International Journal of Research and Innovation in Multidisciplinary (www.ijrim.com)
||Volume 3, Issue 8, August 2026 || Impact Factor: 5.2 |

# Convolutional Neural Network based Deep Learning Model for Smart Meter Electricity Theft Detection

[1]Abhijeet Kumar, [2]Mr. Balram Yadav
*[1]M.Tech Scholar, School of Electrical and Electronics Engineering, SCOPE Global Skills University, Bhopal, India,*
*[2]Associate Professor & HOD, School of Electrical and Electronics Engineering, SCOPE Global Skills University, Bhopal, India*

***Abstract*— Electricity theft is a major challenge for power utilities because fraudulent consumption can increase non-technical losses, reduce revenue, and affect the accuracy of electrical energy accounting in smart-grid environments. The availability of smart meters and Advanced Metering Infrastructure (AMI) provides detailed time-series electricity consumption data that can be utilized for automated identification of abnormal consumption patterns. In this research, a Convolutional Neural Network (CNN)-based deep learning model is developed for smart meter electricity theft detection. The proposed approach utilizes historical smart-meter electricity consumption data and applies data preprocessing, normalization, and relevant feature preparation before model training. The CNN model is designed to extract important local patterns and variations from electricity consumption sequences and classify consumers into normal and potentially fraudulent categories. The developed model is evaluated using standard classification performance measures, including accuracy, precision, recall, F1-score, and confusion matrix analysis. The framework aims to improve the identification of complex electricity theft patterns while reducing dependence on manual inspection and predefined detection rules. The proposed CNN-based approach can provide an efficient and scalable solution for automated electricity theft detection and support power utilities in reducing non-technical losses, improving energy monitoring, and strengthening the reliability and security of modern smart-grid systems.**

***Keywords: Smart Meter, Electricity Theft, CNN, Deep Learning, Smart Grid.***

## I. INTRODUCTION

Smart meter electricity theft detection has become an important research area in modern electrical power systems because electricity theft contributes significantly to non-technical losses and affects the financial and operational performance of power utilities. Unauthorized electricity consumption can occur through meter tampering, meter bypassing, illegal connections, or manipulation of metering equipment, resulting in a difference between the actual energy supplied and the energy recorded by the meter. Such discrepancies can affect electricity billing, energy accounting, distribution-system planning, and utility revenue. The development of smart meters has provided an opportunity to continuously monitor electricity consumption and identify abnormal consumption behavior using recorded electrical load data [1].

Smart meters are an important component of Advanced Metering Infrastructure (AMI), enabling utilities to obtain electricity consumption measurements at regular time intervals. These measurements provide detailed information about the electrical load profile of individual consumers and can be analyzed to identify unusual changes in consumption. Unlike conventional meter inspection methods, smart-meter-based detection can support automated monitoring of a large number of consumers simultaneously. This capability is particularly useful for modern distribution systems where the number of connected consumers and the volume of electricity consumption data are continuously increasing [2].

Electricity theft detection is challenging because fraudulent consumption patterns may be similar to legitimate variations in household electrical demand. Residential electricity consumption changes according to appliance operation, occupancy, time of day, weather conditions, seasonal requirements, and consumer behavior. For example, a sudden decrease in electricity consumption may indicate unauthorized usage, but it may also result from a consumer being away from the premises. Similarly, high electricity consumption may be completely legitimate when air conditioners, water heaters, pumps, or other high-power appliances are operating. Therefore, reliable detection requires the analysis of consumption behavior over a suitable time period rather than relying only on individual meter readings [3].

Traditional electricity theft detection generally depends on physical inspection, predefined rules, threshold-based methods, and statistical comparison of electricity consumption. Physical inspection can identify meter tampering but requires substantial manpower, time, and operational cost. Rule-based methods are relatively simple to implement but are usually designed for known theft conditions and may not perform effectively when new forms of electricity theft appear. Statistical techniques can identify deviations from expected consumption, but their performance may be affected by the nonlinear and time-dependent characteristics of residential load profiles. These limitations have encouraged the use of intelligent data-driven approaches for automated smart-meter electricity theft detection [4].

Machine learning techniques provide an alternative approach by learning the relationship between historical electricity consumption patterns and known normal or fraudulent conditions. Classification algorithms can analyze smart-meter measurements and assign consumption records to legitimate or suspicious categories. Such approaches can reduce dependence on manual inspection and enable utilities to prioritize consumers requiring further investigation. However, smart-meter electricity consumption is fundamentally a time-series phenomenon, and important information about fraudulent behavior may be distributed across multiple consecutive measurements. Therefore, detection models need to capture meaningful patterns from sequential electrical consumption data rather than considering individual observations independently [5].

The increasing availability of large-scale smart-meter datasets has further encouraged the application of deep learning techniques to electricity theft detection. Deep learning models can automatically learn complex and nonlinear representations from electricity consumption data and reduce the requirement for extensive manual feature engineering. This capability is particularly useful when theft patterns are difficult to describe using predefined mathematical rules. Deep learning can analyze variations in active power and energy consumption and identify patterns that differ from normal consumer behavior. Consequently, deep learning-based approaches have become an important direction for automated electricity theft detection in smart-grid environments [6].

Convolutional Neural Networks (CNNs) are particularly promising for smart-meter electricity theft detection because their convolution operation can extract meaningful local patterns from structured electricity consumption data. A sequence of smart-meter measurements can be represented in a suitable input format so that the CNN can identify local changes, fluctuations, and characteristic consumption patterns. Convolutional layers can progressively learn important features from the input electrical load profile, while subsequent classification layers can use these features to distinguish between normal and potentially fraudulent electricity consumption. This capability makes CNN-based approaches suitable for analyzing complex smart-meter load patterns [7].

The preprocessing of smart-meter data is another important component of electricity theft detection. Real-world smart-meter datasets may contain missing observations, measurement noise, abnormal readings, and inconsistent data. Furthermore, electricity theft datasets commonly suffer from class imbalance because legitimate consumption records are generally much more numerous than confirmed theft cases. If these issues are not appropriately addressed, a detection model may become biased toward normal consumption and fail to identify fraudulent cases. Therefore, data cleaning, normalization, feature preparation, and suitable data-balancing techniques are important for improving the reliability of smart-meter electricity theft detection models [8].

Another major challenge is maintaining a low false-positive rate while detecting a high proportion of actual theft cases[9]. A false positive occurs when a legitimate electricity consumer is incorrectly classified as a potential electricity thief. Such incorrect classification can lead to unnecessary field inspections, increased operational expenditure, and inconvenience to consumers. At the same time, failure to detect actual theft can result in continued financial losses for utilities. Therefore, an effective smart-meter electricity theft detection model must provide an appropriate balance between detection capability, precision, recall, and false-positive reduction. This requirement becomes increasingly important when automated detection systems are deployed across large distribution networks [10].

## II. METHODOLOGY

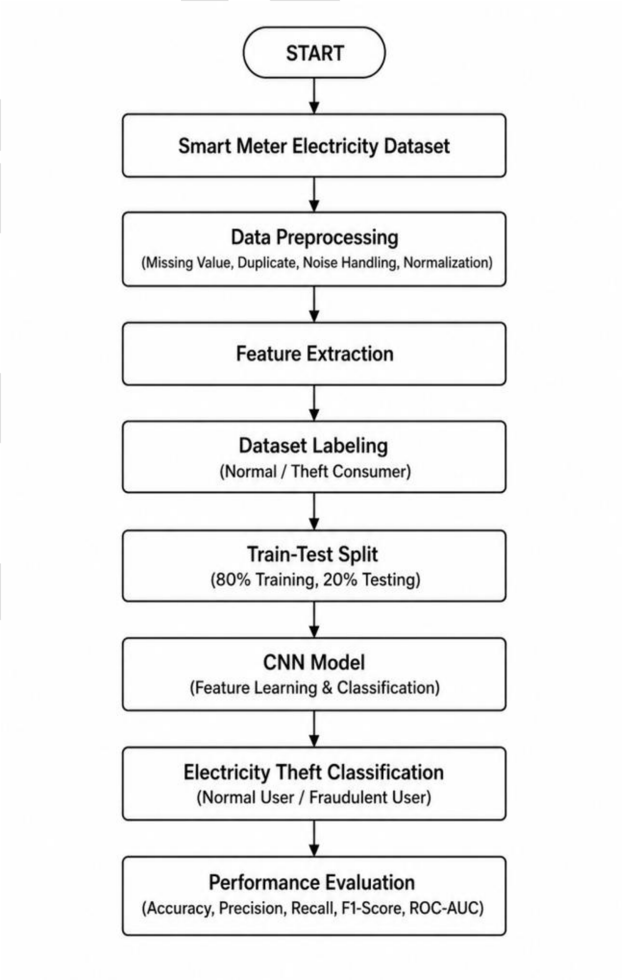


Figure 1: Flow chart

The proposed framework is designed to detect electricity theft from smart-meter electricity consumption data using a Convolutional Neural Network (CNN). The complete process starts with the collection of smart-meter data, followed by preprocessing, feature extraction, CNN-based learning, classification of consumers, and performance evaluation.

**Step 1: Smart Meter Electricity Dataset**

The first step is to obtain electricity consumption data recorded by smart meters. The dataset contains historical electrical consumption measurements of different consumers. These measurements represent the electricity usage pattern over different time intervals and provide the basic information required to identify normal and potentially fraudulent consumption.

The input data can be represented as:

$$X = \{x1, x2, x3, ..., xn\}$$

where x represents the recorded electricity consumption value and n represents the number of observations.

**Step 2: Data Preprocessing**

The collected smart-meter data may contain missing values, duplicate records, measurement noise, and values with different numerical ranges. Therefore, preprocessing is performed before applying the CNN model.

The main operations include:

- Missing-value handling
- Duplicate removal
- Noise removal
- Data normalization

Normalization can be performed using the Min-Max method:

$$X' = (X - Xmin) / (Xmax - Xmin)$$

where X is the original value, Xmin and Xmax are the minimum and maximum values, and **X'** is the normalized value.

This step improves data consistency and allows the CNN model to process the electrical consumption data more effectively.

**Step 3: Feature Extraction**

After preprocessing, useful characteristics are extracted from the smart-meter electricity data. The objective is to identify consumption patterns that can help distinguish normal electricity usage from suspicious or fraudulent usage. Depending on the available dataset, features may Electricity consumption, Active power, Reactive power, Voltage, Current, Time-based consumption patterns. The extracted features are then provided as input to the CNN model.

**Step 4: Dataset Labeling**

The processed data are classified into two basic categories:

**Normal Consumer → 0**

**Theft/Fraudulent Consumer → 1**

The labels provide the CNN model with the information required for supervised learning. During training, the model learns the relationship between the electrical consumption features and their corresponding class labels.

**Step 5: Train-Test Data Split**

The labeled dataset is divided into training and testing subsets. In the proposed framework, 80% of the data are used for training and 20% for testing. The training data are used to teach the CNN model the characteristics of normal and fraudulent electricity consumption. The testing data are kept separate and are used later to determine how well the trained model performs on previously unseen smart-meter observations.

The division can be represented as:

$$D = Dtrain + Dtest$$

where Dtrain represents the training dataset and Dtest represents the testing dataset.

**Step 6: CNN Model**

The CNN is the main deep learning component of the proposed framework. The prepared smart-meter data are supplied to the CNN input layer. The convolutional layers then identify important local patterns and variations within the electricity consumption data.

A simplified convolution operation can be expressed as:

$$C(i) = \Sigma X(i + k) \times W(k) + b$$

where X represents the input electricity consumption data, W represents the convolution filter, b is the bias, and C(i) represents the extracted feature.

The CNN architecture in the flowchart performs feature learning followed by classification.

**Step 7: Electricity Theft Classification**

The features learned by the CNN are passed to the classification stage. The final output determines whether the analyzed electricity consumption pattern belongs to a normal consumer or a potentially fraudulent consumer.

For binary classification, the sigmoid function can be used:

$$P = 1 / (1 + e^{(-z)})$$

where z represents the output of the final neural-network layer and P represents the probability of the consumer belonging to the theft class.

For example, a probability closer to 0 indicates a normal-consumption class, while a probability closer to 1 indicates a potential electricity-theft class.

**Step 8: Performance Evaluation**

Finally, the performance of the CNN-based electricity theft detection model is evaluated using classification metrics. The main metrics shown in the framework are:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

It represents the overall percentage of correctly classified consumers.

Precision = TP / (TP + FP)

It indicates how many consumers identified as theft cases are actually theft cases.

Recall = TP / (TP + FN)

It measures the ability of the model to correctly identify actual electricity-theft cases.

F1 = 2 × (Precision × Recall) / (Precision + Recall)

It provides a combined measure of precision and recall.

ROC-AUC is also used to evaluate the model's ability to distinguish between normal and fraudulent electricity consumption across different classification thresholds.

## III. Simulation Results

The simulation results are performed using the python software.

| | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|
| 1 | R1-PM1:V | R1-PA2:VH | R1-PM2:V | R1-PA3:VH | R1-PM3:V | R1-PA4:IH | R1-PM4:I |
| 2 | 127673.1 | -49.5723 | 127648 | -169.578 | 127723.2 | 65.68961 | 605.911 |
| 3 | 130280.7 | -46.3007 | 130255.6 | -166.278 | 130355.9 | 71.83172 | 483.5935 |
| 4 | 130305.8 | -46.2549 | 130280.7 | -166.232 | 130381 | 71.8088 | 483.5935 |
| 5 | 130581.6 | -45.8996 | 130556.5 | -165.883 | 130656.8 | 72.15258 | 482.8611 |
| 6 | 131083.1 | -45.4241 | 131058 | -165.424 | 131158.3 | 72.1182 | 484.5091 |
| 7 | 131058 | -45.4413 | 131032.9 | -165.442 | 131133.2 | 71.78015 | 486.5233 |
| 8 | 131007.8 | -45.4585 | 131007.8 | -165.453 | 131083.1 | 71.67129 | 487.0726 |
| 9 | 130982.8 | -45.4757 | 130957.7 | -165.476 | 131058 | 71.52805 | 487.9882 |
| 10 | 130932.6 | -45.533 | 130932.6 | -165.522 | 131007.8 | 71.22438 | 489.453 |
| 11 | 130857.4 | -45.7163 | 130857.4 | -165.705 | 130932.6 | 70.60559 | 491.6504 |
| 12 | 130832.3 | -45.8939 | 130832.3 | -165.894 | 130932.6 | 70.00971 | 494.2139 |
| 13 | 130832.3 | -46.0887 | 130807.2 | -166.095 | 130907.5 | 69.56281 | 497.693 |
| 14 | 130807.2 | -46.2893 | 130807.2 | -166.272 | 130907.5 | 69.12163 | 498.7916 |
| 15 | 130832.3 | -46.3122 | 130807.2 | -166.29 | 130907.5 | 69.09298 | 498.9748 |
| 16 | 130832.3 | -46.3408 | 130807.2 | -166.341 | 130907.5 | 69.05287 | 498.9748 |
| 17 | 130832.3 | -46.5815 | 130782.2 | -166.57 | 130907.5 | 68.66899 | 500.2565 |
| 18 | 130832.3 | -46.782 | 130807.2 | -166.782 | 130907.5 | 68.36532 | 500.989 |
| 19 | 130832.3 | -46.7935 | 130782.2 | -166.782 | 130907.5 | 68.38824 | 500.8059 |
| 20 | 130832.3 | -46.8737 | 130807.2 | -166.885 | 130882.5 | 68.23927 | 500.989 |
| 21 | 130832.3 | -46.8909 | 130807.2 | -166.908 | 130907.5 | 68.21063 | 501.1721 |
| 22 | 133289.5 | -42.4906 | 133264.4 | -162.508 | 133364.7 | 78.22593 | 281.4401 |

Figure 2: Dataset

The figure 2 is showing the dataset, which is taken from the Kaggle machine learning website.

| Name | Type | Size |
|---|---|---|
| i | int | 1 |
| inp | engine.keras_tensor.KerasTensor | 1 |
| label_encoder | preprocessing._label.LabelEncoder | 1 |
| nor | Array of int64 | (1100,) |
| pca | decomposition._pca.PCA | 1 |
| pool | engine.keras_tensor.KerasTensor | 1 |
| Predictedval | Array of int32 | (50,) |
| Predictedval1 | Array of int32 | (50,) |
| tpr | Array of float64 | (3,) |
| x | DataFrame | (4966, 128) |
| x_pca | Array of float64 | (4966, 120) |
| x_test | Array of float64 | (1242, 120) |
| x_train | Array of float64 | (3724, 120) |
| x_train11 | Array of float64 | (3724, 120, 1) |
| y | Series | (4966,) |
| y_pred | Array of int32 | (1242,) |
| y_test | Series | (1242,) |
| y_train | Series | (3724,) |

Figure 3: Variable explorer

Figure 3 is showing variable explorer of the dataset. The Variable Explorer displays all variables generated during the implementation of the proposed CNN-based electricity theft detection model.

cm - NumPy object array

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 20 | 0 | 0 | 0 |
| 1 | 1 | 20 | 0 | 0 |
| 2 | 1 | 0 | 0 | 0 |
| 3 | 1 | 0 | 0 | 0 |
| 4 | 1 | 0 | 0 | 0 |
| 5 | 1 | 0 | 0 | 0 |
| 6 | 1 | 0 | 0 | 0 |
| 7 | 1 | 0 | 0 | 0 |
| 8 | 1 | 0 | 0 | 0 |

Format Resize Background color

Save and Close Close

Figure 4: Confusion matrix

The figure 4 represents the Confusion Matrix which is a performance evaluation tool used to compare the actual class labels (Y_test) with the predicted class labels (Y_pred) generated by the CNN model.

Table 1: Result Comparison

| Sr. No. | Parameter | Previous Work [1] | Previous Work [2] | Proposed Work |
|---|---|---|---|---|
| 1 | Method | Transformer-DW-MHC | Transformer-SW-MHC | CNN |
| 2 | Precision (%) | NA | NA | 95.23 |
| 3 | Recall (%) | NA | NA | 99.00 |
| 4 | F measure (%) | 81.82 | 75.25 | 97.56 |
| 5 | Accuracy (%) | 93.59 | 91.26 | 97.56 |
| 6 | Classification error (%) | 6.41 | 8.74 | 2.44 |

## IV. CONCLUSION

The study successfully demonstrates the effectiveness of a Convolutional Neural Network (CNN)-based deep learning model for smart meter electricity theft detection in modern smart-grid environments. The proposed framework uses the Power System Attack Dataset and applies appropriate preprocessing to improve data quality before model training. The CNN automatically extracts significant features from power-system operational data through convolution, activation, and pooling operations, reducing the dependence on manual feature engineering. The developed model achieved 95.23% Precision, 99.00% Recall, 97.56% F1-Score, and 97.56% Accuracy, with only 2.44% Classification Error. Comparative results also demonstrate better overall performance than the Transformer-DW-MHC and Transformer-SW-MHC models. The high recall indicates strong capability in identifying actual attack events while minimizing missed detections. Overall, the proposed CNN framework provides an accurate, efficient, and scalable solution for electricity theft detection and can contribute to reducing non-technical losses, strengthening smart-grid security, and improving reliable power-system operation.

## REFERENCES

[1] X. Chen, C. Huang, Y. Zhang and H. Wang, "LoadGuard: An Adaptive Deep Learning Model for Smart Meter Electricity Theft Detection," 2025 IEEE 23rd International Conference on Industrial Informatics (INDIN), Kunming, China, 2025, pp. 1-6, doi: 10.1109/INDIN64977.2025.11279201.

[2] H. Lin, G. Zhang and Z. Sun, "A Hybrid Machine Learning and Deep Learning Framework for Effective Electricity Theft Detection in Smart Grids," 2025 5th Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), Shenyang, China, 2025, pp. 589-595, doi: 10.1109/ACCTCS66275.2025.00108.

[3] X. Chen, C. Huang, Y. Zhang, and H. Wang, "Smart energy guardian: A hybrid deep learning model for detecting fraudulent pv generation," in 2024 IEEE International Smart Cities Conference (ISC2), 2024, pp. 1–6.

[4] F. An, B. Zhao, B. Cui, Y. Chen, L. Qu, Z. Yu, R. Zeng, "Selective Virtual Synthetic Vector Embedding for Full-Range Current Harmonic Suppression of the DC Collector," IEEE Trans. Power Electronics., vol. 38, no. 2, pp. 2577–2588, Feb., 2023.

[5] Wu, H. Shi, S. Fu, Y. Luo, and M. Xu, "p2detect: Electricity theft detection with privacy preservation for both data and model in smart grid," IEEE Transactions on Smart Grid, vol. 14, no. 3, pp. 2301–2312, 2023.

[6] S. Zidi, A. Mihoub, S. Mian Qaisar, M. Krichen, and Q. Abu Al-Haija, "Theft detection dataset for benchmarking and machine learning based classification in a smart grid environment," Journal of King Saud University - Computer and Information Sciences, vol. 35, no. 1, pp. 13–25, 2023.

[7] Pamir, N. Javaid, S. Javaid, M. Asif, M. U. Javed, A. S. Yahaya, and S. Aslam, "Synthetic theft attacks and long short term memory-based preprocessing for electricity theft detection using gated recurrent unit," Energies, vol. 15, no. 8, 2022, p. 2778.

[8] An, F., Zhao, B., Cui, B., Ma, Y., Zhang, X., Tang, X., & Dong, L. (2022). Asymmetric Topology Design and Quasi-Zero-Loss Switching Composite Modulation for IGCT-Based High-Capacity DC Transformer. IEEE Transactions on Power Electronics, 38(4), 4745–4759.

[9] Pamir, N. Javaid, A. Almogren, M. Adil, M. U. Javed, and M. Zuair, "RFE based feature selection and KNNOR based data balancing for electricity theft detection using BiLSTM-Logit Boost stacking ensemble model," IEEE Access, vol. 10, 2022, pp. 112948–112963.

[10] D. Gu, Y. Gao, K. Chen, J. Shi, Y. Li, and Y. Cao, "Electricity theft detection in AMI with low false positive rate based on deep learning and evolutionary algorithm," IEEE Transactions on Power Systems, vol. 37, no. 6, 2022, pp. 4568–4578.