incorrectly detected 2 normal instances as attacks daily. The FP is 11, which means that the model failed to identify 11 attacks, mistakenly classified as normal.

4.5. Comparison with previous works

Several deep learning methods have been studied in the literature for electricity theft detection for different datasets. Table 4 presents a comparative study of the proposed Bi-LADAM-SVM-GRU model with several other methods that have been used to detect electricity theft in the literature. As shown in Table 4, the performance evaluation reveals that the proposed Bi-LADAM-SVM-GRU model outperforms other models respectively, with an MSE of 0.00692, MAE of 0.062, RMSE of 0.0067, and R^2 of 0.74. After examining this previous work, this study notes the superiority of the proposed model. The relevance of these metrics to theft detection is observed with the necessity to acquire real-time data from smart meters and the data analysis and training, especially for large-scale datasets. These results are justified because hybrid Bi-LADAM-SVM-GRU model incorporates the benefits of both Bi-LSTM and LADAM for feature extraction and selection, and SVM with GRU for feature classification and prediction. However, the efficiency the proposed hybrid model can be improved in the future by reducing the necessary time for electricity theft detection.

Table 4. Comparison with recent deep learning approaches of literature

Author	Years	MSE	MAE	RMSE	R^2	F1-score	Recall	Accuracy	Precision	AUC
Nirmal *et al.* [16]	2025	-	-	-	-	0.78	0.76	0.91	0.86	0.91
Gunduz and Das [17]	2024	-	-	-	-	0.90	0.84	0.87	0.89	0.90
Nirmal *et al.* [19]	2024	0.01	0.34	-	-	0.71	0.85	0.88	0.92	0.79
Taruna *et al.* [23]	2025	0.2	-	0.3	0.89	0.79		0.89	0.91	0.86
Sun *et al.* [24]	2025	0.9	-	-	-	0.84	0.76	-	0.81	0.79
Chuwa *et al.* [25]	2025	-	0.34	0.67	0.81	0.77	0.91	0.83	0.91	0.91
This study	2025	0.00692	0.062	0.0067	0.74	0.92	0.97	0.96	0.98	0.92

5. CONCLUSION

This research presented a comprehensive framework based on deep learning models for electricity theft detection in a smart distribution grid. Therefore, the proposed method, named Bi-LADAM-SVM-GRU model, is made of Bi-LSTM, LADAM optimizer with SVM classifier, and Bi-GRU. In this study, a dataset provided by the SGCC is presented to evaluate the performance of the proposed method. The dataset has been pre-processed and trained to optimize the deep learning model. Thus, the proposed intelligent model allows to detect fraudulent consumers over a three-year period. In addition, to validate the proposed model, simulations were carried out using various performance coefficients such as MSE, RMSE, MAE, and R^2, respectively, 0.00692, 0.0067, 0.062, and 0.74. Moreover, hybrid Bi-LADAM-SVM-GRU model was compared with other deep learning techniques in the literature, such as LSTM-ADAM, LSTM-Nesterov, LSTM-Momentum, LSTM-Adagrad, LSTM-RMSProp, LSTM-mini-batch gradient descent, and LSTM-gradient descent with decaying learning rate, XGBoost decision tree, and auto-encoder. The results derived from this research demonstrate the outperformance of this novel smart model compared with those in the literature for electricity theft detection. With these promising results, the proposed hybrid model can be applied for large-scale smart grids with the implementation of feature extraction and real-time data acquisition. Moreover, further work can be applied in the implementation of optimized deep learning techniques with different datasets and the internet of things in order to obtain better performance. In addition, we can further improve the system with federated generative deep learning algorithms for the enhancement of privacy and to reduce the impact of electricity theft, which can give novel opportunities to electricity distributors.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Department of Electrical Engineering of ENSET, University of Douala, Douala, Cameroon.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Etienne François Mouckomey	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	
Jacques Bikai	✓	✓	✓		✓		✓		✓		✓	✓		
Camille Franklin Mbey		✓		✓		✓	✓			✓		✓	✓	✓
Felix Gislain Yem Souhe		✓	✓		✓			✓	✓		✓		✓	
Alexandre Teplaira Boum	✓	✓		✓		✓		✓	✓	✓	✓	✓		
Vinny Junior Foba Kakeu	✓		✓		✓	✓	✓		✓	✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

Not applicable. This study did not involve human participants, human data, or any personally identifiable information. All data used were either publicly available, fully anonymized, or derived from non-human sources, and therefore no informed consent was required from individuals.

ETHICAL APPROVAL

Not applicable. This research did not involve human subjects, human biological materials, or experimental procedures on animals. The work was conducted solely on computational models, publicly available datasets, or non-sensitive data that did not require intervention with living organisms. Therefore, ethical approval from an institutional review board or animal ethics committee was not necessary for this study.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [EFM], upon reasonable request.

REFERENCES

[1] H. Wen, X. Liu, B. Lei, M. Yang, X. Cheng, and Z. Chen, "A privacy-preserving heterogeneous federated learning framework with class imbalance learning for electricity theft detection," *Applied Energy*, vol. 378, pp. 1–21, 2025, doi: 10.1016/j.apenergy.2024.124789.

[2] W. Liao, D. Yang, L. Ge, Y. Jia, and Z. Yang, "Electricity theft detection in integrated energy systems considering multi-energy loads," *International Journal of Electrical Power and Energy Systems*, vol. 164, pp. 1–12, 2025, doi: 10.1016/j.ijepes.2024.110428.

[3] L. Ge, J. Li, T. Du, and L. Hou, "Double-layer stacking optimization for electricity theft detection considering data incompleteness and intra-class imbalance," *International Journal of Electrical Power and Energy Systems*, vol. 165, pp. 1–15, 2025, doi: 10.1016/j.ijepes.2025.110461.

[4] A. Ullah, I. U. Khan, M. Z. Younas, M. Ahmad, and N. Kryvinska, "Robust resampling and stacked learning models for electricity theft detection in smart grid," *Energy Reports*, vol. 13, pp. 770–779, 2025, doi: 10.1016/j.egyr.2024.12.041.

[5] O. O. Ajayi, L. A Adegboye, A. Adebisi, J. O. Babawale, "Detection of electricity theft in metering systems using convolutional neural network," *Engineering and Technology Journal*, vol. 10, no. 08, pp. 6066–6073, 2025, doi: 10.47191/etj/v10i08.04.

[6] E. Madhuri, D. Sushanth, D. N. Reddy, K. V. Kumar, and G. Swetha, "IoT-based power theft detector," *International Journal of Innovative Research in Engineering*, vol. 6, no. 1, pp. 1–6, 2025, doi: 10.59256/ijire.20250601001.

[7] A. Norouzi, I. Ahmadi, M. Nazari, H. Pourrostami, and H. H.-Kordkheili, "A novel framework to detect electricity theft in regions with traditional metering; applying a hybrid machine learning-based approach to low-sampling-rate data," *Results in Engineering*, vol. 25, 2025, doi: 10.1016/j.rineng.2025.104478.

[8] M. Ammar, N. Javaid, and A. Arishi, "An AI-explained systematic modular approach for enhanced electricity theft detection for urbanized smart grid environment," *Alexandria Engineering Journal*, vol. 129, pp. 998–1023, 2025, doi: 10.1016/j.aej.2025.07.040.

[9] H. K. Imran and M. S. S. Al-kafaji, "Electricity theft detection in smart grids based on machine learning techniques," *Al-Furat Journal of Innovations in Electronics and Computer Engineering*, vol. 4, no. 1, pp. 160-183, 2025, doi: 10.46649/fjiece.v4.1.12a.25.3.2025.
[10] A. H. Massarani, M. M. Badr, M. Baza, H. Alshahrani, and A. Alshehri, "Efficient and accurate zero-day electricity theft detection from smart meter sensor data using prototype and ensemble learning," *Sensors*, vol. 25, no. 13, 2025, doi: 10.3390/s25134111.
[11] M. Adil, N. Javaid, U. Qasim, I. Ullah, M. Shafiq, and J. G. Choi, "LSTM and bat-based rusboost approach for electricity theft detection," *Applied Sciences*, vol. 10, no. 12, pp. 4378–4388, 2020, doi: 10.3390/app10124378.
[12] H. Zhang, "Development of an intelligent intrusion detection system for IoT networks using deep learning," *Discover Internet of Things*, vol. 5, no. 1, pp. 1–27, 2025, doi: 10.1007/s43926-025-00177-7.
[13] W. Zhang, Q. Cao, S. Yang, and H. Guo, "Detection of power theft in sensitive stations based on generalized robust distance metric and multi-classification support vector machine," *Discover Applied Sciences*, vol. 7, no. 5, pp. 1–16, 2025, doi: 10.1007/s42452-025-06834-4.
[14] K. Blazakis, N. Schetakis, M. M. Badr, D. Aghamalyan, K. Stavrakakis, and G. Stavrakakis, "Power theft detection in smart grids using quantum machine learning," *IEEE Access*, vol. 13, pp. 61511–61525, 2025, doi: 10.1109/ACCESS.2025.3558143.
[15] M. S. Iqbal, S. Munawar, M. Adnan, A. Raza, M. A. Akbar, and A. Bermak, "A critical review of technical case studies for electricity theft detection in smart grids: a new paradigm-based transformative approach," *Energy Conversion and Management: X*, vol. 26, pp. 1–37, 2025, doi: 10.1016/j.ecmx.2025.100965.
[16] S. Nirmal, P. Patil, and S. Shinde, "Adversarial measurements for convolutional neural network-based energy theft detection model in smart grid," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 11, pp. 1–9, 2025, doi: 10.1016/j.prime.2025.100909.
[17] M. Z. Gunduz and R. Das, "Smart grid security: an effective hybrid CNN-based approach for detecting energy theft using consumption patterns," *Sensors*, vol. 24, no. 4, 2024, doi: 10.3390/s24041148.
[18] W. Bai, L. Xiong, Y. Liao, Z. Tan, J. Wang, and Z. Zhang, "Detection method for three-phase electricity theft based on multi-dimensional feature extraction," *Sensors*, vol. 24, no. 18, 2024, doi: 10.3390/s24186057.
[19] S. Nirmal, P. Patil, and J. R. R. Kumar, "CNN-AdaBoost based hybrid model for electricity theft detection in smart grid," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 7, 2024, doi: 10.1016/j.prime.2024.100452.
[20] M. S. Iqbal, S. Munawar, M. Adnan, and S. E. G. Mohamed, "An efficient approach for electricity theft detection based on transfer learning in temporal sequence," *Results in Engineering*, vol. 26, 2025, doi: 10.1016/j.rineng.2025.105125.
[21] F. Algarni, L. Nassef, and N. Alowidi, "Deep learning model with hybrid resampling technique for electricity theft detection in smart grid," *Journal of King Abdulaziz University: Computing and Information Technology Sciences*, vol. 14, no. 1, pp. 95–105, 2025, doi: 10.64064/1658-6336.1009.
[22] X. Li, W. Lv, I. U. Khan, B. Xie, and R. Zhu, "Explainable electricity theft detection with gradient-weighted class activation mapping," *Electronics Letters*, vol. 61, no. 1, 2025, doi: 10.1049/ell2.70264.
[23] A. P. Taruna, G. Arisona, D. Irwanto, A. B. Bestari, and W. Juniawan, "Electricity theft detection using machine learning in traditional meter postpaid residential customers: a case study on state electricity company (PLN) Indonesia," *IEEE Access*, vol. 13, pp. 7167–7191, 2025, doi: 10.1109/ACCESS.2025.3526764.
[24] Y. Sun, S. Dang, W. Fang, W. Huang, and Y. Xu, "Design and analysis of electricity anomaly detection and diagnosis models based on generative adversarial networks," *International Journal of High Speed Electronics and Systems*, vol. 34, no. 4, pp. 1–32, 2025, doi: 10.1142/S0129156425403304.
[25] M. Chuwa, D. Ngondya, and R. Mwifunyi, "Use of information-fusion deep-learning techniques to detect possible electricity theft: a proposed method," *The African Journal of Information and Communication*, vol. 35, no. 35, pp. 1–16, 2025, doi: 10.23962/ajic.i35.20652.

BIOGRAPHIES OF AUTHORS

Etienne François Mouckomey received his master's degree in Electrical Engineering in 2021. He is currently a Ph.D. student at the University of Douala. He is the chief information officer for E-government promotion/IT consultant at OIC-Okinawa, Japan. His areas of interest are machine learning, deep learning, and smart grids. He can be contacted at email: emoucko@gmail.com.

Prof. Jacques Bikai is a lecturer at the University of Ngaoundere. He is currently the head of the Department of Electrical Engineering at the University of Ngaoundere. He has published several articles in international journals. His research interests include soft computing, machine learning, electricity transportation, electrical power networks, and intelligent systems. He can be contacted at email: robertys.bikai@gmail.com.

Dr. Camille Franklin Mbey is a lecturer at ENSET of the University of Douala. He received his Ph.D. degree in 2021 at the University of Douala. His research interests include smart grids, electrical power networks, renewable power generation, artificial intelligence, and deep learning. He can be contacted at email: camillembey@yahoo.fr.

Dr. Felix Gislain Yem Souhe is a lecturer at IUT of the University of Douala. He received his Ph.D. degree in 2022 at the University of Douala. His research interests include electrical power grid, electrical consumption forecasting, artificial intelligence, and smart grids. He can be contacted at email: felixsouhe@gmail.com.

Prof. Alexandre Teplaira Boum is a full professor at University of Douala since 2024. He received his Ph.D. degree in 2014. He is currently the head of the Department of Basic Sciences at the University of Douala. His research interests include system optimization, smart power networks, deep learning, and MATLAB programming. He can be contacted at email: boumat2002@yahoo.fr.

Vinny Junior Foba Kakeu is a Ph.D. student at ENSET of the University of Douala. He received his master's degree in 2023 at the University of Douala. His research interests include photovoltaic power generation, artificial intelligence, smart grids, and deep learning. He can be contacted at email: kakeuvinn@gmail.com.