JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Review on Real-time Electricity Theft Detection in Smart Grids using ANN and DNN

[1]P. ELAIYARAJA [2]S.BABU

[1]M.Tech Student, Department of Computer Science and Engineering, Kuppam Engineering College, KES Nagar, Kuppam, Andhra Pradesh, 517425, India

[2]Assistant Professor, Department of Computer Science and Engineering, Kuppam Engineering College, KES Nagar, Kuppam, Andhra Pradesh, 517425, India

Abstract - Electricity theft poses a major problem for utility companies, leading to financial losses, safety risks, and increased costs for legitimate consumers. With the deployment of smart grids and Advanced Metering Infrastructure (AMI), large volumes of consumption data are generated, offering a promising avenue for detecting such anomalies using machine learning techniques. This paper presents a novel approach to electricity theft detection in smart grids, leveraging Artificial Neural Networks (ANN) and Deep Neural Networks (DNN). The proposed methodology involves the collection of consumption data, followed by preprocessing techniques such as data imputation, normalization, and feature extraction in both time and frequency domains. These features are then fed into ANN and DNN models, which classify the data as either normal or indicative of theft. To address challenges like class imbalance and missing data, techniques like synthetic data generation and interpolation are applied. Experimental results show the effectiveness of the proposed method in accurately detecting electricity theft while minimizing false positives, providing a robust and scalable solution for real-time monitoring and theft prevention in smart grids.

Index Terms - Electricity theft, smart grids, machine learning, deep neural networks, artificial neural networks, anomaly detection, advanced metering infrastructure, data pre-processing, feature extraction, class imbalance, synthetic data generation, real-time monitoring etc.

I. INTRODUCTION

1.1 Background

Electricity theft, a significant issue for utility providers globally, leads to annual losses amounting to billions of dollars. This illegal activity not only destabilizes the revenue streams of utility companies but also causes disruptions in the energy market, often resulting in higher prices for legitimate consumers and safety risks, including fire and electrocution hazards [1]. Traditional theft detection approaches, such as manual meter inspections and audits, are labor-intensive, costly, and inadequate in today's expansive and complex power grids [2]. The recent adoption of smart grid technology has provided an opportunity for more effective theft detection through advanced metering infrastructure (AMI), which captures high-resolution consumption data from consumers. With this data, utility companies can deploy automated and data-driven detection systems capable of identifying unusual consumption patterns that may indicate theft [3].

Smart grids employ machine learning (ML) and deep learning (DL) methods to analyze large amounts of real-time data generated by smart meters. These technologies allow for a more nuanced and adaptive approach to theft detection by recognizing subtle irregularities that traditional methods may overlook. In particular, the deployment of convolutional neural networks (CNNs) and ensemble models has enabled advanced anomaly detection methods, with CNNs capable of extracting features from time-series data and ensemble methods improving robustness and accuracy through model diversity [4,5]. However, applying these methods in real-time settings remains challenging due to the high computational demands and latency constraints associated with processing and analyzing massive volumes of data in smart grids.

1.2 Challenges in Existing Literature

Despite the advancements brought by machine learning and deep learning, existing literature highlights several challenges that limit the effectiveness of these methods in real-time electricity theft detection. One of the primary issues is data quality. Electricity consumption datasets often contain missing values and imbalanced classes, with theft cases representing a small fraction of the overall data [6]. Missing

values introduce bias in ML models, reducing prediction accuracy, while class imbalance often leads to models that perform well on non-theft data but fail to detect theft reliably. Techniques such as interpolation [7] and synthetic data generation methods, like SMOTE, have been employed to address these issues, but challenges remain in ensuring data integrity and generalizability [6].

Another challenge is the complexity of theft detection models. While deep learning architectures, such as CNNs and autoencoders, capture intricate patterns in consumption data, they require substantial computational power, which can hinder real-time processing. For example, Zheng et al. [4] demonstrated that wide and deep CNNs are effective for theft detection but may not be suitable for real-time applications due to high resource demands. Additionally, the variety of theft tactics—ranging from meter tampering to bypassing necessitates adaptable models that can detect diverse theft patterns, further increasing computational requirements. Ensemble learning, which combines multiple classifiers to enhance robustness, has been proposed as a solution, but its deployment in real-time environments remains limited due to latency and computational constraints [9,10].

Furthermore, the deployment of these methods in live, large-scale environments faces challenges related to latency and scalability. Real-time applications require fast processing to monitor high-frequency data streams effectively, but existing algorithms often struggle to meet these demands without sacrificing accuracy [6,11]. Edge computing and task offloading strategies have been proposed to address latency, yet they are in early stages of application in theft detection and require further research to validate their efficacy [6].

1.3 Motivation

The limitations of current detection methods underscore the need for more robust, real-time theft detection models capable of handling data quality issues and providing accurate, low-latency responses. Ensemble learning techniques, which have demonstrated success in other anomaly detection fields, offer a promising solution for electricity theft detection in smart grids. By leveraging diverse models, ensemble approaches can mitigate the impact of data noise and class imbalance, while utilizing advanced preprocessing methods to improve real-time application viability. This review is motivated by the need to explore the effectiveness of ensemble-based techniques in overcoming the limitations of existing models and enhancing theft detection in smart grid environments.

1.4 Objectives of the Paper

The primary objectives of this paper are as follows:

1. To review existing electricity theft detection techniques, focusing on ensemble-based approaches, and analyze their strengths and limitations.
2. To examine the challenges posed by data quality issues, such as missing values and class imbalance, and evaluate preprocessing methods for improving detection accuracy.
3. To provide a comprehensive assessment of ensemble models' potential to enhance theft detection performance in real-time applications within smart grids.
4. To identify open challenges and suggest future research directions to further improve detection models in terms of accuracy, latency, and scalability.

1.5 Contributions

This paper makes several key contributions:

1. A systematic review of current methods for electricity theft detection, highlighting the role of ensemble learning techniques in addressing model robustness and accuracy challenges.
2. An evaluation of data preprocessing methods, including interpolation and synthetic data generation, that address common data quality issues in electricity consumption datasets.
3. An analysis of the strengths and weaknesses of ensemble-based models in the context of real-time theft detection, particularly in their capacity to handle high-frequency smart grid data.
4. Recommendations for future research directions, emphasizing improvements in model scalability, adaptability, and the integration of edge computing solutions to support real-time deployment.

This paper is organized into five chapters. Chapter 1 introduces the problem of electricity theft, highlighting its significance and the need for effective detection methods in smart grids. Chapter 2 provides a comprehensive literature review, summarizing existing approaches and techniques used for electricity theft detection, including the challenges and advancements in the field. Chapter 3 delves into the application of machine learning techniques, particularly focusing on their role in identifying anomalous consumption patterns that could indicate theft. Chapter 4 presents the proposed methodology, outlining the steps from data collection and pre-processing to the training of Artificial Neural Networks (ANN) and Deep Neural Networks (DNN) for real-time theft detection. Finally, Chapter 5 concludes the paper by summarizing the findings, discussing the potential impact of the proposed approach, and suggesting future directions for research in this area

II. Literature Review

2.1 Electricity Theft Detection in Smart Grids

Electricity theft presents significant financial and operational challenges for utility providers. Traditional theft detection approaches, such as manual inspections and audits, have proven to be insufficient for addressing large-scale theft in modern power grids. The transition to smart grids has provided utility companies with a range of advanced monitoring tools, including smart meters and AMIs, which collect large volumes of consumer data. This data enables the development of data-driven theft detection approaches,

primarily based on machine learning (ML) and deep learning (DL) models, which analyze consumer usage patterns to detect anomalies indicative of theft [1, 2].

2.2 Machine Learning Approaches for Theft Detection

Several machine learning techniques have been explored to identify abnormal electricity consumption behaviors effectively. Chen et al. [3] conducted a comprehensive review of electricity consumption abnormality detection methods, highlighting the advantages and limitations of various ML techniques. Among these, supervised learning methods, such as decision trees, support vector machines (SVMs), and k-nearest neighbors (KNN), have been widely used but face challenges related to data quality, particularly in managing missing values and class imbalance [4].

To improve detection accuracy, convolutional neural networks (CNNs) have been applied to capture complex spatial and temporal patterns in electricity usage data [4, 5]. Zheng et al. [4] proposed wide and deep CNNs to address these complexities, demonstrating their effectiveness in detecting theft. However, deep learning models often require high computational resources, which can limit their deployment in real-time environments.

2.3 Data Quality Challenges

Data quality issues, such as missing values and class imbalance, are common in electricity consumption datasets. These issues adversely affect the performance of theft detection models, as missing data can introduce biases, and class imbalance often results in low sensitivity to theft instances. Ding et al. [7] proposed an interpolation method to address missing data by leveraging trends in the data. Additionally, synthetic data generation methods, such as the Synthetic Minority Over-sampling Technique (SMOTE), have been used to tackle class imbalance [5, 6].

A significant advancement in this area is the application of reinforcement learning for efficient data handling in intelligent transport and power systems [6]. Such methods enhance the data preprocessing pipeline by making real-time predictions feasible, improving the reliability of ML-based theft detection models.

2.4 Ensemble Learning for Improved Detection

Ensemble learning, which combines multiple classifiers to improve accuracy and robustness, has shown promising potential in electricity theft detection. Liao et al. [9] demonstrated that ensemble models using Euclidean and graph convolutional neural networks outperformed single-model approaches by leveraging diverse data perspectives. Similarly, Yan and Wen [14] applied extreme gradient boosting (XGBoost) in theft detection, showing that ensemble methods can effectively handle complex and high-dimensional data, improving detection accuracy.

Despite the potential of ensemble methods, their computational demands remain a concern, particularly for real-time detection scenarios in smart grids. Cui et al. [13] explored methods to mitigate these issues, proposing a two-step detection strategy using convolutional autoencoders and regression algorithms to enhance economic returns. Ensemble approaches have also been applied with success to other anomaly detection problems, including IoT data streams, further supporting their relevance for smart grid applications [16].

2.5 Addressing Deployment Challenges in Real-Time Settings

Real-time deployment of theft detection models requires efficient algorithms that minimize latency and can handle the high data inflow of smart grids. Researchers have proposed strategies, such as task offloading to edge computing, which help mitigate latency issues while maintaining detection accuracy [6]. Furthermore, the use of memory-augmented autoencoders has been suggested to improve anomaly detection in IoT time series data, offering potential for application in electricity theft detection by enhancing model memory and response time [16].

Cui et al. [17] proposed a strategy that combines convolutional autoencoders with economic analysis for theft detection. This approach optimizes for both detection accuracy and cost-efficiency, which is essential for large-scale deployment in utility companies. Pereira and Saraiva [18] focused on handling unbalanced data within theft detection, comparing various techniques and highlighting ensemble models' ability to maintain accuracy in low-theft instances, a key consideration for reliable real-time deployment.

2.6 Emerging Techniques and Future Directions

Recent advancements in neural network architectures, such as attention mechanisms, are beginning to gain traction in electricity theft detection. Finardi et al. [19] applied self-attention models to enhance anomaly detection capabilities by focusing on relevant portions of the data, which improves both detection accuracy and model interpretability. As the field progresses, incorporating advanced techniques like attention-based models and memory-augmented networks may further enhance real-time theft detection systems.

To summarize, the literature demonstrates a clear progression from traditional ML methods to more sophisticated ensemble and DL-based approaches, which better accommodate the unique challenges of electricity theft detection. Future research should continue to explore ensemble-based models and emerging deep learning architectures to address real-time deployment challenges in smart grids, emphasizing robustness, computational efficiency, and adaptability.

III. Machine Learning Techniques Electricity Theft Detection

Machine learning (ML) techniques are increasingly utilized in electricity theft detection to identify abnormal consumption patterns, which may indicate theft, tampering, or other fraudulent activities. Here's an overview of the most common ML techniques applied in this field:

3.1. Supervised Learning Techniques

Supervised learning techniques are widely used in electricity theft detection as they leverage labeled data (historical consumption data tagged as either normal or fraudulent) to train models. Key supervised learning methods include:

a. Decision Trees and Random Forests

- **Decision Trees** classify data by making decisions based on feature values, which is helpful in capturing patterns in consumption behavior.
- **Random Forests**, an ensemble of decision trees, improve accuracy and generalizability by aggregating results across multiple trees. These models are popular for theft detection due to their interpretability and robustness in handling different theft patterns.

b. Support Vector Machines (SVM)

- SVM is effective for binary classification problems and works by finding a hyperplane that best separates normal and fraudulent consumption behaviors. It is especially useful in detecting theft when the data is linearly separable but requires extensive tuning to handle high-dimensional data.

c. Logistic Regression

- Logistic regression is a probabilistic model that estimates the likelihood of an event (e.g., theft) occurring based on consumption features. Despite its simplicity, it provides a quick baseline model for detecting anomalous behavior in electricity consumption.

d. k-Nearest Neighbors (k-NN)

- This method identifies theft by comparing a consumer's data with its nearest neighbors. If a user's consumption pattern significantly deviates from its neighbors, it may be flagged as suspicious. However, k-NN is computationally intensive, especially on large datasets.

3.2. Unsupervised Learning Techniques

Unsupervised learning methods are beneficial when labeled data is limited, as they can identify abnormal patterns without requiring explicit labels.

a. Clustering (e.g., k-Means)

- Clustering methods like k-Means group users based on similar consumption behaviors. Users in clusters with abnormal patterns may be flagged for further investigation. Clustering is advantageous in detecting new or evolving theft patterns that may not be evident in labeled data.

b. Principal Component Analysis (PCA)

- PCA is a dimensionality reduction technique that identifies the most significant features for analysis, highlighting anomalies that deviate from regular consumption patterns. It's useful for pre-processing large datasets to improve computational efficiency and reveal hidden patterns.

c. Autoencoders

- Autoencoders, which are neural networks used for dimensionality reduction, learn to compress data into a lower-dimensional space and then reconstruct it. When applied to normal consumption data, they reconstruct it well, but theft cases are often reconstructed poorly, making them easy to identify.

3.3. Deep Learning Techniques

Deep learning techniques are increasingly popular for electricity theft detection due to their ability to handle complex patterns and large-scale datasets in smart grids.

a. Convolutional Neural Networks (CNNs)

- CNNs are well-suited for grid data with spatial or temporal dependencies. For theft detection, CNNs extract hierarchical features from consumption patterns, identifying even subtle abnormalities. CNN-based models can process time-series data or grid snapshots, enabling them to detect tampering or irregular usage more accurately.

b. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM)

- RNNs and their variants like LSTM networks are designed for sequential data, making them ideal for analyzing time-series consumption data. LSTMs capture long-term dependencies and detect patterns in energy usage that may indicate theft (e.g., sudden drops or spikes in consumption).

c. Generative Adversarial Networks (GANs)

- GANs can generate synthetic data similar to real consumption patterns, which is useful for addressing data imbalance in theft detection. By training on both real and synthetic data, GANs help create more robust detection models that perform well on minority classes, such as theft cases.

3. 4. Ensemble Learning Techniques

Ensemble techniques combine multiple models to enhance prediction accuracy and robustness, making them suitable for handling the complexity of electricity theft detection.

a. Random Forests and Gradient Boosting Machines (GBM)

- Random Forests and GBMs aggregate decisions from multiple trees to improve the predictive accuracy and reduce overfitting. Gradient Boosting techniques (e.g., XGBoost) iteratively correct errors from previous models, producing more refined and reliable theft detection models.

b. Extreme Gradient Boosting (XGBoost)

- XGBoost is an optimized implementation of gradient boosting that is efficient in handling large datasets and can model complex relationships in consumption data. It has become a popular choice for electricity theft detection due to its high accuracy and speed.

c. Stacking and Voting Classifiers

- Stacking combines the predictions of various base models (e.g., decision trees, logistic regression) and uses a meta-classifier to make the final prediction. Voting classifiers aggregate predictions from multiple models, making the final decision by majority or weighted vote. These ensemble techniques improve accuracy by capitalizing on the strengths of different models.

3.5. Hybrid Approaches

Hybrid models combine multiple ML techniques to leverage their respective strengths. For example, a common approach is combining clustering with supervised learning, where the clustering algorithm identifies groups with potential theft cases, and a classifier (e.g., SVM) is then used to refine the detection.

Example: CNN-RNN Hybrid Models

- Hybrid models that integrate CNNs and RNNs benefit from both spatial feature extraction and temporal sequence analysis. CNNs extract high-level features from consumption data, while RNNs analyze the sequence of these features over time to capture irregular consumption patterns that may indicate theft.

3.6. Anomaly Detection Techniques

Anomaly detection models focus on identifying deviations from established consumption norms without requiring labeled theft data.

a. Isolation Forest

- Isolation Forest isolates observations by randomly selecting features and splitting values. The fewer splits required to isolate a point, the more likely it is an anomaly, making it effective in identifying theft cases as outliers.

b. One-Class SVM

- This method is trained on normal consumption data and identifies theft as deviations from this normal class. One-Class SVMs are useful when only non-theft data is available for training but may be less accurate if theft cases are similar to regular consumption patterns.

c. Hidden Markov Models (HMM)

- HMMs are probabilistic models that can model sequences, making them suitable for time-series consumption data. By learning typical consumption sequences, HMMs detect anomalies when sequences deviate from normal behavior.

7. Reinforcement Learning

Reinforcement learning (RL) is an emerging area for electricity theft detection, where models learn optimal actions (e.g., detecting or ignoring certain consumption patterns) through trial and error, guided by reward mechanisms. RL-based models can improve detection by adapting to new types of theft patterns over time.

Each of these machine learning techniques has its strengths and limitations in detecting electricity theft, and their effectiveness depends on factors such as data availability, computational resources, and real-time processing requirements. Techniques such as ensemble learning and hybrid models show great promise, combining multiple ML approaches to build more accurate and robust theft detection systems. Integrating these models with edge computing and advanced data preprocessing techniques can further support real-time detection and enhance the scalability of these solutions in large-scale smart grids

IV. Proposed Method

The proposed method for electricity theft detection involves several stages, starting with the acquisition of data from smart meters and other sources in the smart grid. The input data typically includes electricity consumption patterns, meter readings, and contextual information like weather and location. The pre-processing stage addresses common issues such as missing data, noise, and class imbalance. Techniques like data imputation, synthetic data generation, and normalization are applied to ensure the dataset is clean and suitable for machine learning. Feature extraction follows, where both time-domain features (such as consumption mean, peak, and variability) and frequency-domain features (like harmonic components) are extracted. These features help capture patterns and anomalies that could indicate theft. The core of the method involves training machine learning models, specifically Artificial Neural Networks (ANN) and Deep Neural Networks (DNN), which are well-suited for detecting complex, non-linear patterns in large datasets. These models learn to classify consumption data as either normal or indicative of theft. Finally, the output consists of real-time predictions, with the model flagging unusual consumption behavior as a potential case of theft. Performance is evaluated using various metrics like accuracy, precision, recall, and F1-score to ensure the model detects theft effectively while minimizing false positives. The proposed approach offers a robust and scalable solution for real-time electricity theft detection in smart grids.

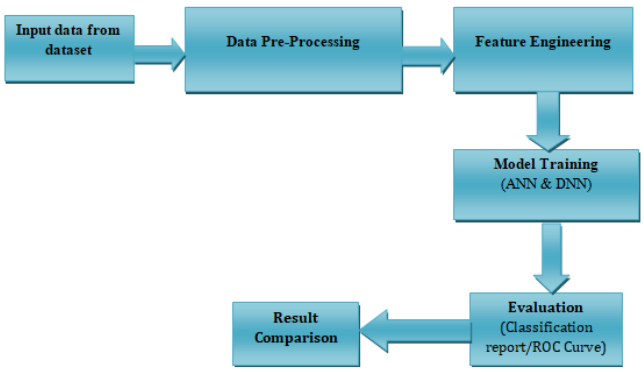

Figure 1: Proposed method Architecture

The process of using machine learning, specifically Artificial Neural Networks (ANN) and Deep Neural Networks (DNN), for electricity theft detection. The approach involves multiple stages, from data acquisition and preprocessing to feature extraction, model training, and output generation.

1. Input from Dataset

- **Data Acquisition**: The primary input for electricity theft detection models comes from the smart grid or Advanced Metering Infrastructure (AMI) system. This data typically includes:
  - **Consumption Data**: Hourly, daily, or monthly electricity usage per customer or meter.
  - **Metering Data**: Meter readings from the smart meters, which include consumption time series and voltage/current values.
  - **Historical Data**: Previous consumption data for customers, which helps to understand typical consumption behavior.
  - **Environmental/Contextual Data**: Factors such as weather, location, and demographic data can be integrated to enhance the model's ability to detect anomalies.
- **Data Characteristics**: The dataset will often contain time-series data that reflects the consumption pattern over time, including regular consumption, spikes, dips, or unusual activity which may suggest theft.

2. Pre-processing

- **Handling Missing Data**: Missing or incomplete data is common in real-world datasets. Several techniques can be applied to handle this:
  - **Imputation**: Missing values can be filled using techniques like mean imputation, interpolation (linear or cubic), or using more sophisticated methods like KNN imputation.
  - **Synthetic Data Generation**: For imbalanced classes (normal vs. theft cases), synthetic data (such as through SMOTE—Synthetic Minority Over-sampling Technique) can be used to augment the minority class (theft cases).
- **Data Normalization/Standardization**: Scaling features to a standard range (e.g., 0 to 1) or normal distribution is critical for neural networks. This helps prevent certain features from dominating the model's learning process, especially when the dataset includes features with different units or magnitudes.
- **Dealing with Class Imbalance**: Since theft cases are typically much rarer than normal consumption patterns, techniques like class weighting, oversampling the minority class (synthetic data), or undersampling the majority class may be applied to balance the dataset and improve the model's performance.
- **Noise Removal**: Any noise in the data (e.g., erroneous readings or outliers) should be filtered to improve the model's robustness and prevent overfitting.

3. Feature Extraction

- **Time Domain Features**: These features capture patterns within the raw time-series data:
  - **Mean Consumption**: Average consumption over a given period (e.g., daily or monthly).
  - **Peak Consumption**: Highest usage within a given time window.
  - **Consumption Variability**: The variance or standard deviation of consumption.
  - **Time-Series Trend**: Long-term consumption trends or seasonal effects.
- **Frequency Domain Features**: These features are derived using techniques like Fast Fourier Transform (FFT) to capture periodic patterns and frequencies within the time-series data:
  - **Frequency Components**: Identify dominant frequencies that might indicate unusual consumption behavior.
  - **Harmonics**: Anomalies in the harmonic spectrum can reveal tampering or other non-typical usage patterns.
- **Statistical Features**: Features such as skewness, kurtosis, and percentiles (e.g., 25th, 50th, and 75th percentiles) can provide a statistical summary of the data and highlight outliers or irregular consumption.
- **Behavioral Features**: Patterns of consumption based on customer behavior, such as sudden spikes or drops in consumption, can be significant indicators of theft. Temporal features like time of day, weekday/weekend, and holidays may also help identify abnormal patterns.
- **Contextual Features**: Environmental or demographic factors such as weather conditions, time-of-day, or neighborhood energy consumption patterns can be used to refine the detection model.

4. Model (ANN & DNN)

- **Artificial Neural Networks (ANN)**:
  - **Architecture**: A basic ANN for electricity theft detection typically consists of an input layer, one or more hidden layers with activation functions like ReLU or Sigmoid, and an output layer (binary classification: theft or no theft).
  - **Training**: The model is trained using backpropagation and gradient descent techniques. The model learns to map the extracted features to the target labels (normal or theft).
  - **Activation Function**: ReLU or Tanh is commonly used in hidden layers for better

learning. The output layer uses a Sigmoid activation function for binary classification.

- **Deep Neural Networks (DNN)**:
  - **Architecture**: DNNs are a more advanced version of ANNs with deeper networks (more hidden layers), which allow for better modeling of complex, non-linear relationships in the data. These are particularly useful when the dataset is large and contains complex consumption patterns.
  - **Training**: DNNs are trained with more advanced techniques such as Adam or RMSprop optimizers to deal with issues like vanishing gradients and overfitting. Dropout regularization or batch normalization may be applied to improve generalization.
  - **Backpropagation**: DNNs use multi-layer backpropagation, which adjusts the weights in a deeper network to minimize the error (loss function) across multiple layers.
- **Model Hyperparameter Tuning**: To achieve optimal results, hyperparameters such as learning rate, number of layers, number of neurons per layer, batch size, and epochs are tuned using techniques like grid search or random search.
- **Loss Function**: The binary cross-entropy loss function is typically used for classification problems to measure how well the predicted labels match the true labels (normal or theft).

5. Output

- **Predictions**: After training, the model outputs a prediction for each input sample. This prediction will be either:
  - **Normal Consumption**: When the model classifies the consumption as regular.
  - **Theft or Anomaly**: When the model detects consumption patterns that deviate from the norm and classify them as abnormal, which may indicate theft or tampering.
- **Thresholding**: Depending on the application, a threshold may be applied to the model's output probabilities to make the final classification. For instance, if the output probability for theft exceeds a certain threshold (e.g., 0.5), the model will classify the observation as a theft case.
- **Output Analysis**: The results can be analyzed in terms of:
  - **Confusion Matrix**: To measure the performance, including precision, recall, F1-score, and accuracy.
  - **False Positives and False Negatives**: The impact of misclassifying a normal consumption case as theft (false positive) or missing an actual theft case (false negative) needs to be evaluated.

6. Model Evaluation

- **Cross-validation**: Cross-validation techniques (e.g., k-fold cross-validation) are used to ensure the model generalizes well across different subsets of data.
- **Performance Metrics**: Evaluate performance using metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix. Special attention is given to the recall metric to ensure that theft cases are detected as accurately as possible.
- **Testing**: After training and tuning, the model is evaluated on a separate test dataset to verify its ability to detect theft in unseen data.

.

V. Conclusion

Electricity theft remains a persistent and costly challenge for utility providers globally, undermining revenue and jeopardizing energy security. With the advent of smart grids and advanced metering infrastructure (AMI), new opportunities have emerged to combat electricity theft using sophisticated, data-driven techniques. This review has examined current approaches to real-time electricity theft detection, focusing on the application of ensemble learning models and their potential to enhance detection accuracy and robustness. The review highlighted that traditional machine learning and deep learning models have shown promise but are often hindered by challenges such as data quality issues, computational demands, and the need for real-time processing. Ensemble learning, which combines multiple models to boost predictive performance, has demonstrated potential in addressing these limitations, particularly in terms of handling missing values, class imbalance, and complex consumption patterns. Studies also show that ensemble methods can offer improved resilience to noise and variability in smart grid data, which is critical for accurate theft detection.

References

1. M. Xing, W. Ding, H. Li, T. Zhang, A power transformer fault prediction method through temporal convolutional network on dissolved gas chromatography data. Secur. Commun. Netw. 2022, 66 (2022)
2. S.S.S.R. Depuru, L. Wang, V. Devabhaktuni, Electricity theft: overview, issues, prevention and a smart meter based approach to control theft. Energy Policy 39(2), 1007–1015 (2011)
3. Q. Chen, K. Zheng, C. Kang, F. Huangfu, Detection methods of abnormal electricity consumption behaviors: review and prospect. Autom. Electr. Power Syst. 42(17), 189–199 (2018)
4. Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, Y. Zhou, Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. IEEE Trans. Ind. Inform. 14(4), 1606–1615 (2017)

5. A. Arif, T.A. Alghamdi, Z.A. Khan, N. Javaid, Towards efficient energy utilization using big data analytics in smart cities for electricity theft detection. Big Data Res. 27, 100285 (2022)
6. H. Gao, W. Huang, T. Liu, Y. Yin, Y. Li, Ppo2: location privacy-oriented task offloading to edge computing using reinforcement learning for intelligent autonomous transport systems. IEEE Trans. Intell. Transp. Syst. 6, 66 (2022)
7. W. Ding, Z. Wang, Y. Xia, K. Ma, An efficient interpolation method through trends prediction in smart power grid. Intell. Mob. Serv. Comput. 66, 79–92 (2021)
8. M.I. Ibrahem, M. Mahmoud, F. Alsolami, W. Alasmary, A.-G. Abdullah, X. Shen, Electricity theft detection for changeand- transmit advanced metering infrastructure. IEEE Internet Things J. 9, 25565 (2022)
9. W. Liao, Z. Yang, K. Liu, B. Zhang, X. Chen, R. Song, Electricity theft detection using Euclidean and graph convolutional neural networks. IEEE Trans. Power Syst. 6, 66 (2022)
10. P. Jokar, N. Arianpoo, V.C. Leung, Electricity theft detection in ami using customers' consumption patterns. IEEE Trans. Smart Grid 7(1), 216–226 (2015)
11. X. Kong, X. Zhao, C. Liu, Q. Li, D. Dong, Y. Li, Electricity theft detection in low-voltage stations based on similarity measure and dt-ksvm. Int. J. Electr. Power Energy Syst. 125, 106544 (2021)
12. Y. Himeur, A. Alsalemi, F. Bensaali, A. Amira, Smart power consumption abnormality detection in buildings using micromoments and improved k-nearest neighbors. Int. J. Intell. Syst. 36(6), 2865–2894 (2021)
13. L. Cui, L. Guo, L. Gao, B. Cai, Y. Qu, Y. Zhou, S. Yu, A covert electricity-theft cyber-attack against machine learningbased detection models. IEEE Trans. Ind. Inform. 6, 66 (2021)
14. Z. Yan, H. Wen, Electricity theft detection base on extreme gradient boosting in ami. IEEE Trans. Instrum. Meas. 70, 1–9 (2021)
15. Y. Huang, Q. Xu, Electricity theft detection based on stacked sparse denoising autoencoder. Int. J. Electr. Power Energy Syst. 125, 106448 (2021)
16. H. Gao, B. Qiu, R.J.D. Barroso, W. Hussain, Y. Xu, X. Wang, Tsmae: a novel anomaly detection approach for internet of things time series data using memory-augmented autoencoder. IEEE Trans. Netw. Sci. Eng. 6, 66 (2022)
17. X. Cui, S. Liu, Z. Lin, J. Ma, F. Wen, Y. Ding, L. Yang, W. Guo, X. Feng, Two-step electricity theft detection strategy considering economic return based on convolutional autoencoder and improved regression algorithm. IEEE Trans. Power Syst. 37(3), 2346–2359 (2021)
18. J. Pereira, F. Saraiva, Convolutional neural network applied to detect electricity theft: a comparative study on unbalanced data handling techniques. Int. J. Electr. Power Energy Syst. 131, 107085 (2021)
19. S. Sharma, M. Saraswat, A.K. Dubey, Fake news detection on twitter. Int. J. Web Inf. Syst. 6, 66 (2022)
20. P. Finardi, I. Campiotti, G. Plensack, R.D. de Souza, R. Nogueira, G. Pinheiro, R. Lotufo, Electricity theft detection with selfattention. arXiv preprint arXiv: 2002. 06219 (2020).