IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 37, No. 6, December, 2023, pp. 1557-1564
Journal homepage: http://iieta.org/journals/ria

Models for Detecting Electricity Thieves Using 1D and 2D Convolutional Neural Networks

Mohamed Nadjib Meadi*, Ferial Ouamane, Abdelhamid Djeffal

Department of Computer Science, LESIA Laboratory, University of Mohamed Khider, Biskra 07000, Algeria

Corresponding Author Email: mohamed_nadjib.meadi@univ-biskra.dz

https://doi.org/10.18280/ria.370620

Received: 9 August 2023
Revised: 17 September 2023
Accepted: 24 October 2023
Available online: 27 December 2023

Keywords:
deep learning, convolutional neural networks, electricity fraud, non-technical losses, smart grids

ABSTRACT

Smart grid systems are vulnerable to electricity theft, which endangers operational safety, sustainable development, and income integrity. This paper explores the use of one-dimensional (1D) and two-dimensional (2D) convolutional neural networks (CNNs) for efficient detection of electricity theft. One notable feature of 1D CNNs is their ability to extract patterns from sequential data, but 2D CNNs are better at handling pictorial data. An inventive method is presented to tackle the problem of missing values in the dataset, improving the performance of the models that are in use. The results show that the performance of electricity theft detection systems is greatly improved by these improved models.

1. INTRODUCTION

Electricity stands as the predominant energy source on our planet. Within this context, electricity companies provide a range of services encompassing energy generation, transmission, distribution, and sales, catering to consumers, businesses, and industries alike. The operations of these entities are closely monitored by various federal agencies. Electricity theft is a prevalent issue that contributes to energy losses. It involves the unauthorized installation of devices or methods to bypass the meters responsible for measuring energy consumption. This illegal practice is carried out with the intention of reducing the recorded consumption amount or evading billing for the energy used. Typically, electricity theft involves the installation of hidden systems or mechanisms that allow consumers to bypass the meter without detection. This form of illicit activity has detrimental effects on power grids, leading to a decline in power supply quality and a decrease in operating profits.

Electricity theft detection is a critical issue due to its widespread prevalence and significant consequences. It is driven by factors such as high energy costs, economic inequality, and weak law enforcement. The impacts of electricity theft are substantial, including financial losses for utility companies, compromised service quality for legitimate consumers, reduced government revenue, system inefficiencies, and safety risks. Effectively addressing electricity theft requires advanced technologies, collaboration among stakeholders, and public awareness campaigns. Implementing robust detection systems and comprehensive strategies is crucial to protect revenue, ensure reliable service delivery, promote fairness, and maintain a sustainable energy distribution system.

By combating electricity theft, utility companies can safeguard their financial viability and invest in infrastructure improvements. Governments can generate sufficient tax revenue, implement energy sector reforms, and provide reliable electricity to their citizens. Moreover, addressing theft promotes a culture of compliance, fairness, and responsible energy consumption, contributing to energy efficiency and environmental sustainability. Overall, tackling electricity theft is essential for the economic, social, and environmental well-being of communities and the stability of energy systems.

Furthermore, the smart grid, representing an advanced iteration of the traditional power grid, presents an opportunity to revolutionize the energy sector and usher in a new era of enhanced dependability. The smart grid incorporates modern technologies and communication systems to enable more efficient and intelligent management of electricity generation, distribution, and consumption. By integrating real-time data monitoring, automation, and advanced analytics, the smart grid facilitates improved power grid stability, reliability, and resilience. It enables better load management, fault detection, and self-healing capabilities, thereby minimizing disruptions and maximizing the overall efficiency of the energy system. The smart grid's ability to optimize energy usage, integrate renewable energy sources, and empower consumers with information and control has the potential to transform the energy sector and meet the increasing demand for reliable and sustainable electricity supply.

Moreover, smart grids, an evolved iteration of conventional grids, present an opportunity to usher in a fresh era characterized by enhanced reliability, accessibility, and efficiency, thereby contributing to economic and environmental prosperity [1]. The smart grid represents a digital innovation that facilitates bidirectional communication between utilities and consumers. Its "smart" attribute stems from sensor networks integrated into transmission lines. The

two-way interactivity inherent in smart grids allows for automated rerouting in response to equipment malfunctions or breakdowns, thereby bolstering the resilience and readiness of electrical energy systems during emergencies.

Convolutional neural networks (CNNs) are extensively employed in contemporary artificial intelligence applications, specifically for image and audio data processing. A standard CNN architecture comprises convolutional layers, followed by a pooling layer and a fully connected layer. These characteristics remain consistent across various dimensions. One-dimensional CNNs (1D CNNs) excel in signal analysis, particularly for fixed-length signals, as they are adept at detecting basic patterns that contribute to the creation of more complex patterns in subsequent layers. When working with fixed-length segments of a dataset where the positional information of features within the segments is not essential, 1D CNNs demonstrate high efficacy. They are commonly utilized in audio signal analysis and find applicability in certain natural language processing.

The choice of dimensions (1D, 2D, or 3D) for a CNN depends on the specific problem at hand. For example, 1D CNNs are typically employed for audio signals, 2D CNNs for images, and 3D CNNs for movies. Despite dimensional disparities, CNNs share similar properties and approaches.

As a result, this investigation centers predominantly on the application of deep learning for detecting fraud within smart grids.

The primary goal of this paper is to explore the potential application of convolutional neural networks (CNNs), specifically in their 1D or 2D variants, for the purpose of identifying electricity consumers engaged in fraudulent activities. By leveraging the capabilities of CNNs, we aim to investigate their effectiveness in detecting and flagging instances of electricity fraud. Through this research, we seek to contribute to the development of advanced techniques that can aid in the identification and prevention of fraudulent behavior within the electricity sector.

Distinguishing itself from prior studies, our approach exhibits the following unique characteristics:

- The problem of missing values has been resolved by imputing the averages of daily electricity consumption for each customer category (fraud or no fraud) into the respective columns.
- Various deep Convolutional Neural Network (CNN) architectures have been proposed.
- The integration of CNNs has yielded notable enhancements in accuracy, AUC, and other assessment metrics, attributed to their capacity for feature extraction and generalization.
- Subsequent to the implementation of these aforementioned measures, the need to address the challenge of highly skewed data, linked to instances of electricity theft, has been obviated.

The paper is structured as follows: In Section 2, we present an in-depth exploration of energy fraud terminology, covering key concepts such as Energy Loss Types, Electricity Theft Methods, and Smart Grids Fraud Methods. This section aims to provide a comprehensive understanding of the various types of fraud involved in the energy sector. Moving on to Section 3, we conduct a thorough review of relevant literature to establish a strong background for our research. By examining previous works and studies, we gain valuable insights and identify any existing gaps in knowledge that our research aims to address. Section 4 focuses on our suggested model design, where we outline the steps involved in its development. This model serves as a framework for effectively combating energy fraud. Additionally, Section 5 presents the findings from our study, offering a comprehensive overview of the research outcomes. We also introduce the implementation tools we utilized, shedding light on their effectiveness in addressing and mitigating energy fraud. Finally, in the concluding section, we summarize the key findings, discuss the implications of our study, and highlight potential avenues for future research and improvement in the field of energy fraud prevention.

2. THEORETICAL BACKGROUND

Energy plays a crucial role in advancing economies and technological development. The infrastructure required for service provision encompasses extensive networks of pipelines or transmission lines, reaching millions of meters for monitoring individual customer consumption.

2.1 Energy loss types

Within the energy sector, a persistent challenge is the discrepancy between energy billing and supply, known as energy loss. These losses are typically categorized as Technical Losses (TL) and Non-Technical Losses (NTL) [2].

Technical Losses (TLs) occur due to energy dissipation in the electricity transmission system, specifically the joule effect on power lines and transformers made of copper or iron [3]. TLs are a normal part of the system, relatively constant, and independent of consumer behavior. They have a tolerable impact on the economy. Calculating TLs is complex, as it involves determining the point of loss and estimating the amount of energy lost. While it is not possible to completely eliminate TLs, they can be reduced by implementing modulation techniques throughout the system [4, 5].

Non-Technical Losses (NTLs) are the residual losses that remain unaccounted for after subtracting the calculated TLs. NTLs represent abnormal power consumption patterns that cannot be explained theoretically. Estimating NTLs is challenging as they are typically caused by external factors outside the power system, such as electricity theft. NTLs can significantly harm power providers' economic performance and lead to safety issues, including equipment damage, power outages, and injuries [6].

2.2 Electricity theft methods

Electricity theft is the practice of tampering with electrical equipment or circumventing energy meters in an effort to lower usage or avoid being charged. By putting in secret methods to go around the meter, this illegal technique aims to stop the proper recording of electricity usage. Power grids suffer from the effects of electricity theft, which lowers operational revenues and has negative effects on the quality of the power supply.

Figure 1 illustrates various methods of electricity theft, highlighting the diverse ways through which illegal consumption of electricity can occur. Among these methods, meter tampering stands out as the most prevalent form of electricity theft. Meter tampering involves fraudulent manipulation of the meter reading on an electromechanical meter device. By tampering with the meter, fraudsters aim to deceive the system and avoid accurate measurement of their

actual energy usage. Another method of theft is through unregistered connections, where consumers do not report their meter readings to the electricity supply company. Direct hooking from the mainline of a high transmission line (HTL) is another prevalent technique, accounting for 80% of worldwide electricity theft. In this method, consumers tap directly into the HTL without using the electricity meter panel [7].

Modifying the meter itself is another way to steal electricity. This can involve pressing outer materials into the meter, creating holes in the electromechanical meter, using highly viscous liquids to reset meter readings, damaging the rotating density coil with meter screws, or using solid neodymium magnets to disrupt the disk. Electromechanical meters are susceptible to tampering using these methods. Additional methods of meter modification include inverse meter reading, where intruders reverse the actual meter reading. This is typically done by opening the protective shield cover of the electricity meter.

These are just a few examples of the different techniques used in electricity theft, highlighting the need for effective fraud detection and prevention measures to safeguard the integrity of the power system [8].

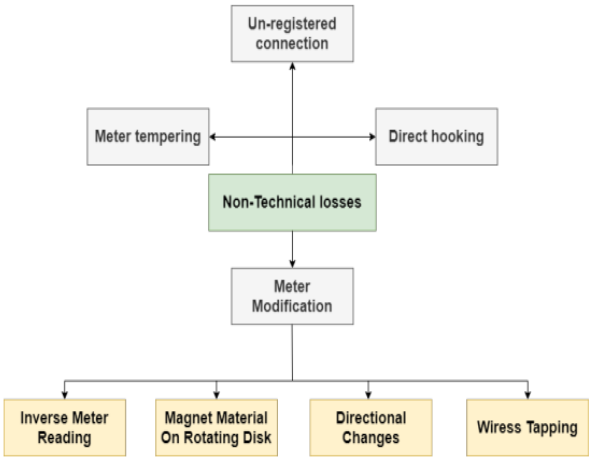

Figure 1. Energy consumption fraud methods

2.3 Smart grids fraud methods

In order to prevent electricity theft and reduce losses in energy transmission and distribution, certain countries have implemented the transformation of their conventional power grids into smart grids. This transition not only delves deeper into the concept of smart grids but also addresses potential fraudulent practices associated with them.

In a bid to counteract conventional electricity theft techniques, numerous nations have undertaken the transformation of their conventional grids into smart grids. A smart grid can be defined as the utilization of digital information technology to optimize the generation, distribution, and utilization of electrical power [9]. This term encapsulates the integration of communication and control functionalities into the traditional grid, comprising transmission lines, substations, transformers, and other key components responsible for the transmission of electricity from power plants to residences and commercial establishments [10]. Regrettably, despite the advancement in technology, instances of electrical network breaches and subsequent electricity theft persist. Upon a thorough examination of security concerns surrounding smart grids, certain vulnerabilities have been identified that may lead to instances of electricity theft [11]:

• Fraudsters have the capability to manipulate or alter meter firmware, enabling complete control over the meter's functions. Organized criminal entities exploit firmware updates, generating revenue by marketing meter hacking kits. However, a more significant apprehension revolves around the surreptitious consumption of electricity without triggering detection mechanisms.

• Password exploitation: Smart meters store passwords that can be communicated through messages, allowing fraudsters to obtain these passwords, thus facilitating non-technical losses (NTL) fraud.

• Man-In-the-Middle (MIM) attacks involve intercepting confidential data like network passwords and keys. Such attacks occur when an intruder infiltrates communication between smart meters or between a smart meter and the head end of a smart grid, deceiving them into believing they are engaged in direct interaction.

• Key spoofing involves the illicit acquisition of encryption keys for the purpose of decrypting messages or subverting authentication processes. This fraudulent activity includes extracting keys from smart meters and relies on diverse side-channel attacks, such as channel timing attacks and cold boot attacks.

2.4 Electricity fraud detection impacts and methods

The identification of electricity theft is a pervasive issue with serious ramifications that call for better solutions. Numerous elements, such as economic inequality, high energy costs, and lax law enforcement, have an impact on the frequency of electricity theft. These circumstances foster an environment where people and businesses turn to illicit means of obtaining electricity in order to avoid paying for it. The effects of energy theft are vast and include operational, social, and financial ramifications. Due to significant financial losses, utility firms experience diminished revenue, impaired profitability, and a lack of resources for infrastructure maintenance and upgrades. In consequence, this may lead to recurrent power outages and an unstable energy supply for lawful users. Governments also experience a decline in tax revenue, which limits their capacity to spend money on public services and pursue energy sector changes. Additionally, since law-abiding consumers must pay higher prices to cover the losses brought on by theft, power theft threatens the justice and equity of the energy system.

Addressing electricity theft detection presents several challenges that need to be overcome. Traditional detection methods, such as manual inspections, are often inefficient, time-consuming, and prone to errors. Perpetrators employ sophisticated techniques to bypass detection, making it difficult to accurately identify instances of theft. Moreover, weak law enforcement, corruption, and societal acceptance of theft in some regions further complicate the issue. To tackle these challenges, there is a need for better solutions. Advanced technologies, such as smart metering, data analytics, and machine learning algorithms, hold promise for more accurate and efficient detection of theft patterns and anomalies. Collaborative efforts among utility companies, governments, and regulatory authorities are crucial to strengthen enforcement, implement stricter penalties, and raise public awareness about the detrimental effects of electricity theft.

Additionally, addressing the underlying socioeconomic factors that drive theft, such as economic inequality and high energy costs, can contribute to long-term solutions. By developing comprehensive strategies that combine technological advancements, robust enforcement, and social interventions, stakeholders can work together to mitigate electricity theft, protect revenue, and ensure reliable and equitable energy distribution. Long-term solutions can also be aided by tackling the underlying socioeconomic issues, such as economic inequality and high energy costs, that motivate theft. Stakeholders can collaborate to reduce electricity theft, protect income, and guarantee dependable and equitable energy distribution by adopting comprehensive solutions that integrate technology developments, strict enforcement, and social interventions.

The present techniques for Non-Technical Loss (NTL) detection can be classified as follows:

• Hardware-Based Solutions: Approaches reliant on the creation and design of devices aimed at detecting and estimating fraudulent activities [12].

The implementation of options such as meter reversal and disconnection offers the advantage of completely eradicating theft. Additionally, advanced systems provide the capability to detect non-technical losses (NTLs) both at the meter and within the network, improving overall efficiency. Moreover, these systems enable the identification of various NTLs throughout the grid. However, there are disadvantages to consider, such as the high cost associated with installing hardware in numerous households and the financial burden of equipment expenses. Furthermore, the effectiveness of anti-theft measures heavily relies on the availability and integration of smart metering systems.

• This approach is centered around identifying alternatives to hardware-based methods, which are often impractical for numerous Power Distribution Companies (PDCs), especially those in underdeveloped nations where creating new infrastructure can be cost-prohibitive. The proposed solution involves utilizing software-based classification techniques to determine and estimate the existence of Non-Technical Losses (NTLs) based on analysis of consumers' electricity consumption data [12].

3. LITERATURE REVIEW

This section will examine previous studies that have utilized various machine learning and deep learning techniques to develop systems for detecting fraudulent behavior among smart grid customers:

3.1 Machine learning-based works

Nagi et al. [13] employed Support Vector Machines (SVMs) to evaluate fraudulent activities within a power grid. The study utilized historical customer data extracted from TNBD's electronic Customer Information Billing System (e-CIBS), encompassing a dataset of 265,870 customers spanning 25 months. The researchers constructed an SVM classifier, training it with 330 profiles representing legitimate usage patterns and 53 profiles associated with fraudulent activities. This training yielded an accuracy of 86.43%.

In the realm of fraud detection, various supervised learning algorithms, including K-Nearest Neighbors (KNN), Logistic Regression, Support Vector Machines (SVMs), and Extreme Gradient Boosted Trees (XGBoost), were applied [14]. Testing was conducted on real data from Spain's prominent distribution company, Endesa, yielding an impressive 91% Area Under the Curve (AUC) score through the employment of the XGBoost classifier.

Pereira and Saraiva [15] compared various data balancing techniques, including cost-sensitive learning, random undersampling, random oversampling, K-Medoids-based undersampling, SMOTE, and Cluster-Based Oversampling (CBOS). Machine learning techniques like Logistic Regression, Random Forest, SVMs, and ANNs were employed to detect power theft.

3.2 Neural networks and deep learning based works

Ford et al. [16] utilized Artificial Neural Networks (ANNs) to identify fraudulent activity within the smart grid. Their study consisted of two main phases. Initially, they trained the neural network using energy consumption data from the European Central Bank (ECB). Next, they developed a model simulating a scenario where a malevolent actor tampered with a smart meter, causing temporary disruptions in readings. Their findings demonstrated an accuracy of 84.37%.

Costa et al. [17] introduced a method for identifying energy fraud via a multilayer perceptron. Their research harnessed data from a Brazilian electric power distribution firm encompassing over seven million consumers. Their outcomes indicated an accuracy of 87.17%, precision of 65%, and recall of 29.5%.

A CNN and LSTM based strategy for detecting electricity theft was proposed [5]. To address data gaps, the authors devised an inventive data preparation algorithm founded on local values. Results showcased an F1-score of 94%, accuracy of 89%, precision of 92%, and recall of 96%.

In the context of electricity fraud identification, Rouzbahani Aldegheishem et al. [18] proposed two distinct models. The initial model employed Synthetic Minority Oversampling Technique (SMOTE) and processed nearest neighbor algorithm for data balance while leveraging AlexNet for dimensionality reduction and feature extraction. Light Gradient Boosting served as the classification algorithm, achieving an AUC of 90.6%. The second model employed a Generative Adversarial Network (GAN) with gradient penalty and utilized GooLeNet for dimensionality reduction, with adaptive boosting as the classification algorithm. This model excelled, attaining an AUC of 96%.

A methodology combining CNN and Bidirectional Gated Recurrent Unit (BiGRU) was put forth [19], using data from a Colombian electric provider. The reported accuracy, precision, recall, and AUC values were 92.9%, 88.5%, and 96.6% respectively.

To address data imbalance in the SGCC dataset, authors [20] proposed a strategy involving random bagging to create balanced subsets. A model composed of nine deep CNNs with a voting system achieved a precision of 90%, recall of 91%, and accuracy of 89%.

Based on the cited works, it can be observed that studies utilizing simple machine learning algorithms have generally achieved lower accuracy compared to those employing more complex architectures based on deep learning approaches. The use of deep learning techniques allows for the extraction of intricate patterns and features from the data, leading to improved performance in electricity theft detection. However, it is important to note that even with deep learning models,

additional techniques are often required to address the challenge of imbalanced data.

In summary, the works cited above suggest that utilizing complex deep learning architectures can improve accuracy in electricity theft detection compared to simple machine learning algorithms. However, it is necessary to employ additional techniques to tackle the challenge of imbalanced data and ensure robust and reliable detection performance.

By leveraging the inherent capabilities of CNNs in extracting spatial and temporal features from data, our approach successfully overcomes the challenges associated with imbalanced data. The CNN architecture's capacity to learn and represent intricate patterns within the data likely contributes to its ability to achieve higher accuracy in electricity theft detection. While our approach does not incorporate additional techniques tailored for imbalanced data, it is worth exploring the potential benefits of integrating such techniques to further enhance the performance and robustness of the model in scenarios with varying degrees of data imbalance.

4. PROPOSED METHODOLOGY

In this study, we introduced a novel approach utilizing convolutional neural networks (CNNs) for the purpose of identifying instances of electricity theft. The architecture of our model is visually represented in Figure 2.

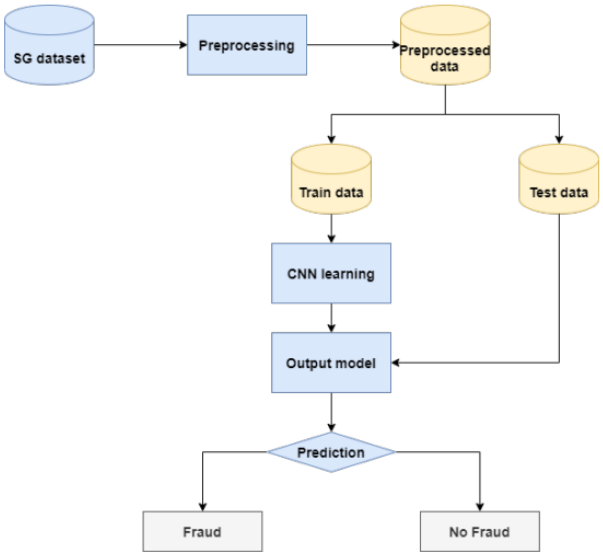

Figure 2. Schematic representation of the proposed system

Our approach commences with the acquisition of the Smart Grid (SG) dataset, which is a one-dimensional dataset requiring refinement and filtration for effective training suitability. The preprocessed dataset is then partitioned into two segments: the training set and testing sets. The training set is employed to facilitate the training and construction of predictive models using the 1D/2D-CNN algorithms. Following the training phase, the derived models are subsequently employed to analyze the testing dataset.

4.1 Dataset description

We employed an authentic electricity consumption dataset made accessible by the State Grid Corporation of China (SGCC) [3]. The dataset encompasses electricity consumption records for 42,372 customers spanning 1,034 days (from January 1, 2014, to October 31, 2016). Among these records, approximately 9% (3615 individuals) were identified as engaging in fraudulent activities (refer to Table 1).

To prepare the dataset for analysis, we conducted cleaning, filtering, and addressed missing data.

Table 1. Dataset description

	Normal	Fraud	Total
Before Cleaning	38757	3615	42372
After Cleaning	36679	3579	40258

4.2 Preprocessing

Data preprocessing plays a crucial role in enhancing data quality, making it more amenable to extracting meaningful insights. This phase encompasses a series of tasks, including (see Figure 3):

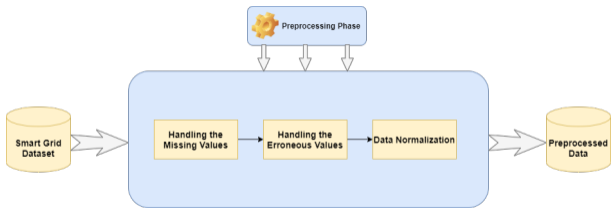

Figure 3. Preprocessing phase

• Removing Null Rows: Within the SGCC dataset, certain entries are found to be empty. Consequently, a decision was made to eliminate these rows, resulting in the removal of 2,114 rows during this process (refer to Table 1).

• Addressing Missing Values: A significant challenge within the SGCC dataset is the presence of missing values. Various factors contribute to these gaps, including meter and power failures, unscheduled maintenance, sensor damage, and cyberattacks [20]. A distinctive contribution of our work involves an innovative approach for addressing these gaps in the dataset. Our strategy is rooted in the assumption that electricity consumption patterns of fraudulent customers might exhibit similarities. As a result, we propose categorizing the training set into two groups: those involved in fraud and those not engaged in fraudulent activities. Subsequently, for each consumer category, we employ an approach to impute the missing values for each day using the average daily electricity consumption. To prevent data leakage, we ensure that the averages from the training dataset are employed to fill in the missing values in the testing dataset.

• Data Normalization: Employing data normalization enhances the consistency of diverse entities within a data model. The data normalization process encompasses multiple tasks. Initial focus is on purging any duplicated entries within the dataset. Subsequently, emphasis is laid on logically structuring the data. The neural network's effectiveness can be impacted by the variability of data values. To mitigate this concern, dataset normalization becomes essential. Our approach involves the utilization of the MAX-MIN scaling technique, as described by the following equation:

$$F(x_i) = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

4.3 Convolutional neural networks learning

Convolutional neural networks (CNNs) are widely utilized in various Computer Vision tasks [21], encompassing image processing and natural language processing, among others. The term "convolution" denotes a mathematical operation that combines multiple functions. A traditional CNN architecture comprises one or more convolutional layers, followed by a pooling layer and a fully connected layer [22].

The attributes of CNNs remain consistent across different dimensions—be it 1D, 2D, or 3D. These dimensions share common traits and methodologies. Nevertheless, the primary distinction lies in the input data's dimensionality and the manner in which filters, also known as convolution kernels or feature extractors, traverse the data.

Given our focus on the smart grid domain, the available data pertains solely to one-dimensional (1D) power consumption data, which manifests as sequential data. Our approach centers around the use of the 1D-CNN model for identifying fraudulent customers. Furthermore, we explore the potential of the 2D-CNN model, which converts one-dimensional input into a two-dimensional matrix, thereby enabling the transformation of data.

4.4 Evaluation metrics

Numerous metrics are available for assessing the effectiveness of classification models. The primary method for evaluation and validation involves utilizing a confusion matrix. This matrix summarizes the performance of the classification algorithm, computing the following:

• True Positives (TP) and True Negatives (TN): Count of correctly predicted samples.

• False Positives (FP) and False Negatives (FN): Count of misclassified samples.

Furthermore, our models were evaluated using metrics (defined in Table 2) such as Precision, Recall, F1-score, and AUC.

Table 2. Evaluation metrics for classification models

Metric	Definition	Formula
Precision	Assesses the model's overall predictive performance across multiple classes.	$Precision = \frac{TP}{TP + FP}$
Recall	Analyzes the model's ability to discover all positive individuals.	$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$
F1-score	A metric that combines precision and recall values.	$\text{F1} - \text{score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$
AUC	AUC stands for Area Under the Curve. AUC curve represents the relationship between false positives (FPR) and true positives (PR).	$PR = \frac{TP}{TP + FN}$ $FPR = \frac{FP}{FP + TN}$

5. RESULTS AND DISCUSSION

We executed various components of this project utilizing the Kaggle platform. Additionally, open-source machine-learning libraries like Tensorflow, Keras, and Pandas were leveraged. Our models were compiled with a categorical Cross-Entropy Loss Function and an Adam Optimizer. To address missing values in the dataset, the SimpleImputer function from the sklearn library was applied. The models were trained for 100 epochs.

Validation accuracy assessment was performed using the test dataset as a validation set. Throughout the learning phase, we preserved the best-performing model that exhibited optimal results for consumer class prediction.

The initial architecture employed (see Figure 4) comprises a 1D-CNN design, featuring two convolution layers, a batch normalization layer, a fully connected layer, and a dense layer. The progression of learning and validation rates is depicted in Figure 5.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_2 (Conv1D)            (None, 1033, 32)          96
_________________________________________________________________
conv1d_3 (Conv1D)            (None, 1032, 64)          4160
_________________________________________________________________
batch_normalization_1 (Batch (None, 1032, 64)          256
_________________________________________________________________
flatten_1 (Flatten)          (None, 66048)             0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 66049
=================================================================
Total params: 70,561
Trainable params: 70,433
Non-trainable params: 128
```

Figure 4. Proposed 1D-CNN structure

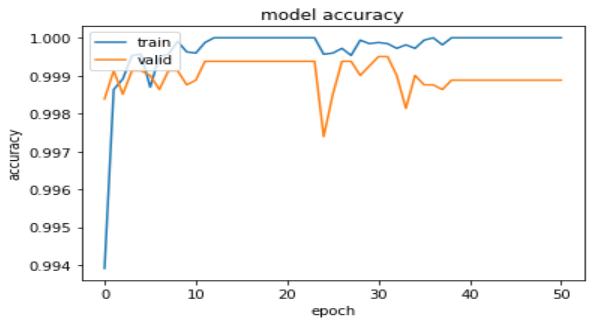

Figure 5. 1D-CNN model accuracy evolution

The second CNN model (refer to Figure 6) employed is a 2D-CNN architecture. In this configuration, we converted our 1D data into a 2D format, yielding a 33x33 matrix. To ensure a column count that is a multiple of 33, we introduced 55 additional zero columns. Our 2D-CNN model consists of three convolution layers, followed by a fully connected layer and a dense layer. Figure 7 illustrates the progression of learning and validation accuracy rates throughout the training process.

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 33, 33, 32)        160
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 33, 33, 64)        8256
_________________________________________________________________
flatten_1 (Flatten)          (None, 69696)             0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 69697
=================================================================
Total params: 78,113
Trainable params: 78,113
Non-trainable params: 0
```

Figure 6. Proposed 2D-CNN Structure

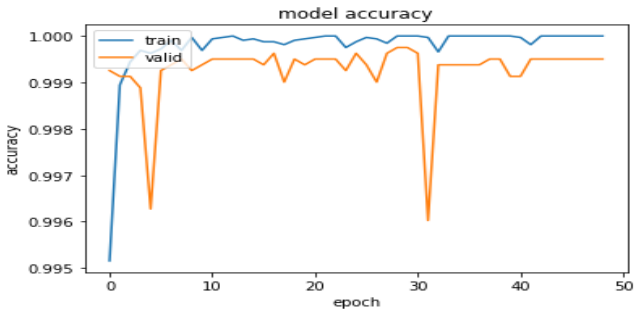

Figure 7. 2D-CNN model accuracy evolution

The confusion matrix is shown in Figure 8 for a 1D-CNN model that uses 80% training data. This matrix reveals the need for enhanced accuracy in classifying 4 out of 703 fraudulent customers, while accurately identifying all normal customers. Consequently, the recall measure for this model stands at 99.42%, with precision reaching 100%. The F1-score and AUC measurements register at 99.71%, as detailed in Table 2, signifying the model's excellence.

In contrast, Figure 9 portrays the confusion matrix for the 2D-CNN model. In this scenario, only two fraudulent customers are misclassified, leading to a recall rate of 99.71%, precision of 100%, and an F1-score of 99.86%. The AUC value, as evidenced in Table 3, nearly reaches one, indicating the effectiveness of the 2D-CNN model.

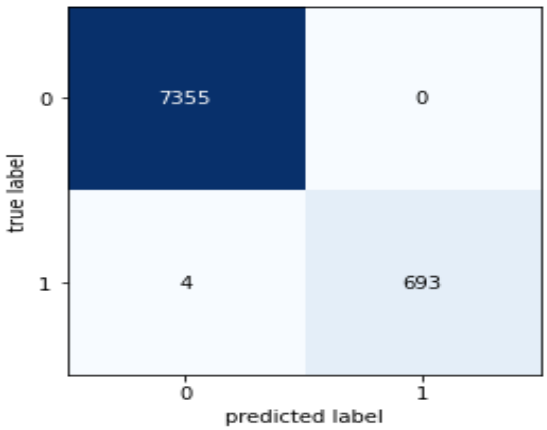

Figure 8. Confusion matrix evaluating our 1D-CNN model

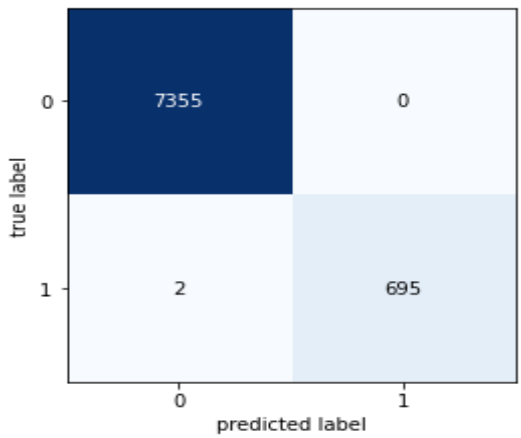

Figure 9. Confusion matrix evaluating our 2D-CNN model

The outcomes of the experiments, with adjustments made to the training and test set ratios (60:40, 70:30, 80:20, and 90:10), are presented in Tables 3 and 4. These tables reveal a slight performance advantage of the 1D-CNN-based model (Table 3) over the 2D-CNN-based model (Table 4). Notably, altering the training set size did not yield any noticeable impact on the superior classification prowess exhibited by the proposed models.

Table 3. Results achieved by the 1D-CNN model

	90:10	**80:20**	**70:30**	**60:40**
Accuracy	99.93%	99.95%	99.94%	99.91%
Recall	99.17%	99.42%	99.44%	99.14%
Precision	100%	100%	99.90%	99.78%
F1-score	99.58%	99.63%	99.67%	99.45%
AUC	99.59%	99.71%	99.71%	99.56%

Table 4. Results attained from the 2D-CNN model

	90:10	**80:20**	**70:30**	**60:40**
Accuracy	99.93%	99.97%	99.95%	99.93 %
Recall	99.17%	99.71%	99.44%	99.36%
Precision	100%	100%	100%	99.79%
F1-score	99.58%	99.85%	99.72%	99.57%
AUC	99.58%	99.85%	99.72%	99.67%

Table 5. A contrast with prior research

Approach	**Recall**	**Precision**	**F1-Score**	**AUC**
2D-CNN	**99.44%**	**100%**	**99.72%**	**99.72%**
1D-CNN	**99.44%**	**99.90%**	**99.67%**	**99.71%**
EDCNN [19]	91%	90%	89%	99.30%
CNN-LSTM [5]	91%	87%	89%	-
Wide-deep CNN [3]	-	97%	-	78.60%

In Table 5, a comparison is drawn between our proposed models and prior investigations employing the SGCC dataset and a training set ratio of 70%. Notable among these studies are Wide and Deep CNN [3], CNN-LSTM [5], and EDCNN [19]. The results underscore the superior performance of our suggested models in detecting electricity theft, surpassing all preceding endeavors.

6. CONCLUSIONS AND FUTURE WORKS

The objective of this study was to develop a deep learning-based classification model to detect fraudulent behavior in power usage, with a specific focus on the State Grid Corporation of China (SGCC) as a case study. To mitigate NTLs, we employed deep convolutional neural networks (CNNs) in our models, we proposed a new method, to fill in the missing values, based on consumers' behavior. It should be noted that this proposal led to an amazing improvement in the results obtained.

The results demonstrated that our CNN models surpassed earlier studies discussed, with 2D-CNNs (99.97%) yielding more accurate outcomes compared to 1D-CNNs (99.95%). Consequently, convolutional neural networks offer interesting implications, as they can be applied to identify fraudulent energy consumers. Overall, our models exhibited favorable classification results. So, our proposed models for electricity theft detection have significant real-world applications and impact. They empower utility companies to protect their revenue streams, allocate resources efficiently, and enhance service quality.

Moving forward, this research aims to explore avenues for

improving the model, including the utilization of alternative datasets, investigating various deep learning approaches, and exploring different preprocessing methods.

REFERENCES

[1] Alotaibi, I., Abido, M.A., Khalid, M., Savkin, A.V. (2020). A comprehensive review of recent advances in smart grids: A sustainable future with renewable energy resources. Energies, 13(23): 6269. https://doi.org/10.3390/en13236269

[2] Coma-Puig, B., Carmona, J., Gavalda, R., Alcoverro, S., Martin, V. (2016). Fraud detection in energy consumption: A supervised approach. In 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Canada, pp. 120-129. https://doi.org/10.1109/DSAA.2016.19

[3] Zheng, Z., Yang, Y., Niu, X., Dai, H.N., Zhou, Y. (2017). Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. IEEE Transactions on Industrial Informatics, 14(4): 1606-1615. https://doi.org/10.1109/TII.2017.2785963

[4] Saeed, M.S., Mustafa, M.W., Hamadneh, N.N., Alshammari, N.A., Sheikh, U.U., Jumani, T.A., Khalid, S.B.A., Khan, I. (2020). Detection of non-technical losses in power utilities-A comprehensive systematic review. Energies, 13(18): 4727. https://doi.org/10.3390/en13184727

[5] Hasan, M.N., Toma, R.N., Nahid, A.A., Islam, M.M., Kim, J.M. (2019). Electricity theft detection in smart grid systems: A CNN-LSTM based approach. Energies, 12(17): 3310. https://doi.org/10.3390/en12173310

[6] Yang, Y., Song, R., Xue, Y., Zhang, P., Xu, Y., Kang, J., Zhao, H. (2022). A detection method for group fixed ratio electricity thieves based on correlation analysis of non-technical loss. IEEE Access, 10: 5608-5619. https://doi.org/10.1109/ACCESS.2022.3141610

[7] Hussain, Z., Memon, S., Dhomeja, L., Abbasi, S. (2017). Analysis of non-technical electrical power losses and their economic impact on Pakistan. Sindh University Research Journal-SURJ (Science Series), 49(2).

[8] Hussain, Z., Memon, S., Shah, R., Bhutto, Z.A., Aljawarneh, M. (2016). Methods and techniques of electricity thieving in Pakistan. Journal of Power and Energy Engineering, 4(9): 1-10. http://dx.doi.org/10.4236/jpee.2016.49001

[9] Bayindir, R., Colak, I., Fulli, G., Demirtas, K. (2016). Smart grid technologies and applications. Renewable and Sustainable Energy Reviews, 66: 499-516. https://doi.org/10.1016/j.rser.2016.08.002

[10] Kabalci, Y. (2016). A survey on smart metering and smart grid communication. Renewable and Sustainable Energy Reviews, 57: 302-318. https://doi.org/10.1016/j.rser.2015.12.114

[11] Han, W., Xiao, Y. (2016). Non-technical loss fraud in advanced metering infrastructure in smart grid. In International Conference on Cloud Computing and Security, pp. 163-172. https://doi.org/10.1007/978-3-319-48674-1_15

[12] Viegas, J.L., Esteves, P.R., Mel´ıcio, R., Mendes, V., Vieira, S.M. (2017). Solutions for detection of non-technical losses in the electricity grid: A review. Renewable and Sustainable Energy Reviews, 80: 1256–1268. https://doi.org/10.1016/j.rser.2017.05.193

[13] Nagi, J., Yap, K.S., Tiong, S.K., Ahmed, S.K., Mohamad, M. (2009). Nontechnical loss detection for metered customers in power utility using support vector machines. IEEE Transactions on Power Delivery, 25(2): 1162-1171. https://doi.org/10.1109/TPWRD.2009.2030890

[14] Buzau, M.M., Tejedor-Aguilera, J., Cruz-Romero, P., G´omez-Exp´osito, A. (2018). Detection of non-technical losses using smart meter data and supervised learning. IEEE Transactions on Smart Grid, 10(3): 2661-2670. https://doi.org/10.1109/TSG.2018.2807925

[15] Pereira, J., Saraiva, F. (2021). Convolutional neural network applied to detect electricity theft: A comparative study on unbalanced data handling techniques. International Journal of Electrical Power & Energy Systems, 131: 107085. https://doi.org/10.1016/j.ijepes.2021.107085

[16] Ford, V., Siraj, A., Eberle, W. (2014). Smart grid energy fraud detection using artificial neural networks. In 2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG), Orlando, FL, USA, pp. 1-6. https://doi.org/10.1109/CIASG.2014.7011557.

[17] Costa, B.C., Alberto, B.L., Portela, A.M., Maduro, W., Eler, E.O. (2013). Fraud detection in electric power distribution networks using an ANN-based knowledge-discovery process. International Journal of Artificial Intelligence & Applications, 4(6): 17-23. http://doi.org/10.5121/ijaia.2013.4602

[18] Rouzbahani Aldegheishem, A., Anwar, M., Javaid, N., Alrajeh, N., Shafiq, M., Ahmed, H. (2021). Towards sustainable energy efficiency with intelligent electricity theft detection in smart grids emphasizing enhanced neural networks. IEEE Access, 9: 25036-25061. https://doi.org/10.1109/ACCESS.2021.3056566

[19] Duarte Soares, L., de Souza Queiroz, A., López, G.P., Carreño-Franco, E.M., López-Lezama, J.M., Muñoz-Galeano, N. (2022). BiGRU-CNN neural network applied to electric energy theft detection. Electronics, 11(5): 693. https://doi.org/10.3390/electronics11050693

[20] Rouzbahani, H.M., Karimipour, H., Lei, L. (2020). An ensemble deep convolutional neural network model for electricity theft detection in smart grids. In 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Canada, pp. 3637-3642. https://doi.org/10.1109/SMC42975.2020.9282837

[21] Mukkapati, N., Anbarasi, M.S. (2022). Brain tumor classification based on enhanced CNN model. Revue d'Intelligence Artificielle, 36(1): 125-130. https://doi.org/10.18280/ria.360114

[22] Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. Insights into Imaging, 9(4): 611-629. https://doi.org/10.1007/s13244-018-0639-9