Turkish Journal of Electrical Engineering and Computer Sciences

Volume 32 | Number 4 Article 8

7-26-2024

Detection and classification of unauthorized use of irrigation motors in agricultural irrigation

ÖNDER CİVELEK

SEDAT GÖRMÜŞ

HALİL İBRAHİM OKUMUŞ

ORHAN GAZİ KEDEROGLU

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2024) 32: 605 – 622
© TÜBİTAK
doi:10.55730/1300-0632.4090

Detection and classification of unauthorized use of irrigation motors in agricultural irrigation

Önder CİVELEK[1]*, **Sedat GÖRMÜŞ**[2], **H.İbrahim OKUMUŞ**[3], **Orhan GAZİ KEDEROĞLU**[4]
[1]Trabzon Vocational High School, Karadeniz Technical University, Trabzon, Turkiye
[2]Department of Computer Engineering, Faculty of Engineering, Karadeniz Technical University, Trabzon, Turkiye
[3]Department of Electrical and Electronics Engineering, Karadeniz Technical University, Trabzon, Turkiye
[4]Fırat Elektrik Dağıtım A.Ş., Elazığ, Turkiye

Received: 05.02.2024 • **Accepted/Published Online:** 25.06.2024 • **Final Version:** 26.07.2024

Abstract: The decarbonisation of electricity generation requires real-time monitoring and control of grid components to efficiently and timely dispatch demand. This highly automated system, known as the Smart Grid, relies on smart or sensor-equipped distribution network components to optimise energy flow and minimise losses. However, energy theft, a major obstacle to efficient resource utilisation, poses a significant challenge to achieving this goal. This study proposes and evaluates a real-time telemetry and control system designed to mitigate energy theft in agricultural irrigation applications. The system increases energy efficiency by tracking the energy use in agricultural irrigation. The key challenge is to identify the source of illegal electricity consumption, classify it, and localise it. To address these difficulties, two distinct classification problems are addressed through the utilisation of machine learning methodologies. The initial classification task concerns the categorisation of loads that consume illegal electricity in agricultural irrigation. The subsequent classification problem pertains to the categorisation of feeder branches where such loads are activated. Therefore, a pilot distribution grid feeder has been simulated, and irrigation motors have been used as illegal loads which are activated at different points along the distribution feeder. The data collected from these simulations are used to create a data set where three-phase current data are collected from the transformer substation. The generated data set is employed to train machine learning models for the classification of illegal loads and feeder branches. The performance results of machine learning methods is obtained using the following metrics: accuracy, precision, recall, and F1-score. The results of the classification of loads stealing electricity in agricultural irrigation demonstrate that the bagged trees (BAT) algorithm achieves 99.64% in each criterion. In branch classification, the algorithm achieves the best results, with 97.64%, 97.40%, 96.22%, and 96.81%, respectively. Both classification performance results indicate that the proposed algorithm is effective for solving both classification problems.

This research demonstrates the efficacy of ML-powered real-time monitoring and control in combating energy theft and promoting efficient resource utilisation within agricultural irrigation networks. It is a pioneering study in the field of determining and classifying illegal loads in agricultural irrigation. Further research will investigate the potential for expanding the system's capabilities to include different load types and exploring alternative ML techniques for broader applicability within the context of low voltage distribution network monitoring.

Key words: Smart grid, load classification, machine learning, electricity theft, distribution network

*Correspondence: ondercivelek@ktu.edu.tr

1. Introduction

Energy losses in a distribution network are quantified as the discrepancy between the energy delivered to consumers and the metered amount reflected in billing. There are two primary sources of energy loss in a distribution network: technical loss (TL) and nontechnical loss (NTL). Technical loss encompasses power losses in electrical system components, such as distribution lines and transformers. In contrast, nontechnical loss arises from unforeseen external actions against the power system. Nontechnical loss represents the primary source of commercial losses, largely due to the difficulty in accurately measuring and rectifying such losses. One of the most significant issues facing NTL is the illegal use of electricity. This can take the form of bypassing electricity meters and direct interference with the distribution line. The global cost of electricity theft is estimated at 89.3 billion USD annually, of which 58.7 billion USD occurs in developing countries. India, Brazil, and Russia are the countries most economically affected by this phenomenon. In particular, the illegal connection of large loads such as irrigation motors directly to the grid has a detrimental impact on the lifespan of the grid and increases the risk of secure access to electricity.

In this study, a pilot region is selected in Malatya Province of Türkiye, where 226 measurement points are connected to 10 substations. Over the period from May 2021 to October 2021, a loss rate of 42% was observed in the pilot region. Over the 6-month period between May and October 2021, a total of 987,452 kWh of distributed energy is recorded, while a total of 568,826 kWh is billed. Of this, 418,626 kWh is not invoiced, indicating a significant discrepancy. It is estimated that approximately 90% of this unauthorised energy use is due to theft from the low voltage network. It has been established through field controls that the activation of loads, particularly for agricultural irrigation, plays a significant role in this unauthorised use.

Electricity theft in the low-voltage network is predominantly manifested as direct interference with the meter and the distribution line. Studies addressing electricity theft on the distribution line typically focus on unauthorised activities such as manipulation of consumer meters. Examples of these studies include [11–13, 15–23]. However, there are few studies on the detection of theft with direct contact to the low-voltage network [4–7]. In these studies, simulations are conducted without utilising actual network parameters. Deterministic methods are employed as the preferred approach. Due to the variable load profile in a real low-voltage network, the success rate of deterministic methods is likely to be low. Furthermore, the phase angles of the currents employed in deterministic methods are not considered when large power loads are activated. This may result in errors in the detection of unauthorised use. However, with the installation of advanced metering equipment on distribution lines, the use of smart meters and wireless networks, the detection and classification of such direct contact illegal loads has become possible with data-driven algorithms.

The electricity distribution company can detect the illegally activated loads within a limited period during the day, typically using the regional scanning method with the help of technical personnel. However, detection along long distribution lines in rural areas is very difficult and time consuming. Developing decision support systems that process data from end devices using machine learning algorithms can help prevent such issues. Machine learning algorithms provide advantages over traditional methods in terms of accuracy, efficiency, time consumption, precision and labour. In the event of electricity theft, machine learning algorithms are capable of detecting and classifying instances of electricity theft by analysing the pattern differences in the network size. This represents the inaugural study to utilise a wireless IoT network and machine learning to detect illegal electricity use in agricultural irrigation. Furthermore, it is a significant contribution to the field as it is the first study to detect and classify unauthorised uses that occur in the form of direct interference to a real distribution line with intelligent algorithms.

The rest of the article is organized as follows: the relevant background work in the studies on electricity theft in Section 2. Section 3 covers system architecture and simulation studies. The research findings and discussion are presented in Section 4. Finally, Section 5 outlines concluding remarks and possible future directions.

2. Related work

The studies on electricity theft can be classified into four main categories: game theory-based methods, power grid analysis-based methods, hardware-based methods, and machine learning-based methods.

The detection of electricity theft can be approached from a game-theoretic perspective, which employs the principles and techniques of game theory to analyse the dynamics between utility companies and electricity thieves. However, this strategy necessitates a comprehensive understanding of the participants' tactics and goals, which presents challenges in identifying an appropriate model to depict the intricate relationship between utility companies and electricity thieves [1]. Wei et al. [2] analyse the data marked as suspicious in smart meter data using the Benford method. They formulate a Stackelberg game model to analyze the strategic interactions between the grid and multiple theft locations. The Stackelberg equilibrium provides a likelihood ratio test (LRT) to perform sampling rate and threshold. The proposed method is validated against real usage data for four theft scenarios, and a successful theft detection rate of over 95% is achieved for each smart meter.

Methods based on power grid analysis detect abnormal electricity consumption behaviours by analysing power grid parameters such as current and line voltage. Kim et al. [3] propose an intelligent power distribution network model based on intermediate measurement monitors (IMMs) to analyze power flow in detail and effectively detect NTLs. The hardware architecture of IMMs is first defined, and an NTL detection algorithm is proposed to solve the energy balance linear equation (LSE) established between IMMs and collectors. A fairness coefficient is defined to measure consumers' susceptibility to energy theft. Simulation results demonstrate that the proposed NTL detection algorithm achieves a classification accuracy of 95%.

Leite et al. [4] propose a method based on a multivariate control chart that monitors voltage and current variances to detect NTLs in distribution networks. The main objective of this study is to develop a cost-effective methodology that can identify and locate NTLs caused by different types of cyber attacks. For this purpose, they use a path-finding method based on the A-Star algorithm to determine the consumption point where energy theft occurs, and integrate it with a geographic information system application to visualize the targeted consumption point affected by the cyber attack. The numerical results demonstrate the efficiency of the proposed methodology when applied to monitor a real distribution network.

There are studies that detect and localize electricity theft using deterministic methods in smart grids. Uvais [5] proposes a controller-based detection and localization system. Voltage and current measurements are taken from the feeder outputs of the distribution transformer and from the household meters. The collected data from the meters, along with the additional voltage drop and current on the distribution line, are used in a deterministic method to determine the location of the theft by activating circuit breakers. However, the fact that the study is only presented in a simple simulation environment and the potential harm to other subscribers when the circuit breakers are activated makes it challenging to implement the system in the field. An effective strategy against electricity theft is proposed by Raza et al. [6] to detect and locate the theft. The study uses a distribution line model that includes multiple transformers. To detect the theft, the real voltages of the distribution transformer feeders are collected from the advanced metering infrastructure. The voltages of the transformers are also calculated from the simulated system. The measured voltages are compared to the real voltages, and if the difference exceeds the average difference at any point, electricity theft is detected and its location is determined. Power flow analysis is performed to calculate NTL in the proposed algorithm. However, applying the algorithm to a single-transformer low-voltage distribution grid is challenging since accurate power flow analysis and NTL calculations cannot be performed. Metaliya et al. [7] propose a deterministic method based on voltage differences at nodes and line parameters for detecting electricity theft. The proposed method is implemented in the IEEE 33 distribution system and satisfactory results are obtained under various operating conditions. However, the study uses simulated data and has not been tested in the field. A method based on line parameters is proposed for detecting and localizing electricity theft in a low-voltage distribution grid by Nta et al. [8]. The method compares the mainline current with the current values of the meters to detect theft. For location detection, the voltages at the nodes where the meters are connected are compared. Additionally, the currents at the nodes are compared with the meter currents to identify the branch where the theft occurs. The method is applied to a real network feeder, and a success rate of 94% is achieved for feeder branch of the distribution system classification.

Hardware-based methods focus on developing protection devices or designing algorithms for detecting electricity theft based on specific equipment. Engelbrecht et al. [9] propose an anomaly detection device that identifies users engaged in electricity theft by comparing the deviation between real-time measurements recorded by a voltage and current detection circuit and estimated values obtained from a support vector machine (SVM) model. There are studies that aim to reduce the deployment costs of monitoring and control devices through optimization methods. Installing smart monitoring and control equipment such as feeder remote terminal units (FRTUs) or digital protective relays in a distribution network is an effective way to detect energy theft.

Liao et al. [10] present the strategic deployment of feeder remote terminal units (FRTUs) in the primary network, considering the cyber security of the distribution secondary network. They assume that the detection of historical abnormal load profiles in the secondary network is observable. These irregularities in historical energy usage can be identified from consumer billing centres using the proposed cyber security metrics. Although the number of FRTUs that can be deployed is constrained by budgetary limitations, the proposed algorithm identifies crucial locations for the installation of FRTUs at different time horizons. Simulation results demonstrate that the implementation of infrastructure improvements using the proposed multistage method enhances investment planning for distribution systems.

Machine learning-based methods analyze costumers' historical power consumption data along with other external data to detect abnormal electricity consumption behavior. Machine learning-based methods are divided into two groups: supervised learning and unsupervised learning. Supervised learning-based methods attempt to build a model using labeled data. Support vector machines (SVM) and artificial neural networks (ANN) are the most commonly used supervised learning methods for electricity theft detection.

Saeed et al. [11] present a detailed review of state-of-the-art methodologies for nontechnical losses published in the ACM Digital Library, Science Direct, and IEEE since 2000. The study mainly focuses on the proposed solutions, criteria, and drawbacks. The literature review in the present paper is primarily focused on nonhardware based solutions, with 79 out of 91 papers belonging to this category. The literature review revealed a significant gap in the category of theoretical methods, with a clear need to investigate the causes of NTLs in developing countries. In developed countries, the proportion of NTLs is much lower, but the consequences are significant. There is a lack of assessment of the financial viability of hardware-based methods.

Jindal et al. [12] present a new method that uses a hybrid of decision tree (DT) and SVM classifiers. Various features such as household size, season, time slot, and temperature are provided as inputs to DT. The DT algorithm calculates the expected electricity consumption for a consumer over a specific period. SVM classifier receives inputs such as household size, season, time slot, temperature, expected electricity consumption, and actual electricity consumption values. The SVM classifier classifies consumers as normal or theft consumers. The results demonstrate that the proposed method achieves a 92.5% accuracy rate and a low false positive rate of 5.12% in identifying theft consumers. Buzau et al. [13] propose a supervised learning-based method that utilizes smart meter data and additional information such as geographic location to analyse abnormal electricity consumption behaviour. Moreover, using a real dataset provided by Endesa, the leading company in Spain, it was shown that the XGBoost method outperforms k-nearest neighbours (kNN), SVM, and linear regression (LR) methods in terms of detection performance. An ANN model under different loads in the absence of electricity theft is trained Handique [14]. The smart meter electrical parameters are used as input data for the model. This model performs power estimation under different loads and detects electricity theft by comparing the estimated power value with the actual power value. In their study, the detection process for electricity theft is performed by comparing the current values from the nodes with the main current value. For the location detection of theft occurring in different regions of the distribution line, a deterministic method utilizing line parameters is proposed. The method is applied to a field network, achieving a location accuracy of 94%.

Khan et al. [15] present a hybrid deep learning model for the effective detection of electricity thieves in smart grids. First, preprocessing techniques are employed to clean the data from smart meters. Subsequently, feature extraction techniques such as AlexNet address the dimensionality issue. The effectiveness of the proposed method is evaluated through simulations using a real dataset of Chinese smart meters. Various benchmark models are also applied to perform a comparative analysis. The proposed model achieved accuracy, precision, recall, and F1 values of up to 86%, 89%, 86%, and 84%, respectively.

Gunduz et al. [16] attempt to detect energy theft using CNN-based models based on real electricity consumption data from 2104 residential users. The models attempt to predict both honest and malicious energy consumption patterns of the users. CNN is used to identify attack patterns within consumer behaviour. In particular, six attack vectors are used to generate malicious readings from honest samples in a real energy consumption dataset. CNN-based detectors are then proposed to detect energy theft. Furthermore, the GAN algorithm is employed to address the problem of unbalanced data. In the CNN-based model, machine learning algorithms such as SVM, RF, LR, kNN, and DT are applied to the problem as benchmarks. The results show that the proposed CNN+LR and CNN+RF models are very promising classification methods for energy theft detection.

In the literature, there are significant studies on identifying potential attack vectors and malicious costumers using smart meter consumption data for energy theft detection.

Jokar et al. [17] employ SVM in advanced metering infrastructure to propose an energy theft detection system based on consumption patterns, achieving a classification accuracy of 94%. Additionally, they consider a range of cyber attack vectors associated with energy theft, which are well-documented in the literature. The authors of [18] introduce a two-stage energy theft detection system utilising DTs and SVM, achieving an accuracy of 92.5%. However, there is no indication regarding the balance of the dataset. Researchers present an energy theft detection method using ML models, aiming to enhance detection rate and reduce error rates by combining various ML techniques. The results demonstrate that the ensemble ML approach outperforms boosting, although not to the extent of other methods.

In [19], although based on simulated data, a neural network implementation achieve a notable success with an overall detection rate of 93%. The authors of [20] proposed a novel approach employing a multilayer perceptron artificial neural network to discern energy theft in distribution systems, achieving an average detection rate of 93.4%. However, the dataset balance information is lacking. In [21], a hybrid deep neural network approach is proposed, combining CNNs, gated recurrent units, and particle swarm optimization. Despite its complexity, the proposed hybrid model does not exhibit superior accuracy and tends to overfit. In [22], a deep recurrent neural network classifier utilising gated recurrent units achieves a detection rate of up to 93%, although the dataset balance is unspecified. Meanwhile, [23] employs convolutional long short-term memory, achieving an accuracy of up to 98%, without specifying the dataset balance.

3. Materials and methods

Machine learning is an artificial intelligence technique that enables computers to learn from experiences. Machine learning algorithms employ computational methodologies to derive information directly from data, eschewing the use of a predefined equation as a model. As the number of available examples for learning increases, algorithms demonstrate an ability to adaptively improve their performance. Machine learning is divided into two main categories: supervised learning and unsupervised learning. In supervised learning, an algorithm utilises a known input dataset and the corresponding known responses (output) to train a model that can make reasonable predictions for new data. Supervised learning is employed when there is a known dataset for the output being predicted. Supervised learning employs classification and regression techniques to develop machine learning models. If the data can be labelled, categorised, and divided into specific groups or classes, classification methods can be employed on this data. Some of the most commonly employed supervised learning algorithms include linear regression, logistic regression, decision trees, random forest, support vector machines (SVM), naive Bayes, and neural networks. In contrast, unsupervised learning identifies patterns within data that are not explicitly stated. The objective of unsupervised learning is to make inferences from input datasets without labelled responses. Clustering is the most prevalent unsupervised learning technique. Other methods, such as dimensionality reduction algorithms, association rule learning, and generative models, may also be considered. The present study employs supervised learning techniques for the classification of illegal loads and branches.

3.1. Ensemble methods

The ensemble method is a technique that aims to obtain more accurate and reliable predictions by combining multiple machine learning models. This method attempts to reduce model errors by utilising the strengths of the models. Ensemble methods provide more accurate and reliable predictions by combining multiple models. By combining the strengths of different models, these methods reduce model bias and keep variance under control. These methods can be analysed in three groups as bagging, boosting, and stacking.

3.1.1. Bagging method

The bagging method serves as the foundation for numerous well-known algorithms, including the random forest algorithm. A random forest is an ensemble of multiple decision trees. Decision trees are susceptible to overlearning issues, which can be mitigated by constructing a lower variance prediction model through the formation of an ensemble. A model with a lower variance is more likely to generalise effectively. When a high variance model learns the data very well, it can overfit the training data set and lead to incorrect predictions in the test data. This situation is referred to as overfitting. The bagging method consists of two steps to reduce model errors: sampling and merging. In sampling, subsets are created from the data set and these subsets form bootstrap data sets using repeated sampling. Each bootstrap dataset is used to train a model. In fusion, each model makes predictions independently. The predictions are then combined to obtain an overall prediction. A common fusion method for classification problems is the maximum voting method. In this method, each model makes a prediction, and the most frequent prediction is selected as the overall prediction.

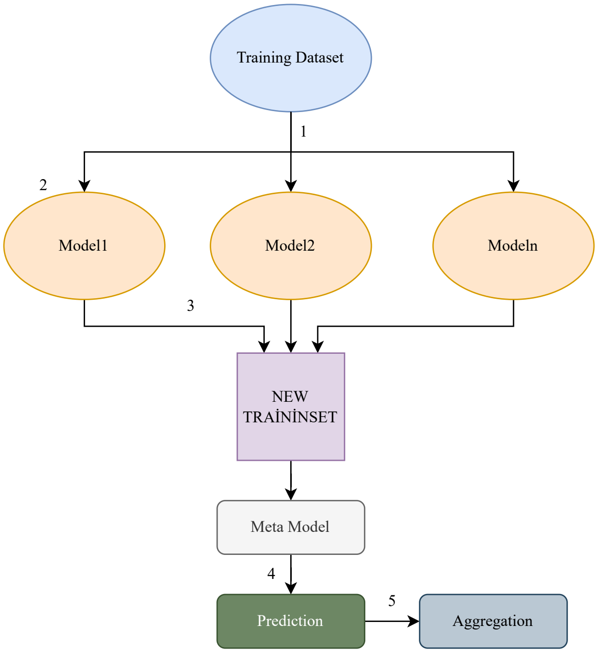


Figure 1. The steps involved in the bagging method.

The steps involved in the bagging method, as illustrated in Figure 1, can be summarized as follows: 1. Preparation of a training dataset comprising n samples. 2. Selection of m subsets from the dataset, with N samples taken from each subset. 3. Training of models. 4. Estimation of models. 5. The overall prediction value is obtained by averaging the predictions for a regression problem. In this study, the bagging method (BAT) is employed for the classification of illegal loads and feeder branches of the distribution system. The hyperparameters of the algorithm are as follows: maximum number of splits: 1649, number of learners: 30.

3.2. System architecture

In this study, a monitoring system based on an RF Mesh network is used to detect instances of theft in agricultural irrigation systems. In this setup, the meters periodically send instantaneous power drawn by the customer loads. By leveraging this wireless IoT network, the total periodical energy consumption reported by the smart meters is compared with the total demand observed in the distribution substation. The diagram of this system is provided in Figure 2. As illustrated in Figure 2, the measurement components of the system are integrated into both the distribution transformer and the customer meters. The embedded computer system positioned at the substation continuously monitors the current and voltage values of the distribution feeders. Once an illegal usage event is detected, machine learning (ML) based algorithms are triggered to classify the load. The detection process involves a two-stage active power control. Firstly, a comparison is made between the total active power drawn from the transformer and the total active power drawn from the feeder meters. This helps determine whether any unauthorised load has come online. Secondly, a feeder-based detection process is carried out. For this purpose, the active power drawn from the feeder is compared to the active power drawn from the feeder meters. Assuming a 7% technical loss, a threshold value of 0.93 is set. To accurately capture the instance of the motor activation, each substation feeder is sampled at 1.4 kHz using analog-to-digital converters capable of 10-bit sampling. The sampled data is used to train and test the machine learning (ML) algorithms evaluated in this work.

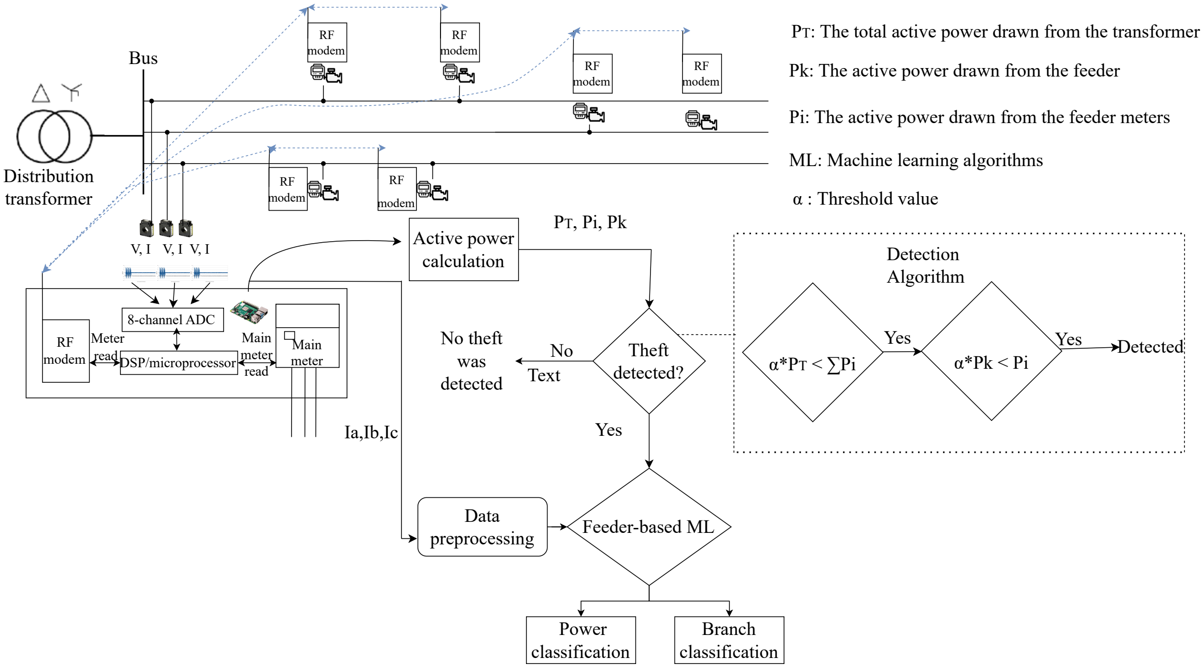


Figure 2. System architecture

3.3. Simulation study

To validate the proposed method, a pilot distribution line in Malatya/Türkiye is modelled in MATLAB/Simulink environment as shown in Figure 3. In addition, the residential loads used in the modelled feeder are given in Table 1. Electromagnetic transient simulations are created on the model using the simulation parameters given in Table 2. For the results of machine learning algorithms, the regression learner tool is used.

Table 1. Household loads used in the simulation.

P(W)	1000	2000	1000	2000	3000	2000	3000	1000	1500	750
Q(VAR)	750	1500	750	1500	2250	1500	2250	750	1000	500

Table 2. The dataset parameters

Location (m)	188, 298, 405, 511, 983, 1137, 539, 640, 694, 562, 593
Motor Power (kw)	1.5, 4, 7.5, 15, 22
Torque (% nominal torque)	20,40,60,80,100
Data	2 seconds three-phase current data
Branch	B1 (L2-L7), B2 (L8-L12), B3 (L10-L11)

The feeder is divided into zones of varying lengths, with irrigation motors activated in each zone. The electrical characteristics of the motors (power and torque) and their activation location result in distinct patterns in the three-phase current waveforms. Therefore, three-phase current data are recorded from the sending end of the feeders. Waveform windows of 2 s are used to train the machine learning algorithms. A scenario is constructed on the modeled feeder to simulate the use of illegal electricity by various residential loads and motors of different powers. A total of 2200 events are simulated for the scenario, with three-phase current data covering a 2-second window collected at the sending end of the feeder at a sampling frequency of 1.4 kHz for each event. Consequently, a dataset with dimensions of 2200 × 8400 is generated, with rows representing the total number of events and columns corresponding to the length of the three-phase current data.

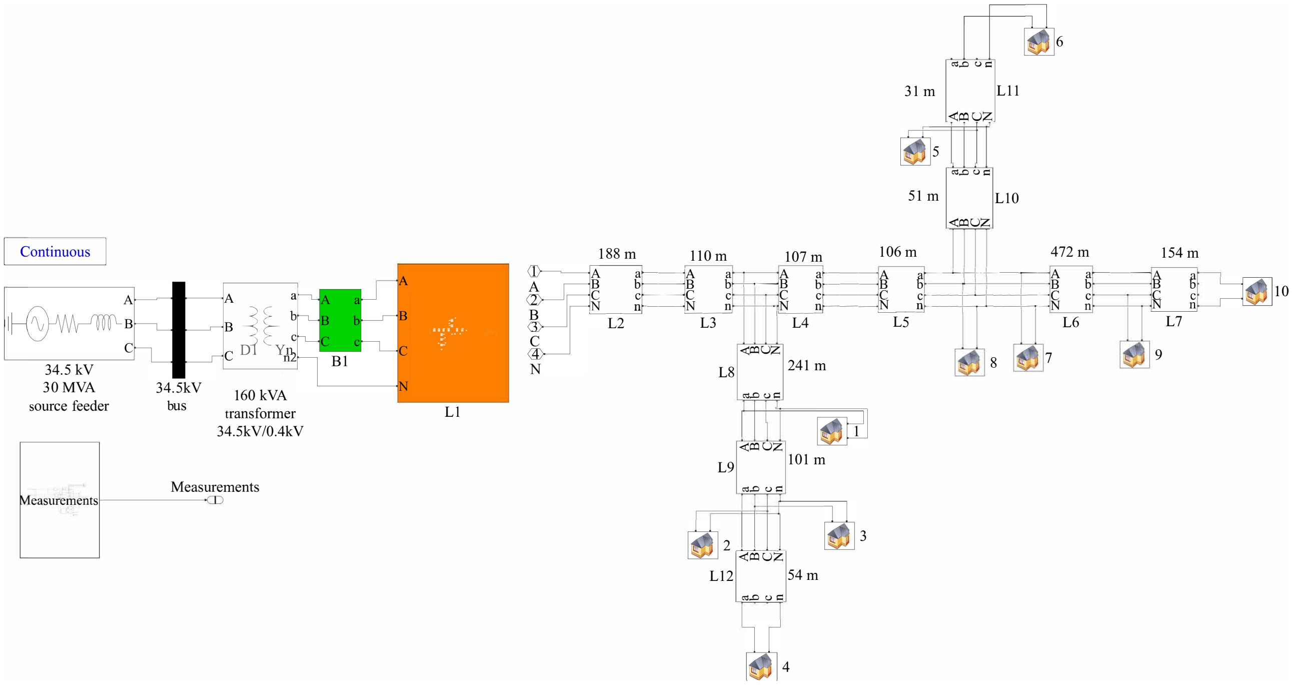


Figure 3. Simulation of the Pilot Feeder

Twenty-five percent of the data set is used as test data. In this study, to protect against overfitting, a 5-fold cross-validation method is employed. In this method, the dataset is shuffled randomly and divided into 5 groups. One group is selected as the validation set while the remaining 4 groups are used as the training set. A model is built using the training set and evaluated using the validation set to determine the error rate.

Figure 4 presents voltage and current waveforms measured from the transformer feeder under normal conditions. Figures 5–7 show the waveforms of current and voltage values measured at the transformer feeder when a 7.5 kW motor is connected as an unmetered load at different locations. As seen from the graphs, voltage drops and changes in the current waveform are noticeable. As the location of the motors connected as unmetered loads increases, the voltage value decreases depending on the characteristics of the line, and significant changes can be observed in the current waveform.

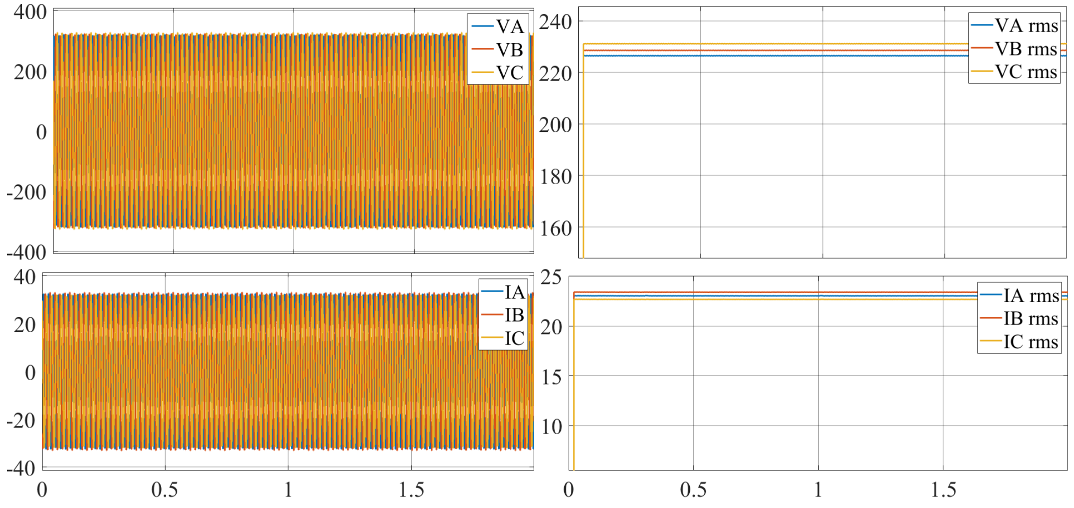


Figure 4. Current and voltage values measured at the transformer feeder under stationary load

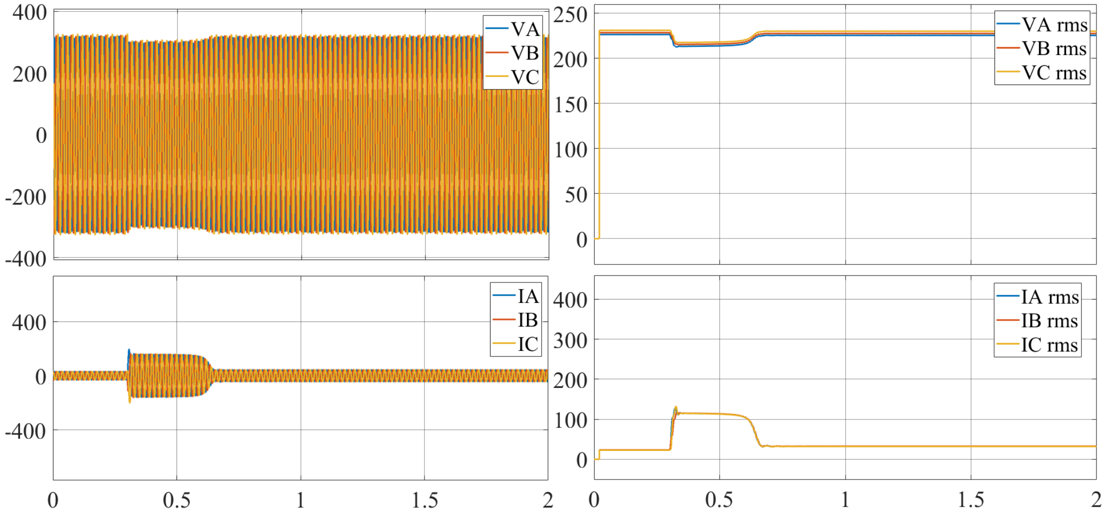


Figure 5. Current and voltage values measured at the transformer feeder when an unmetered 7.5 kW irrigation motor is activated at a distance of 188m

Our research in the pilot region reveal that the loads used in agricultural irrigation systems are submersible pump motors with power ratings generally ranging from 1.5 kW to 22 kW. Submersible water pumps driven by asynchronous motors are commonly used. The pump section provides the load torque to the asynchronous motor. These motors are connected to the distribution line via a direct line connection. Instead of focusing on pump characteristics such as flow rate, static pressure, and head, this study focuses on obtaining electrical characteristics from the asynchronous motor. Since changes in pump characteristics result in different load torques on the asynchronous motor, different torque values have been used as load torque in the data set.

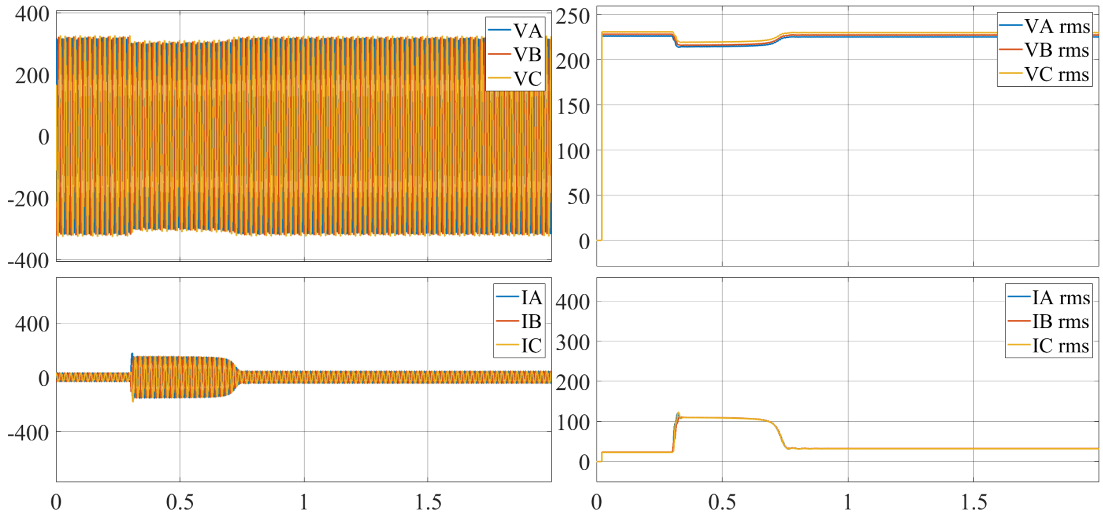


Figure 6. Current and voltage values measured at the transformer feeder when a 7.5 kW irrigation motor is illicitly used at a distance of 511m

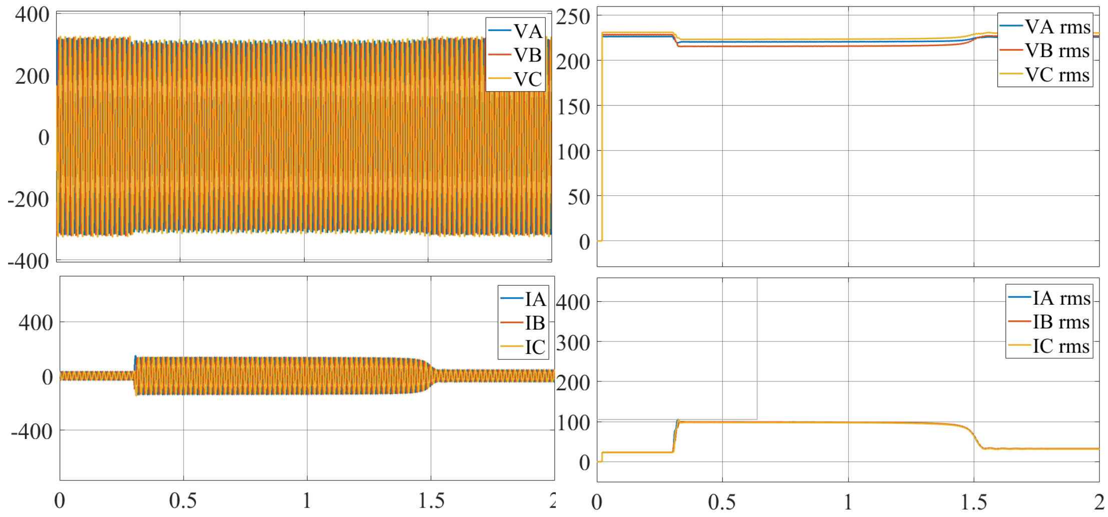
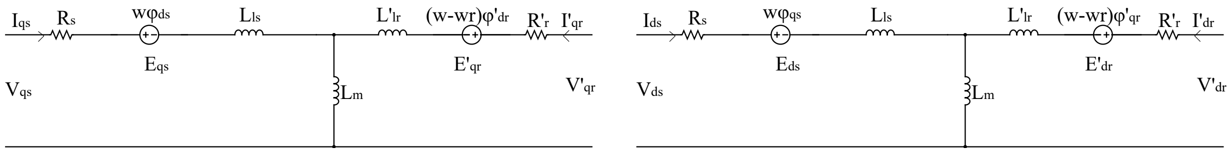


Figure 7. Current and voltage values measured at the transformer feeder when a 7.5 kW irrigation motor is illicitly used at a distance of 1137m

3.4. The dynamic model of an asynchronous motor

The arbitrary (dq) rotating reference frame model is commonly used in the mathematical modeling of asynchronous motors as given in Figure 8.

When writing the model equations, it is assumed that the system operates in a reference frame rotating at an arbitrary speed. This simplifies the differential equations by making their coefficients time-varying. The equations for the d and q equivalent circuits used in the dynamic model are given below. Additionally, Table 3 provides explanations for the symbols used. Stator voltage in the q-axis as presented in Equation 1, Stator voltage in the d-axis is presented in Equation 2, rotor voltage in the q-axis is presented in Equation 3, and rotor voltage in the d-axis is presented in Equation 4. Finally, electromagnetic tork is presented in Equation 5.



(a) d-axis

(b) q-axis

Figure 8. Equivalent circuit model of an asynchronous motor in an arbitrary reference frame.

$$V_{qs} = R_s\ I_{qs} + \frac{d}{dt}\varphi_{qs} + w\varphi_{ds} \tag{1}$$

$$V_{ds} = R_s\ I_{ds} + \frac{d}{dt}\varphi_{ds} + w\varphi_{qs} \tag{2}$$

$$V'qr = R'rI'qr + \frac{d}{dt}\varphi'_{qr} + (w - w_r)\varphi'_{dr} \tag{3}$$

$$V'dr = R'rI'dr + \frac{d}{dt}\varphi'_{dr} + (w - w_r)\varphi'_{qr} \tag{4}$$

$$T_e = 1.5\ (\varphi_{ds}\ I_{qs} - \varphi_{qs}\ I_{ds}) \tag{5}$$

Table 3. The symbols used in the model equations.

V_{qs}, I_{qs}	q-axis stator voltage and current
V_{ds}, I_{ds}	d-axis stator voltage and current
V'_{qr}, I'_{qr}	q-axis rotor voltage and current
V'_{dr}, I'_{dr}	d-axis rotor voltage and current
R_s, R'_r	Stator ve rotor resistance
w, w_r	Electrical angular velocities
φ_{ds}, φ_{qs}	Stator q-axis and d-axis fluxes
φ'_{ds}, φ'_{qs}	Rotor q-axis and d axis fluxes
T_e	Electromagnetic torque

Models of pump motors with powers of 1.5 kW, 4 kW, 7.5 kW, 15 kW, and 22 kW (asynchronous) are employed in the MATLAB program. The electrical parameters of these motors are given in Table 4.

3.5. Model evaluation metrics

Confusion matrix is the most important metric used to evaluate the performance of classification models. It compares the predicted class labels with the true class labels and shows the number of correct and incorrect classifications for each class. For a binary classification model, the confusion matrix and its components are provided in Table 5. These values are used to assess the performance of the model.

TP: true positive, FP: false positive, FN: false negative, TN: true negative

Table 4. The electrical parameters of asynchronous motors.

Power (kW)	1.5	4	7.5	15	22
Speed (rpm)	1428	1430	1440	1460	1440
Frequency (Hz)	50	50	50	50	50
Stator current (A)	3.64	9	16	31	44
Stator resistance (Ω)	4.85	1.405	0.7384	0.2147	0.12
Rotor resistance (Ω)	3.805	1.395	0.7402	0.2205	0.18
Stator inductance (L)	0.274	0.005839	0.003045	0.000991	0.12
Rotor inductance (L)	0.274	0.005839	0.003045	0.000991	0.12
Mutual inductance (Lm)	0.258	0.1722	0.1241	0.06419	0.0035
Number of poles	2	2	2	2	2
Moment of inertia (kg/m2)	0.031	0.0131	0.0343	0.102	0.33
Friction coefficient (Nm.s/rad)	0.00114	0.002985	0.000503	0.009541	0.02791

Table 5. Confusion matrix for a binary classification model.

	Predicted Negative	**Predicted Positive**
Actual Negative	TN	FP
Actual Positive	FN	TP

True positive rate (TPR): Also known as sensitivity, TPR represents the rate of correctly detected true positive instances. It is calculated as:

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

Accuracy (ACC): Accuracy is the ratio of correctly classified instances to the total number of instances. It is calculated as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

False positive rate (FPR): FPR represents the rate of false positive instances among the total negative instances. It is calculated as:

$$FPR = \frac{FP}{FP + TN} \tag{8}$$

These metrics provide additional insights into the performance of a classification model. The receiver operating characteristic (ROC) curve illustrates the relationship between TPR and FPR for different threshold values of classification scores. It is an important metric used in classification problems. The ROC curve plots TPR on the y-axis against FPR on the x-axis, with each point on the curve representing a different threshold for classifying instances as positive or negative. By varying the threshold, the trade-off between TPR and FPR can be visualized. The area under the ROC curve (AUC) is a significant metric used to measure the classifier's discriminatory power. The AUC value ranges between 0 and 1, with a higher value indicating better classifier performance. A high AUC value indicates a better ability of the classifier to distinguish between positive and negative instances.

Precision: Precision is defined as the ratio of true positive predictions to total predictions. In other words, it indicates the proportion of true positive predictions made by a classification model. The higher the precision, the lower the number of false positive predictions. It is calculated as:

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

Recall: Recall is defined as the rate at which true positives are detected. It indicates the proportion of true positive examples correctly identified by a classification model. The higher the recall, the lower the number of false negatives. It is calculated as:

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

F1 score: The F1 score is a metric that represents the harmonic mean of the precision and recall metrics. It is used to evaluate the performance of a model by considering both the precision and recall metrics. The F1 score is particularly useful in situations where both false positives and false negatives need to be reduced in a balanced way. It is calculated as:

$$F1score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{11}$$

4. Results and discussion

The labelling information of the irrigation motor used as illegal loads and the branches where the motors are activated are provided in Table 6. Table 7 shows how well the motors can be classified according to their use of illegal electricity. The best results are produced by decision trees, SVM algorithms, and artificial neural networks. The algorithms achieve 99.64% success in terms of accuracy, precision, recall, and F1-score. Table 8 presents the results of feeder branch classification. The BAT algorithm achieves 97.64%, 97.40%, 96.22%, and 96.81% success rates for accuracy, precision, recall, and F1-score, respectively. These results demonstrate that the BAT algorithm produces the most optimal results.

Table 6. Labels for a binary classification model.

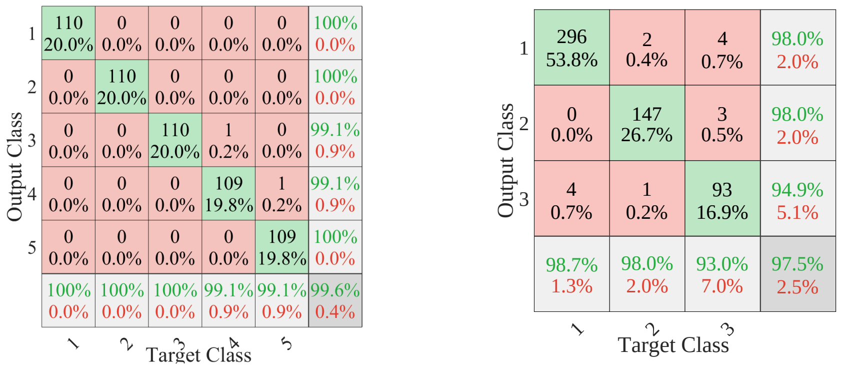
P(Kw)	Branch	Label
1.5	B1 (L2-L7)	1
4	B2 (L8-L12)	2
7.5	B3 (L10-L11)	3
15		4
22		5

Figure 9(a) presents the confusion matrix table, which illustrates the results of the proposed method for classifying illegal loads. The diagonal cells indicate the number and percentage of correct classifications made by the trained network. For instance, 110 events are correctly classified as belonging to class 1, which corresponds to 20% of all 550 events. Similarly, 110 events are correctly classified as belonging to class 2. A total of 110 incidents are correctly classified as belonging to class 3, 109 incidents are correctly classified as belonging to class 4, while one incident is classified as class 3. This corresponds to 0.2% of all events. While 109 incidents belonging to class 5 are correctly classified, one incident is classified as class 4. Likewise, this corresponds to

0.2% of all events. Overall, 99.6% of the predictions are correct, with 0.4% being incorrect. The confusion matrix in Figure 9(b) illustrates the performance of the proposed method in terms of feeder branch predictions. Of the 300 events belonging to Class 1, 98.7% of the predictions are correct, with 1.3% being incorrect. Of the 150 events belonging to Class 2, 98% of the predictions are correct, while 2% are incorrect. Of the 100 events belonging to Class 3, 93% of the predictions are correct, while 7% are incorrect. Overall, 97.5% of the predictions are correct, while 2.5% are incorrect.

Table 7. Results of illegal load classification.

Model	Accuracy	Precision	Recall	F1-score
Fine tree (FT)	0.9964	0.9964	0.9964	0.9964
Medium tree (MT)	0.9964	0.9964	0.9964	0.9964
Coarse tree (CT)	0.9964	0.9964	0.9964	0.9964
Linear discriminant	0.9945	0.9946	0.9945	0.9946
Efficient logistic regression	0.9945	0.9945	0.9945	0.9945
Efficient linear SVM (ELSVM)	0.9964	0.9964	0.9964	0.9964
Gaussian naive Bayes (GNB)	0.9927	0.9928	0.9927	0.9928
Kernel naive Bayes (KNB)	0.9782	0.9799	0.9782	0.9791
Linear SVM (LSVM)	0.9964	0.9964	0.9964	0.9964
Quadratic SVM (QSVM)	0.9964	0.9964	0.9964	0.9964
Cubic SVM	0.9964	0.9964	0.9964	0.9964
Fine Gaussian SVM (FGSVM)	0.9618	0.9664	0.9618	0.9641
Medium Gaussian SVM (MGSVM)	0.9964	0.9964	0.9964	0.9964
Coarse Gaussian SVM (CGSVM)	0.9964	0.9964	0.9964	0.9964
Fine kNN (FkNN)	0.9909	0.9911	0.9909	0.9910
Medium kNN (MkNN)	0.9891	0.9892	0.9891	0.9891
Coarse kNN (COkNN)	0.9782	0.9782	0.9782	0.9782
Cubic kNN (CUkNN)	0.9818	0.9820	0.9818	0.9819
Weighted kNN (WkNN)	0.9927	0.9928	0.9927	0.9928
Bagged trees (BAT)	0.9964	0.9964	0.9964	0.9964
Narrow neural network (NNN)	0.9964	0.9964	0.9964	0.9964
Medium neural network (MNN)	0.9964	0.9964	0.9964	0.9964
Wide neural network (WNN)	0.9945	0.9945	0.9945	0.9945



	1	2	3	4	5	
1	110 20.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	110 20.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	110 20.0%	1 0.2%	0 0.0%	99.1% 0.9%
4	0 0.0%	0 0.0%	0 0.0%	109 19.8%	1 0.2%	99.1% 0.9%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	109 19.8%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	99.1% 0.9%	99.1% 0.9%	99.6% 0.4%

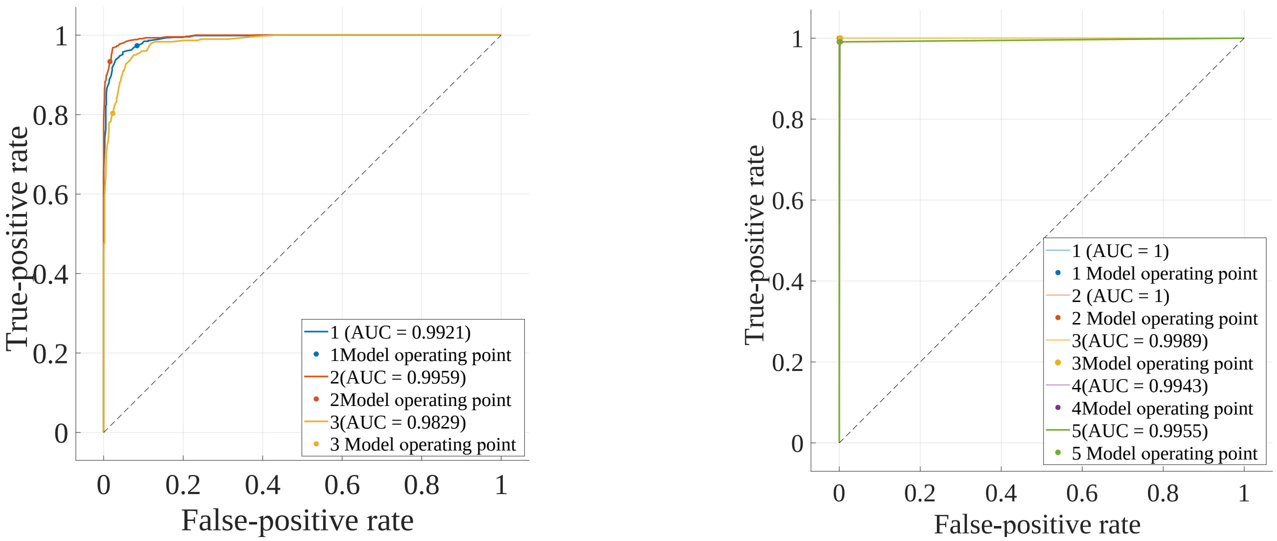
(a)

	1	2	3	
1	296 53.8%	2 0.4%	4 0.7%	98.0% 2.0%
2	0 0.0%	147 26.7%	3 0.5%	98.0% 2.0%
3	4 0.7%	1 0.2%	93 16.9%	94.9% 5.1%
	98.7% 1.3%	98.0% 2.0%	93.0% 7.0%	97.5% 2.5%

(b)

Figure 9. Confusion matrix illustrating the clasification performance of the BAT algorithm for (a) illegal load and (b) feeder branch of the distribution system.

The ROC curve in Figure 10(a) represents the performance of the BAT algorithm for the classification of illegal loads. In contrast, the ROC curve in Figure 10(b) illustrates the performance of the feeder branch classification. From both ROC curves, it can be inferred that the AUC area on the ROC curves is approximately equal to one. This also indicates that the BAT algorithm achieves high performance in both classifications.



(a) ROC Curve Illustrating the Performance of the BAT algorithm for illegal load clasification

(b) ROC Curve Illustrating the Performance of the BAT algorithm for feeder branch of the distribution system clasification

Figure 10. ROC curve illustrating the performance of the BAT algorithm.

5. Conclusions and future work

This study proposes a system to significantly reduce the rate of electrical theft in agricultural irrigation systems. First, an architecture is established to enable the real-time monitoring of electricity consumption data. In this architecture, the metered and dispatched active power data are compared to identify the feeder where the unmetered load is connected. Subsequently, the acquired three-phase current data are processed using machine learning algorithms to achieve high accuracy in the classification of illegal loads and the distribution feeder branches. The proposed system has demonstrated its effectiveness in detecting tapping-based thefts in the distribution network. In particular, the system equipped with the BAT algorithm has been proven to be effective in providing accurate classification in terms of load power level and feeder branch of the distribution system. The results obtained highlight the importance of utilising machine learning algorithms in smart grids.

Future studies will focus on different scenarios with additional load types. In particular, scenarios with different types of inductive loads that consume illegal electricity will be studied. In addition, deterministic methods will be used to determine the loads that are activated and the results will be compared with the results of the proposed method. In order to assess the efficacy of the proposed method in a real-world setting, field measurements will be taken in a variety of scenarios. The performance of the method in these scenarios will be analysed and compared with the results obtained from simulations. Furthermore, the performance of deep learning algorithms in terms of load type and feeder branch location classification will be investigated.

Table 8. Results of feeder branch of the distribution system classification.

Model	Accuracy	Precision	Recall	F1-score
Fine tree (FT)	0.9236	0.9098	0.8978	0.9038
Medium tree (MT)	0.8055	0.8169	0.7356	0.7741
Coarse tree (CT)	0.6200	0.7582	0.4267	0.5460
Linear discriminant (LD)	0.9109	0.8908	0.8944	0.8926
Efficient logistic regression (ELR)	0.5964	0.6864	0.4100	0.5133
Efficient linear SVM (ELSVM)	0.5418	0.5774	0.4000	0.4726
Gaussian naive Bayes (GNB)	0.5655	0.4988	0.4267	0.4599
Kernel naive Bayes (KNB)	0.6273	0.6269	0.4933	0.5222
Linear SVM (LSVM)	0.6909	0.7604	0.5433	0.6338
Quadratic SVM (QSVM)	0.8509	0.8379	0.8122	0.8249
Cubic SVM (CSVM)	0.8964	0.8821	0.8822	0.8822
Fine Gaussian SVM (FGSVM)	0.6400	0.7311	0.4900	0.5868
Medium Gaussian SVM (MGSVM)	0.6257	0.8251	0.4878	0.6131
Coarse Gaussian SVM (CGSVM)	0.5964	0.7589	0.4033	0.5267
Fine kNN (FkNN)	0.7891	0.7661	0.8011	0.7832
Medium kNN (MkNN)	0.6273	0.6141	0.4822	0.5402
Coarse kNN (COkNN)	0.5673	0.5788	0.3711	0.4522
Cubic kNN (CUkNN)	0.6164	0.6223	0.4633	0.5312
Weighted kNN (WkNN)	0.7964	0.7787	0.7889	0.7837
Bagged trees (BAT)	0.9764	0.9740	0.9622	0.9681
Narrow neural Network (NNN)	0.9018	0.8826	0.8756	0.8791
Medium neural Network (MNN)	0.9345	0.9206	0.9256	0.9231
Wide neural Network (WNN)	0.9091	0.8978	0.8922	0.8950

Acknowledgment

This research is supported by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) Research Fund under project number 7220360.

References

[1] Liao W, Yang Z, Bak-Jensen B, Pillai R, Krannichfeldt L et al. Simple data augmentation tricks for boosting performance on electricity theft detection tasks. IEEE Transactions on Industry Applications 2023; 59 (4): 4846–4858. https://doi.org/10.1109/TIA.2023.3262232

[2] Wei L, Sundararajan A, Sarwat A, Biswas S, Erfan I. A distributed intelligent framework for electricity theft detection using Benford's law and Stackelberg game 2017; Resilience Week; Wilmington, DE, USA. pp. 5-11, https://doi.org/10.1109/RWEEK.2017.8088640

[3] Kim J, Hwang Y, Sun Y, Sim I, Kim D et al. Detection for non-technical loss by smart energy theft with ıntermediate monitor meter in smart grid. In: IEEE Access 2019; 7: 129043-129053. https://doi.org/10.1109/ACCESS.2019.2940443

[4] Leite B, Mantovani S. Detecting and locating non-technical losses in modern distribution networks. IEEE Transactions on Smart Grid 2018; 9 (2): 1023-1032. https://doi.org/10.1109/TSG.2016.2574714

[5] Uvais M. Controller based power theft location detection system. International Conference on Electrical and Electronics Engineering (ICE3); Gorakhpur, India, 2020. pp. 111-114, https://doi.org/10.1109/ICE348803.2020.9122940.

[6] Raza H, Imran K, Khattak A, Ulasyar A, Ilyas A. Strategy to detect quantify and locate power theft in a distribution network. 6th International Electrical Engineering Conference; NEDUET, Karachi, Pakistan, 2021.

[7] Metaliya A, Deshpande A. Electricity theft detection scheme using energy loss and voltage estimation in distribution network. International Journal of Electrical Engineering and Technology 2021; pp. 31-40.

[8] Nta E, Udofia K, Okpura N. Development of an energy theft detection and location system for low voltage power distribution networks. Journal of Multidisciplinary Engineering Science and Technology 2022; 9 (4): 15240-15249.

[9] Engelbrecht J, Hancke P, Osifeko M. Design and implementation of an electrical tamper detection system. 45th Annual Conference of the IEEE Industrial Electronics Society; Lisbon, Portugal. 2019. pp. 2952-2957. https://doi.org/10.1109/IECON.2019.8927476

[10] Liao C, Ten C, Hu S. Strategic FRTU deployment considering cybersecurity in secondary distribution network. IEEE Transactions on Smart Grid 2013; 4 (3): 1264-1274. https://doi.org/10.1109/TSG.2013.2256939

[11] Saeed MS, Mustafa MW, Hamadneh NN, Alshammari NA, Sheikh UU et al. Detection of non-technical losses in power utilities—a comprehensive systematic review. Energies 2020; 13: 4727. https://doi.org/10.3390/en13184727

[12] Jindal A, Dua A, Kaur K, Singh M, Kumar N et al. Decision tree and SVM-based data analytics for theft detection in smart grid. IEEE Transactions on Industrial Informatics 2016; 12 (3): 1005-1016. https://doi.org/10.1109/TII.2016.2543145

[13] Buzau M, Tejedor-Aguilera J, Cruz-Romero P, Gómez-Expósito A. Detection of nontechnical losses using smart meter data and supervised learning. IEEE Transactions on Smart Grid 2019; 10 (3): 2661-2670. https://doi.org/10.1109/TSG.2018.2807925

[14] Handique M, Kalita Q, Das G. Design and simulation of electricity theft detection in radial distribution system. ADBU Journal of Electrical and Electronics Engineering (AJEEE) 2019; 3 (2): 44-49.

[15] Khan N, Raza M, Ara D, Bouzguenda B. A deep learning technique Alexnet to detect electricity theft in smart grids. Frontiers in Energy Research 11: pp.1287413. https://doi.org/10.3389/fenrg.2023.1287413

[16] Gunduz MZ, Das R. Smart grid security: an effective hybrid CNN-based approach for detecting energy theft using consumption patterns. Sensors (Basel). 2024;24 (4):1148. https://doi.org/10.3390/s24041148

[17] Jokar P, Arianpoo N, Leung VCM. Electricity theft detection in AMI using costumers' consumption patterns. IEEE Transactions on Smart Grid 2016; 7: 216–226. https://doi.org/10.1109/TSG.2015.2425222

[18] Gunturi SK, Sarkar D. Ensemble machine learning models for the detection of energy theft. Electric Power Systems Research 2021; 192: 106904. https://doi.org/10.1016/j.epsr.2020.106904

[19] Alromih A, Clark JA, Gope P. Electricity theft Detection in the presence of prosumers using a cluster-based multi-feature detection model; Proceedings of the 2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm); Aachen, Germany; 2021; pp. 339–345.

[20] Souza MA, Pereira JLR, Alves GDO, de Oliveira BC, Melo ID et al. Detection and identification of energy theft in advanced metering infrastructures. Electric Power Systems Research 2020; 182:106258. https://doi.org/10.1016/j.epsr.2020.106258

[21] Ullah A, Javaid N, Samuel O, Imran M, Shoaib M. CNN and GRU based deep neural network for electricity theft detection to secure smart grid. Proceedings of the 2020 International Wireless Communications and Mobile Computing (IWCMC); Limassol, Cyprus 2020; 1598–1602.

[22] Nabil M, Ismail M, Mahmoud M, Shahin M, Qaraqe K et al. Deep recurrent electricity theft detection in AMI networks with random tuning of hyper-parameters. Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR); Beijing, China 2018; 740–745.

[23] Zheng Z, Yang Y, Niu X, Dai HN, Zhou Y. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. IEEE Transactions on Industrial Informatics 2018; 14: 1606–1615. https://doi.org/10.1109/TII.2017.2785963