




Article

Electricity Theft Detection in Smart Grid Systems: A CNN-LSTM Based Approach

Md. Nazmul Hasan ¹, Rafia Nishat Toma ¹, Abdullah-Al Nahid ¹, M M Manjurul Islam ² and Jong-Myon Kim ^{2,*}

¹ Electronics and Communication Engineering Discipline, Khulna University, Khulna 9208, Bangladesh

² School of Electrical, Electronics and Computer Engineering, University of Ulsan, Ulsan 44610, South Korea

* Correspondence: jmkim07@ulsan.ac.kr; Tel.: +82-52-259-2217

Received: 24 July 2019; Accepted: 26 August 2019; Published: 28 August 2019



Abstract: Among an electricity provider's non-technical losses, electricity theft has the most severe and dangerous effects. Fraudulent electricity consumption decreases the supply quality, increases generation load, causes legitimate consumers to pay excessive electricity bills, and affects the overall economy. The adaptation of smart grids can significantly reduce this loss through data analysis techniques. The smart grid infrastructure generates a massive amount of data, including the power consumption of individual users. Utilizing this data, machine learning and deep learning techniques can accurately identify electricity theft users. In this paper, an electricity theft detection system is proposed based on a combination of a convolutional neural network (CNN) and a long short-term memory (LSTM) architecture. CNN is a widely used technique that automates feature extraction and the classification process. Since the power consumption signature is time-series data, we were led to build a CNN-based LSTM (CNN-LSTM) model for smart grid data classification. In this work, a novel data pre-processing algorithm was also implemented to compute the missing instances in the dataset, based on the local values relative to the missing data point. Furthermore, in this dataset, the count of electricity theft users was relatively low, which could have made the model inefficient at identifying theft users. This class imbalance scenario was addressed through synthetic data generation. Finally, the results obtained indicate the proposed scheme can classify both the majority class (normal users) and the minority class (electricity theft users) with good accuracy.

Keywords: smart grid; electricity; energy; non-technical loss; data analysis; machine learning; convolutional neural network (CNN); long short-term memory (LSTM)

1. Introduction

Because of the high cost of acquiring energy, as well as the limited amount of energy resources, efficient and operative use of energy resources is a very important aspect of social and economic development for any country. The smart grid has become a key solution for making the greatest use of future energy monitoring. The smart grid system can be described as an entire electricity network consisting of the power system infrastructure and computers to manage and monitor the energy usage, along with an intelligent monitoring system that tracks the usage pattern and mode of action of all consumers connected with the system [1]. The smart grid provides the utilities' and customers' facility to monitor, control, and predict energy use by integrating modern digital equipment with the existing electrical system. In this system, the collector device delivers usage readings to the operational center using the internet, and the power transmission company performs the billing process depending on these readings. At the same time, the operation center collects user readings from neighborhood customers' periodic updates through a wireless network. The main target is to reduce losses due to energy wastage and provide viable, cost-effective, and secure electricity supplies [2]. The device that

performs usage reporting is known as a smart meter; it is a computerized version of an ancient meter. The processor, nonvolatile storage, and communication facilities, along with the ability to maintain widespread customer energy generation, make smart meters an important part of smart grid systems.

Today, electric power loss has become one of the most conspicuous issues affecting both conventional power grids and smart grids. From the statistics, it has been shown that transmission and distribution losses increased from 11% to 16% between the years 1980 to 2000 [3]. The electricity losses vary from country to country. The losses in the USA, Russia, Brazil, and India were 6%, 10%, 16%, and 18%, respectively, of their total energy production [4]. The difference between the energy produced in one system and the metered energy delivered to the users is known as the power loss. To determine the amount of electricity loss, smart meters in smart grids play a prominent role. Advanced energy meters obtain information from the consumers' load devices and measure the consumption of energy in intervals of an hour. The energy meter provides additional information to the utility company and the system operator for better monitoring and billing, and provides two-way communications between the utility companies and consumers [5]. However, it is also possible to limit the maximum amount of electricity consumption, which can terminate as well as re-connect the supply of electricity from any remote place [6].

Electricity loss is mainly classified into two categories, namely technical loss (TL) and non-technical loss (NTL). TL occurs because of the joules effect on power lines and transformer loss during the transportation of electricity [7]. The calculation of TL is quite complex, making it difficult to locate the point of loss and estimate the amount of energy destroyed. The TL cannot be stopped completely, but it can be reduced by applying some modification techniques throughout the system. The NTL can be defined simply as the difference between a total loss and TL [8]. The main causes of NTL are billing delay and irregularities, energy theft, faulty energy meters, fraud, and unpaid bills [9]. Many researchers mentioned a different class of loss, named gratis, which occurs when electricity is provided free of charge [10]. In recent times, cyber tampering has become a way to malevolently change the consumption data patterns of the smart meter, which decreases the bill for the users [11]. Though such electricity loss is caused by a small number of consumers, it decreases the profitability and energy efficiency of the utility companies, resulting in increased costs for all users and creating problems such as load-shedding, industrial routine hampering, and inflation. NTL costs a huge amount of money for both developed and developing countries such as USA, UK, Brazil, Malaysia, and India [12,13]. For example, these losses in the power sectors of India and Brazil cost approximately US \$44.5 and US \$3 billion per year respectively [14]. So, for these countries, the preclusion of electricity energy theft is a major challenge [15] when working to strengthen the economic state. Real-time fraud detection is the only way to get rid of this problem.

Among the various types of NTL, electricity theft viciously damages the revenue of the power sector company, which causes a significant loss of energy resources and damages the economy of any country. To mitigate this problem, several techniques have been implemented through prolific research to identify and abate the theft issue. Researchers categorize energy theft detection (ETD) systems into three basic groups, state-based, game theory-based and classification-based detection [16]. The use of upgraded devices and sensors in state-based detection results in higher accuracy in ETD, as proposed in earlier studies [9,17,18]. The main limitations of applying this detection system are vulnerability, the higher cost of hardware devices, and maintenance of the devices. Cardenas et al. presented a game theory-based detection to find an optimal solution, which was based on formulating various potential strategies [19]. Computing the utility function among distributors, regulators, and thieves is the biggest challenge in this process. The daily electricity consumption patterns of users can be analyzed by using machine learning algorithms to establish classification models, which include decision trees, random forest (RF), support vector machines (SVM) [20], neural networks (NN) [21], and so on.

The periodic description of power consumption plays a major role because the electricity consumption patterns of each user are very significant in the field of ETD. Depending on evaluation of

smart meter data, many researchers have proposed some techniques to identify fraud data. For example, Gue et al. [11] formed a three-level framework, Giani et al. [22] presented a phasor measurement units (PMU)-based security system to counterfeit cyber-attacks, and Najmeddine et al. [23] developed a matrix pencil approach for theft detection solutions. Along with that, the rough set and rule-based models were also applied to detect NTLs [20,24,25]. The conventional electricity theft detection method has been addressed with statistical techniques [26,27], comparing the abnormal and normal meter readings, fuzzy networks, and rough sets [28,29]. In a very recent paper [30], Bernat Coma-Puig and Josep Carmona utilized XGBoost, LightGBM, and CatBoost learning algorithms to detect NLT. In recent times, the new idea of smart grids brings a new era in unraveling electricity theft. In most cases, data from smart meters is being used for further implementation.

Though the amount of data held by power utility companies is too large in most cases, machine learning-based classification has achieved significant attention in recent times. The daily consumption data is applied to find theft patterns, which also preserves the privacy of consumers [16]. In [31–33], SVM was applied to detect anomalies and irregularities in the collected data by clustering and classification. In many algorithms, clustering is applied as a primary and secondary step, which makes this technique the most appropriate for modeling and identification of any energy consumption profile. Not only SVM, but other techniques such as one-class SVM, optimum path forest, and C4.5 decision tree were combined in [34], by combining the decision tree and the SVM [20], fuzzy logic with SVM [13], generic algorithm-support vector machine (GA-SVM) [35,36], extreme learning machine (ELM), and online sequential ELM (OS-ELM) [37–39] to achieve higher accuracy in theft detection.

The extension and availability of the internet increase the concern of cyber-attacks in the smart grid system. With the help of the Markle Hass concept, Wei et al. [35] proposed a security framework. In [40], the smart grid metering infrastructure could be linked with the security threads which make data more confidential and secure. Unsupervised methods such as fuzzy classification were performed with the computation of Euclidean distance to the cluster center in [41]. In [42], a wavelet-based technique was applied to analyze features to identify fraudulent consumers using an artificial neural network (ANN)-based approach. Another technique in [21] has shown better fraud detection efficiency in smart grid systems where an ANN is incorporated with an SVM to construct a hybrid scheme. Load profiles have become an alternative and cost-effective approach to identify fraudulent consumers [37]. Researchers have worked with different pattern recognition methods, which have been used as load profiling tools based on the local recorded pattern [43–45].

In addition to the classical machine learning approaches, deep learning approaches have achieved huge success in areas like image classification and computer vision [46], and speech recognition [47]. Due to its ability to handle and control huge amounts of data, automate feature extraction, and its classification process, deep learning techniques are utilized to build models to work with smart meter data originating from the smart grids. A wide and deep CNN structure was proposed in [7] to detect electricity theft in smart grid environments. Hybrid deep learning techniques have been utilized in recent times for load forecasting. A combination of CNN and LSTM structures was proposed in [48], where the model was used for short-term load forecasting. The proposed model performed quite well in comparison to other approaches. Another similar model was suggested in [49] for similar purposes. This CNN-LSTM hybrid structure is also used in predicting electricity price [50] and household power consumption [51]. In all of these applications, the CNN-LSTM model exhibited very good performance. In contrast, in [50,51], an electricity price was predicted using raw data without any preprocessing. The quality of data can be affected by various facets of the sensors, such as sensor power scale, noise level, and so on. However, it is inevitable to apply an appropriate preprocessing technique to achieve a generalized performance. These studies used hybrid structures like a CNN-LSTM model on electricity consumption datasets to develop a regression model. In this study, we combined the hybrid model, CNN-LSTM, with a preprocessing technique on the electricity consumption signature dataset to solve the classification problem. This encouraged us to exploit this hybrid structure to detect electricity theft by analyzing the irregular and abnormal consumption patterns of consumers.

The remainder of this paper is organized as follows. In Section 2, the overall methodology is presented, including the data preprocessing technique and the proposed CNN-LSTM model. The experimental results, with robustness and reliability analysis using numerous evaluation matrices, are also presented in Section 3. Finally, we conclude this paper in Section 4.

2. Materials and Methods

This work was intended to identify electricity theft from the power consumption pattern of users, utilizing CNN-LSTM-based deep learning techniques. This classifier model was trained utilizing a dataset consisting of daily power consumption data of both normal and fraudulent users in a supervised manner. First, the data was prepared by a data preprocessing algorithm to train the model. The preprocessing step also involved synthetic data generation for better performance. At the next stage, the proposed model was hypertuned and finally, the optimized model was evaluated via the test data. The overall methodology is depicted in Figure 1.

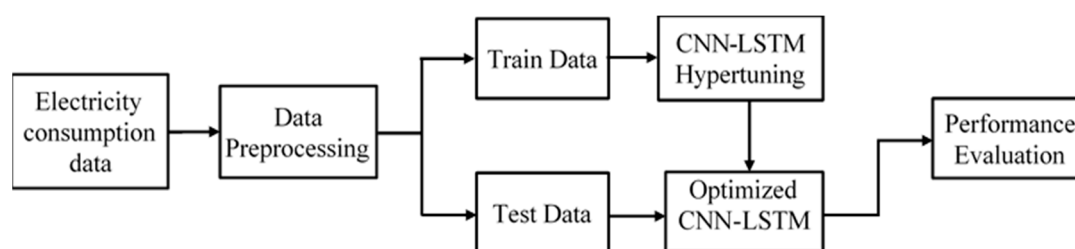


Figure 1. Block diagram of the proposed electricity theft detection system.

2.1. Electricity Theft Data

The study was carried out on a collection of real electricity consumption data of consumers, which was made available by the State Grid Corporation of China (<http://www.sgcc.com.cn/>). Table 1 presents the metadata information about the dataset. The dataset is composed of electricity consumption signatures of 9655 consumers over 1 year. Our primary observation revealed that a normal user and an abnormal user generate different patterns of electricity consumption. Figure 2 displays the fortnightly consumption of two consumers; one is a normal user and the other is an electricity thief. The consumption trend indicates that an abnormal or electricity theft user has a pattern that fluctuates more than a normal user.

Table 1. Metadata information of the electricity theft dataset.

Description	Value
Time window of data collection	1 January 2015–31 December 2015
Total number of consumers	9956
Number of normal users	8562
Number of aberrant user or electricity thieves	1394

Electricity consumption data is generally acquired through smart meters or various sensors located at the user end. The data is then aggregated to any central location through a data communication network. In this scenario, there is a possibility of smart meter failure, sensor malfunctioning, or faults in data transmission and the storage server. It is inherent that missing or erroneous data will be present in the electricity consumption datasets. In this dataset, we found numerous missing values. If those missing instances are just discarded, the size of the dataset shrinks considerably, and thus reliable analysis becomes difficult. To avoid downsizing the dataset, we proposed a data preprocessing algorithm. With the help of this algorithm, we have filled in the missing values in the dataset.

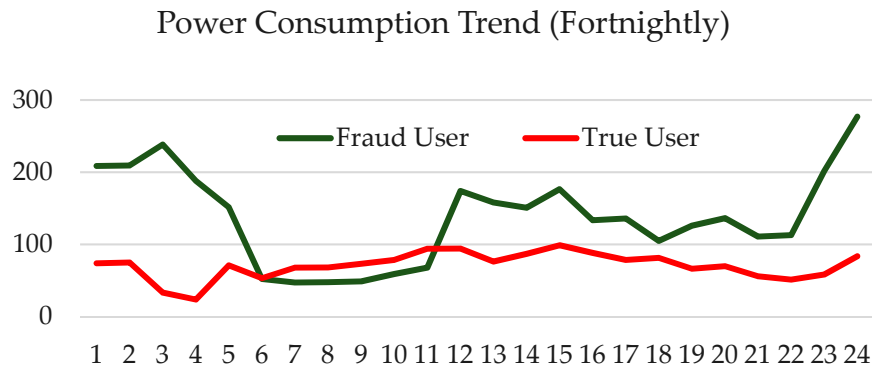


Figure 2. Fortnightly electricity power consumption pattern.

The electricity theft dataset is a typical example of an imbalanced dataset, a dataset where the instances of one class are significantly lower than the other class. The distribution of the two classes, the normal user and the theft user, is presented in Figure 3.

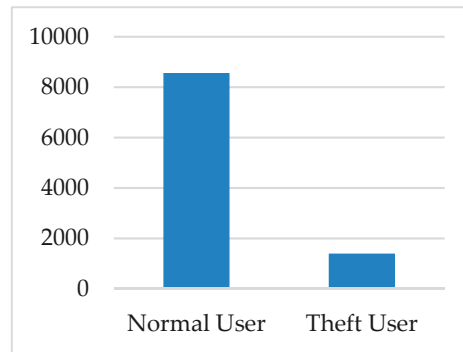


Figure 3. Distribution of normal and theft users in the dataset.

The distribution shows that the number of theft users is very small compared to that of the normal users. This is called a class imbalanced problem. The consequences of using such an imbalanced dataset to train a model are described in the results section. The resulting model can successfully classify only the majority class. This class imbalance problem was addressed by generating synthetic data to increase the minority class. Then, the classification model was trained with the balanced dataset.

2.2. Data Preprocessing

The data preprocessing task was divided into two sections. Initially, the missing data was computed using a preprocessing algorithm, and finally, the generation of synthetic data points was used to solve the class imbalance problem.

2.2.1. Computing the Missing Values

Our proposed preprocessing algorithm utilized the local average value of the consumed power to calculate the missing values. We employed the preprocessing method used in [7]. If there was a missing value or *NaN* value at the position x_i , the value was computed as:

$$f(x_i) = \begin{cases} \sum_{k=i-5}^{i+5} P_k U_k \times Average_{local} & \text{if } x_i = NaN \\ x_i & \text{if } x_i \neq NaN. \end{cases} \quad (1)$$

The local average was given as,

$$Average_{local} = \frac{1}{10} \times \sum_{i=-5}^{i+5} f(x_i). \quad (2)$$

The parameter U_k holds a binary value based on thresholding of the value at entry k . The thresholding was performed as

$$U_k = \begin{cases} 1 & \text{if } x_k \geq Average_{local} \\ 0 & \text{if } x_k < Average_{local}. \end{cases} \quad (3)$$

We experimentally found that the local values have an equal probability of occurrence, and the value of P_k was chosen as 0.10. One special case needed to be tackled, in which there was a continuous entry of *NaN*. Such cases were handled by inserting the row average in those entries before preprocessing was performed.

2.2.2. Generation of Synthetic Data Points

In this dataset, the number of fraudulent observations is significantly lower than non-fraudulent observations, as depicted in Figure 2. A classification model created from such datasets could have a bias towards the majority class. Though the model showed good accuracy, it was prone to misclassify the minority class. This class imbalance problem was counteracted using one of two major approaches: the cost function-based approach and the sampling-based approach. In this paper, we utilized the sampling-based approach. Sampling-based approaches perform under-sampling or oversampling on the imbalanced dataset to reduce the disparity in the amount between the two classes of data. Under-sampling randomly discards the majority class entries to reduce the majority class instances. This technique shrinks the size of the dataset, which is advantageous from a computation perspective, but the random removal may discard potentially significant information, and the remaining data may not be a proper representation of the population. The model developed may produce a less accurate result with the test data.

On the other hand, the over-sampling technique replicated the minority class to increase the number of minority instances. Though no information was lost in this approach, due to the replication of data points, the model developed is likely to suffer from overfitting. By generating synthetic data rather than just replicating the minority class for balancing the dataset, the overfitting problem can be avoided. In this paper, we used the synthetic minority over-sampling technique (SMOTE) to generate synthetic data using minority instances [52]. SMOTE introduces synthetic data points along with the line segments adjoining any or all of the k nearest neighbors of the minority class in the feature space. If (x_1, x_2) is an instance of minority class and if its nearest neighbor is chosen as (x'_1, x'_2) then the data point is synthesized as follows:

$$(X_1, X_2) = (x_1, x_2) + \text{random}(0, 1) \times \Delta, \quad (4)$$

where $\Delta = \{(x'_1 - x_1), (x'_2 - x_2)\}$ and $\text{random}(0, 1)$ provides a random number between 0 and 1.

Using the SMOTE algorithm, the minority class instance was increased to balance the theft dataset. The distribution of two classes after utilizing SMOTE is shown in Figure 4, where the count of normal users and theft users was made equal.

Algorithm 1: Data Pre-processing Algorithm

1. Compute local average:

$$Average_{local} = \frac{1}{10} \times \sum_{i=5}^{i+5} f(x_i)$$

2. Set threshold value:

$$U_k = \begin{cases} 1 & \text{if } x_k \geq Average_{local} \\ 0 & \text{if } x_k < Average_{local} \end{cases}$$

3. Calculate the missing value:

$$f(x_i) = \begin{cases} \sum_{k=i-5}^{i+5} P_k U_k \times Average_{local} & \text{if } x_i = NaN \\ x_i & \text{if } x_i \neq NaN \end{cases}$$

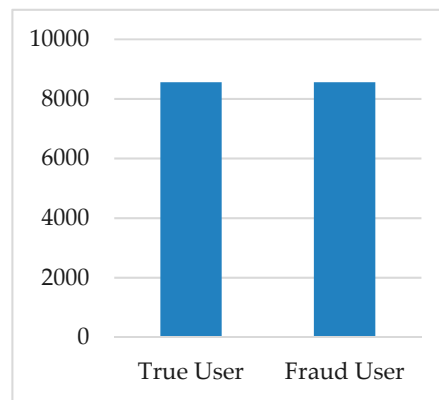


Figure 4. Distribution of normal and theft users after applying synthetic minority over-sampling technique (SMOTE).

2.3. Proposed CNN-LSTM Architecture for Smart Grid Data Classification

In this work, the integration of a convolutional neural network (CNN) and long short-term memory (LSTM) was utilized to solve a classification task. The CNN has an automatic feature extraction ability from the given dataset and LSTM performs better in the case of sequential data. The combination of both has been investigated in different applications, such as text from image or video, sentiment analysis, and natural language processing. In this paper, a CNN-LSTM model was used to solve a binary classification problem. In this study, we utilized 7 hidden layers, where the first 4 hidden layers performed a convolutional operation. Each of the convolutional layers consisted of twenty feature sets. The rest of the hidden layers performed the LSTM operation. The first, second and third LSTM layers consisted of 10, 5 and 100 neurons, respectively.

2.3.1. CNN Model

A CNN is a subclass of neural network proposed by Y. Le Cun et al [53], which is inspired by the working principle of using the human visual cortex for object recognition. CNNs were designed for identifying objects, as well as their classes, in an image. CNN differs from conventional machine learning algorithms in the context of feature extraction, where CNN extracts features globally through a number of stacked layers.

Generally, CNN architecture consists of several convolution layers and pooling layers. These layers are followed by one or more fully connected (FC) layers. The convolutional layer is the

principal building block of a CNN. Convolution is a mathematical operation that acts upon two sets of information. The operation can be addition, multiplication, or a derivative such as

$$\mathbf{y} = \mathbf{x} \times \mathcal{F} \rightarrow \mathbf{y}[\mathbf{i}] = \sum_{j=-\infty}^{+\infty} \mathbf{x}[\mathbf{i} - \mathbf{j}] \mathcal{F}[\mathbf{j}]. \quad (5)$$

In the case of CNNs, the two sets of information are the input data and a convolution filter, which is also called the kernel. The convolutional operation is performed by sliding the kernel over the entire input, which produces a feature map. In practice, different filters are utilized to perform multiple convolutions to produce distinct feature maps. These feature maps are finally integrated to formulate the final output from the convolution layer.

Activation functions are used after the convolution operation to introduce non-linearity to the model. Different activation functions such as linear function, sigmoid, and tanh are used, but the rectified linear unit (ReLU) was used in the proposed CNN since it can train the model faster and ensure near-global weight optimization. The ReLU activation function is defined as follows:

$$f(x_i) = \max(0, x_i). \quad (6)$$

The pooling layer appears next to the convolution layer. This layer down-samples each feature map to reduce their dimensions, which in turn reduces overfitting and training time. The max pooling is widely used in CNNs which just selects the maximum value in the pooling window.

The FC layer is essentially a fully connected artificial neural network. In a nutshell, in a CNN, the convolution and pooling layers extract low-level features such as edges, lines, ears, eyes, and legs, and the fully connected layer performs the classification task based on these low-level features. The activation function used in this final classification layer is typically a SoftMax function, which assigns a probability value to each class which adds up to 1. The SoftMax function is defined as

$$P(y = j | \varphi^{(i)}) = \frac{\mathcal{I}^{\varphi^{(i)}}}{\sum_{j=0}^k \mathcal{I}^{\varphi_k^{(i)}}}. \quad (7)$$

If the weight matrix is denoted as W and the feature matrix by X , then φ in the above equation is generalized as

$$\varphi = \sum_{i=0}^k W_i X_i = W^T X. \quad (8)$$

2.3.2. LSTM

Long short-term memory (LSTM) networks are a special class of recurrent neural networks (RNN) designed to avoid the short-term memory problem of RNNs. LSTMs are capable of remembering and propagating significant information from the initial stages of the network towards the final stage. In this work, we used the fundamental structure of an LSTM, as shown in Figure 5. An LSTM has a similar repetitive structure to that of an RNN, but the modules have different internal components, as seen in Figure 5. The important part of an LSTM is the cell state, which conveys information along the chain. The information in the cell state is dropped or modified by several units called gates. An LSTM module consists of three gates, the forget gate, the input gate, and the output gate.

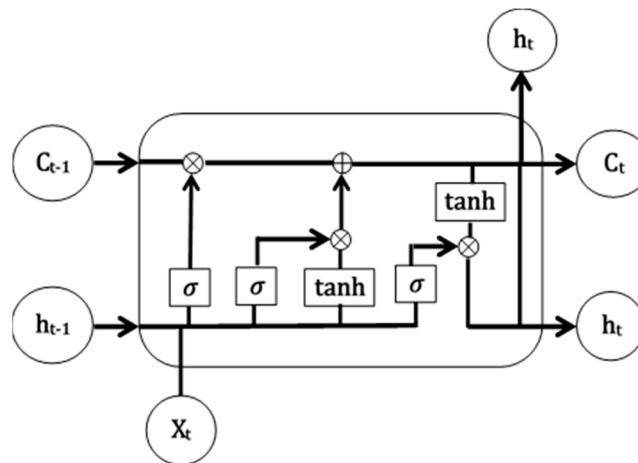


Figure 5. Long short-term memory (LSTM) architecture.

The forget gate consists of a sigmoid layer that takes the previous hidden state (h_{t-1}) and the current input (x_t) to produce an output between 0 and 1. This layer actually decides what information should be kept or discarded. A zero value means to forget the previous information and one means to keep the previous information. The forget gate output is written as

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \quad (9)$$

Later, the forget gate utilizes a sigmoid function and a tanh function to decide what information is going to be added in the cell state. Both the functions take h_{t-1} and x_t as the input. The output of the sigmoid determines whether the current information is important or not, whereas the tanh function regulates the network by squashing the value between -1 and $+1$. Finally, both the outputs are multiplied.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (10)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (11)$$

With the output from the forget gate and input gate, the information in the cell state is updated. It is done through pointwise multiplication of the current cell state and the forget gate output. If f_t is 0, the multiplication will also result in zero, which means total dropout of the previous value. Otherwise, if f_t is 1, it is retained. Later, the pointwise addition updates the cell state as

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t. \quad (12)$$

In the final stage, the output gate determines the final output. This output also acts as the next hidden state, h_t . In this gate, a sigmoid function takes h_{t-1} and x_t as the input and the current cell state C_t is passed through a tanh function. Then, the sigmoid output and the tanh output are multiplied to determine what information the hidden layer is going to carry.

$$\begin{aligned} S_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= S_t \times \tanh(C_t) \end{aligned} \quad (13)$$

Therefore, our proposed wide and deep CNN-LSTM model, depicted in Figure 6 in which the CNN layers precede the LSTM layers, is highly efficient and robust when used for smart grid data classification.

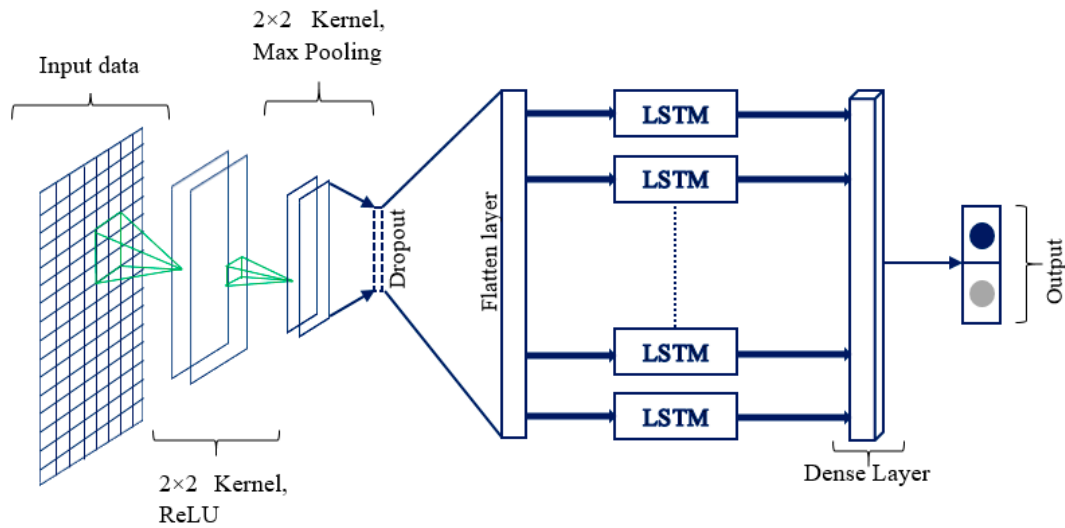


Figure 6. Proposed convolutional neural network based LSTM (CNN-LSTM) model.

3. Results and Discussion

In this section, we verify the efficacy of the proposed CNN-LSTM, with novel missing data treatment for identifying deception users in the benchmark smart grid dataset. To ensure reliability and robustness of the experimental analysis, we used underneath performance matrices.

3.1. Performance Matrices

In this work, the model performance is evaluated using several performance metrics. They are briefly described below.

3.1.1. Binary Cross-Entropy

Cross-entropy is a loss function widely used for classification problems. When P_i indicates the true probability distribution and Q_i is the probability for the classes predicted by a machine learning model, then the cross-entropy is given by

$$\mathcal{H}(P_i, Q_i) = - \sum_i P_i \log Q_i. \quad (14)$$

The cross-entropy decreases toward zero as the prediction becomes more and more accurate. When the classification model is used to classify only two classes, the loss calculation involves only two probabilities. This is called binary cross entropy. Mathematically it is defined as

$$BCE = P_i \log Q_i - (1 - P_i) \log(1 - Q_i). \quad (15)$$

In this work, the BCE was utilized as the cost function to determine how well the model is doing for the training and test datasets.

3.1.2. Matthews Correlation Coefficient (MCC)

This metric is primarily designed for examining the performance of binary classification problems. MCC is a single number that is extracted from the parameters of a confusion matrix. A confusion matrix, presented in Figure 7, is a technique for computing the accuracy of a classification model.

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Figure 7. Confusion matrix.

For binary classification, the Matthews correlation coefficient is given by

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (16)$$

The value of the MCC ranges between -1 and 1 , with 1 being a perfect prediction and -1 a completely incorrect prediction.

3.1.3. F-measure

F-measure, or F1 score, is a useful measure, as compared to accuracy, in cases where the dataset has an imbalanced class distribution. In such a scenario, high accuracy does not imply a robust model. The F1 score is defined as:

$$F1 = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (17)$$

where precision and recall are expressed as follows:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (18)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}. \quad (19)$$

It can be seen that the F1 score encompasses both false positives and false negatives. This implies that the F1 score is a more useful measure than accuracy for any imbalanced dataset.

3.2. Experimental Results

Initially, we trained the model with the weekly, fortnightly, and monthly consumption patterns. The results are presented in Figure 8a–d. The plots in Figure 8 show that the change in the parameter's value remained almost constant but had a slight upward trend. This trend encouraged the use of the whole dataset for training the model.

Here, we present the performance comparison of the proposed model based on the raw dataset and on the transformed dataset after preprocessing and synthetic data augmentation. Table 2 summarizes the performance parameters in these two cases. In both cases, we set the training ratio to 80%. Before adding the synthetic data into the dataset, the number of fraud users was comparatively small with respect to the true users. The overall accuracy is almost the same in both cases. However, due to class imbalance, the model could not classify the fraud users very successfully, which is evident from the values of precision, recall, and F1 score. On the other hand, the inclusion of synthetic data in the dataset improved the model's ability to classify the fraud users.

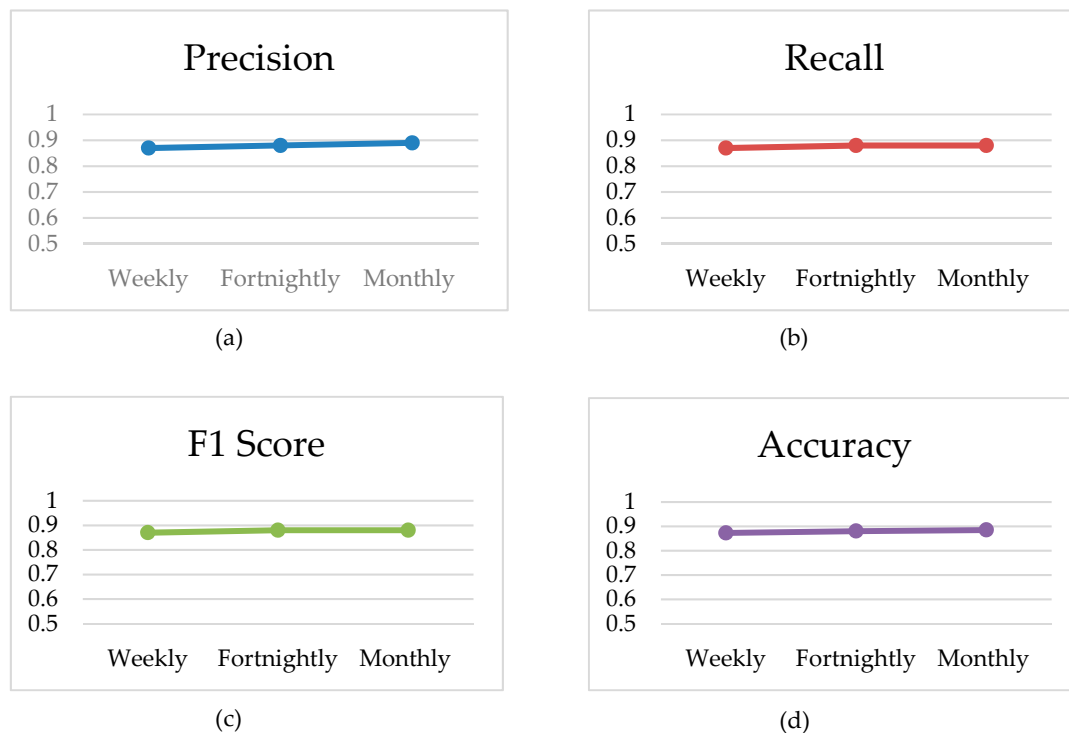


Figure 8. Effect of weekly, fortnightly, and monthly consumption patterns on the model. (a) Precision; (b) Recall; (c) F1 Score; (d) Accuracy.

Table 2. Model performance on the class-imbalanced and balanced datasets.

Parameters		Before applying SMOTE	After Applying SMOTE
Precision	Normal user	0.92	0.90
	Theft user	0.62	0.87
Recall	Normal user	0.96	0.87
	Theft user	0.45	0.91
F1-Score	Normal user	0.94	0.89
	Theft user	0.52	0.89
Overall Accuracy		0.89139	0.8882

Figure 9 shows three other performance parameters plotted against the number of epochs. As the number of epochs was increased, the train loss and accuracy also increased, but the validation loss increased, and accuracy had a downward pattern. This is a case of overfitting of the model, as it was not generalized for the unseen data. Another metric widely accepted for binary classification is the Matthews correlation coefficient (MCC). The MCC is typically used in cases of imbalanced datasets. A value of +1 indicates perfect prediction, but as with the patterns of loss and accuracy, the MCC also decreases for the test set. As shown in Table 2, although the overall accuracies were very similar for both before and after SMOTE, the graphs presented provide contradictory results regarding the accuracy values for both normal users and theft users without SMOTE. This is because even though the accuracy is good, the model does not classify the test data well because of the imbalanced dataset. After implementing the SMOTE to generate synthetic data, the model is trained using the new dataset. The training performance for the balanced dataset can be shown in Figure 10.

The similar parameters, in this case, do not have divergent patterns for the train and test sets. The test accuracy has an unchanged pattern after 400 epochs, which achieved an accuracy rate of 89% at 500 epochs. Further, the test loss seemed to remain constant after the same number of epochs. The

MCC for the test set also attained a value of 0.8, which is indicative of the good prediction capability of the model.

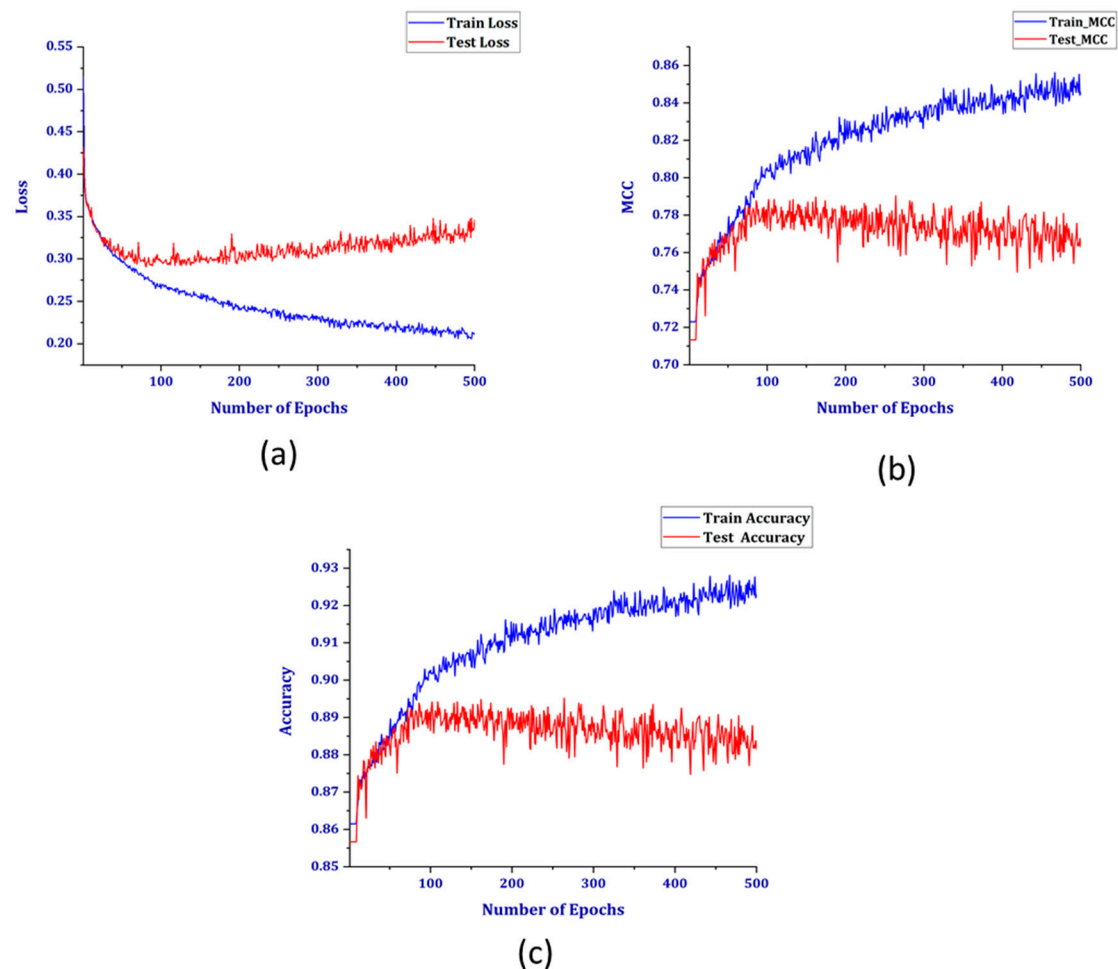


Figure 9. (a) Loss, (b) Matthews correlation coefficient (MCC), and (c) accuracy for the unbalanced dataset.

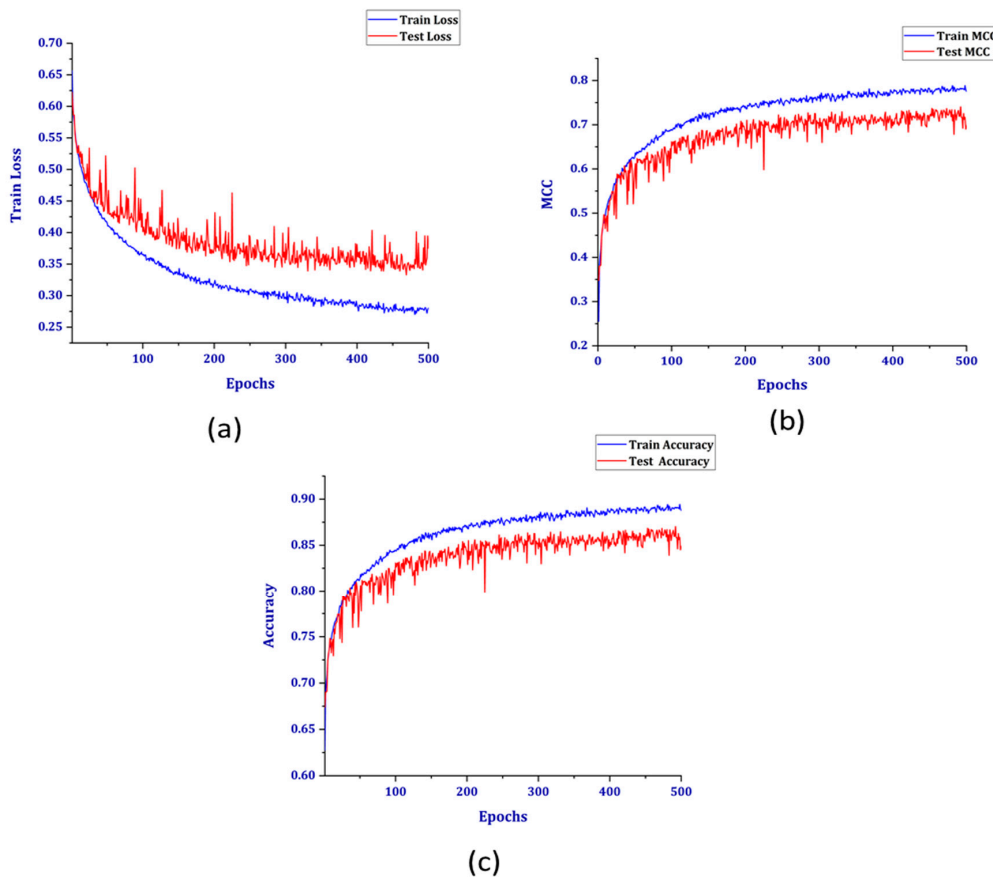
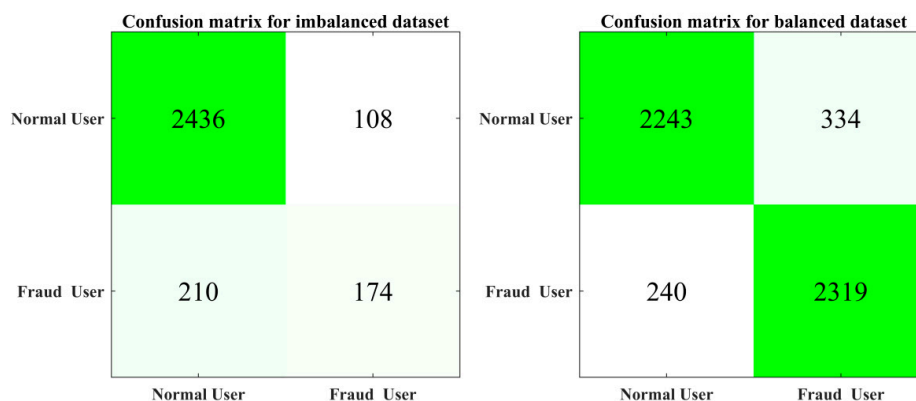
The confusion matrices are compared in Figure 11 for both cases, where the impact of increasing the number of instances in the data set is clearly reflected. In the first scenario, the model had a severe performance loss when classifying fraud users. This is because in the initial dataset were for electricity thieves. Later, after processing, those instances were increased, and the model was able to classify between the minority class and the theft users considerably well. In every instance, 80% of the data were used for training and the remaining 20% were used as validation data. The confusion matrices for the corresponding methods are given in Figure 11.

Performance Comparison

We trained a logistic regression model, the basic model for the binary classification problem, with our dataset. The support vector machine (SVM) algorithm, another popular classification algorithm used in several studies for electricity theft detection, was also used to build the classification model with the same dataset. The results are presented in Table 3.

Table 3. Comparison among LR, Support Vector Machines (SVM), and CNN-LSTM models.

Methods	Precision	Recall	F1-Score	Accuracy
Logistic regression	0.7	0.69	0.69	69%
SVM	0.79	0.79	0.78	79%
CNN-LSTM	0.92	0.96	0.94	89%

**Figure 10.** (a) Loss, (b) MCC, and (c) accuracy for the balanced dataset.**Figure 11.** Confusion matrices for the imbalanced and balanced dataset.

Different machine learning and deep learning algorithms have been investigated for electricity theft detection on various datasets. However, this dataset was not studied using other classification algorithms. Therefore, we present a comparison of several other works, based on different datasets, which were used to perform electricity theft detection. The comparison is presented in Table 4.

Table 4. Summary of models, data sets and performance measures.

#	Reference	Techniques applied	Number of Customers	Accuracy (%)	Precision	Recall	Source
1	[20]	DT coupled SVM	NA	92.50	NA	NA	
2	[34]	Combination functions (SVM, OPF, C4,5 tree)	NA	86.20	54.4	64	Uruguayan Electric Company (UTE)
3	[20]	Regression	30	78	NA	1	Electricity Thefts Surge in Bad Times,” March 16, 2009, USA Today, via Factiva, © 2009 USA Today.
5	[33]	SVM	36176	60	NA	NA	Tenaga Nasional Berhad Distribution (TNBD), Sdn. Bhd.
6	[13]	SVM-based fraud detection model (FDM) with the introduction of a fuzzy inference system (FIS)-SVM-FIS	36176	72	NA	NA	Tenaga Nasional Berhad Distribution (TNBD), Sdn. Bhd.
7	[36]	Genetic- SVM	186,968	62	NA	NA	Tenaga Nasional Berhad Distribution (TNBD),
8	[13]	SVM -fuzzy	36173	72	NA	NA	Tenaga Nasional Berhad Distribution (TNBD)
9	[54]	Fuzzy logic		55	NA	NA	TNB Metering Services database
10	[41]	Fuzzy classification		74.50	NA	NA	brazil
11	[55]	Neural networks (NN)		83.5	24.929.8		Light S.A. Company, Brazil
12	[56]	Neuro-fuzzy	4159	68.2	51.2		Light S.A. Company, Brazil
13	[57]	Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) and Stacked Autoencoder.	12,180	96.9	NA	NA	
14	[7]	Wide and Deep CNN	42372		94.04	NA	State Grid Corporation of China (SGCC)
15	This work	Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM)	17120	89	0.90	0.87	

4. Conclusions

In this work, a robust CNN-LSTM model was investigated for electricity theft detection using historical power consumption data for 10,000 users. The dataset initially had numerous missing data points. A new data preprocessing algorithm was introduced to fill in the missing data rather than just exclude the instances associated with the missing data. Additionally, the dataset had a small number of instances for theft users, similar to other power consumption datasets. Due to this class imbalance characteristic, the model's performance of classifying the theft users was not found satisfactory. Later, the synthetic minority over-sampling technique (SMOTE) was implemented to generate new data points in order to address the class imbalance problem. The model's performance improved after the inclusion of the augmented dataset in the training process. An overall 89% classification accuracy was achieved, which is higher than SVM and logistic regression for the same dataset. This model was also compared with some similar theft detection approaches implemented on different datasets. It showed better performance in terms of accuracy, precision, and recall. There is no time attribute included in this dataset. In the future, we will explore abnormal consumptions based on datasets that have a time attribute and extend our method to real-time energy fraud detection.

Author Contributions: All the authors contributed equally to the conception and ideas, the design of the materials and methods, the analysis and interpretation of the results, and the writing and reviewing of the manuscript.

Funding: This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20192510102510).

Conflicts of Interest: The authors state that there is no conflict of interest.

References

1. Yu, X.; Cecati, C.; Dillon, T.; Simoes, M.G. The new frontier of smart grids. *IEEE Ind. Electron. Mag.* **2011**, *5*, 49–63. [\[CrossRef\]](#)
2. Mavridou, A.; Papa, M. A situational awareness architecture for the smart grid. In *Global Security, Safety and Sustainability & e-Democracy*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 229–236. [\[CrossRef\]](#)
3. Bank, W. *World Development Indicators 2003*; The World Bank: Washington, DC, USA, 2003; Volume 1.
4. Bank, T.W. *Electric Power Transmission and Distribution Losses (% of output)*; IEA: Paris, France, 2016.
5. Zheng, J.; Gao, D.W.; Lin, L. Smart meters in smart grid: An overview. In Proceedings of the 2013 IEEE Green Technologies Conference (GreenTech), Denver, CO, USA, 4–5 April 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 57–64. [\[CrossRef\]](#)
6. Depuru, S.S.S.R.; Wang, L.; Devabhaktuni, V.; Gudi, N. Smart meters for power grid—Challenges, issues, advantages and status. In Proceedings of the 2011 IEEE/PES Power Systems Conference and Exposition, Phoenix, AZ, USA, 20–23 March 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 1–7. [\[CrossRef\]](#)
7. Zheng, Z.; Yang, Y.; Niu, X.; Dai, H.-N.; Zhou, Y. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. *IEEE Trans. Ind. Inform.* **2018**, *14*, 1606–1615. [\[CrossRef\]](#)
8. Agüero, J.R. Improving the efficiency of power distribution systems through technical and non-technical losses reduction. In Proceedings of the PES T&D 2012, Orlando, FL, USA, 7–10 May 2012; pp. 1–8. [\[CrossRef\]](#)
9. McLaughlin, S.; Holbert, B.; Fawaz, A.; Berthier, R.; Zonouz, S. A multi-sensor energy theft detection framework for advanced metering infrastructures. *IEEE J. Sel. Areas Commun.* **2013**, *31*, 1319–1330. [\[CrossRef\]](#)
10. Smith, T.B. Electricity theft: A comparative analysis. *Energy Policy* **2004**, *32*, 2067–2076. [\[CrossRef\]](#)
11. Cabral, J.E.; Gontijo, E.M. Fraud detection in electrical energy consumers using rough sets. In Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583), The Hague, The Netherlands, 10–13 October 2004; IEEE: Piscataway, NJ, USA, 2004; pp. 3625–3629. [\[CrossRef\]](#)
12. Depuru, S.S.S.R.; Wang, L.; Devabhaktuni, V.; Green, R.C. High performance computing for detection of electricity theft. *Int. J. Electr. Power Energy Syst.* **2013**, *47*, 21–30. [\[CrossRef\]](#)
13. Nagi, J.; Yap, K.S.; Tiong, S.K.; Ahmed, S.K.; Nagi, F. Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system. *IEEE Trans. Power Deliv.* **2011**, *26*, 1284–1285. [\[CrossRef\]](#)
14. Bhatia, G.; Gulati, M. *Reforming the Power Sector: Controlling Electricity Theft and Improving Revenue*; The World Bank: Washington, DC, USA, 2004.
15. Abbott, D. Keeping the energy debate clean: How do we supply the world’s energy needs? *Proc. IEEE* **2009**, *98*, 42–66. [\[CrossRef\]](#)
16. Li, B.; Xu, K.; Cui, X.; Wang, Y.; Ai, X.; Wang, Y. Multi-scale DenseNet-based electricity theft detection. In Proceedings of the International Conference on Intelligent Computing, Wuhan, China, 6 July 2018; pp. 172–182. [\[CrossRef\]](#)
17. Lo, C.-H.; Ansari, N. CONSUMER: A novel hybrid intrusion detection system for distribution networks in smart grid. *IEEE Trans. Emerg. Top. Comput.* **2013**, *1*, 33–44. [\[CrossRef\]](#)
18. Xiao, Z.; Xiao, Y.; Du, D.H.-C. Non-repudiation in neighborhood area networks for smart grid. *IEEE Commun. Mag.* **2013**, *51*, 18–26. [\[CrossRef\]](#)
19. Cárdenas, A.A.; Amin, S.; Schwartz, G.; Dong, R.; Sastry, S. A game theory model for electricity theft detection and privacy-aware control in AMI systems. In Proceedings of the 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 1–5 October 2012; pp. 1830–1837. [\[CrossRef\]](#)
20. Jindal, A.; Dua, A.; Kaur, K.; Singh, M.; Kumar, N.; Mishra, S. Decision tree and SVM-based data analytics for theft detection in smart grid. *IEEE Trans. Ind. Inform.* **2016**, *12*, 1005–1016. [\[CrossRef\]](#)
21. Depuru, S.S.S.R.; Wang, L.; Devabhaktuni, V.; Nelapati, P. A hybrid neural network model and encoding technique for enhanced classification of energy consumption data. In Proceedings of the 2011 IEEE Power and Energy Society General Meeting, San Diego, CA, USA, 24–29 July 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 1–8. [\[CrossRef\]](#)
22. Giani, A.; Bitar, E.; Garcia, M.; McQueen, M.; Khargonekar, P.; Poolla, K. Smart grid data integrity attacks: Characterizations and countermeasures π . In Proceedings of the 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), Brussels, Belgium, 17–20 October 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 232–237. [\[CrossRef\]](#)

23. Najmeddine, H.; Drissi, K.E.K.; Pasquier, C.; Faure, C.; Kerroum, K.; Jouannet, T.; Michou, M.; Diop, A. Smart metering by using “Matrix Pencil”. In Proceedings of the 2010 9th International Conference on Environment and Electrical Engineering, Prague, Czech Republic, 16–19 May 2010; pp. 238–241. [\[CrossRef\]](#)
24. Cabral, J.E.; Pinto, J.O.; Pinto, A.M. Fraud detection system for high and low voltage electricity consumers based on data mining. In Proceedings of the 2009 IEEE Power & Energy Society General Meeting, Calgary, AB, Canada, 26–30 July 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1–5. [\[CrossRef\]](#)
25. Ugurlu, U.; Oksuz, I.; Tas, O. Electricity price forecasting using recurrent neural networks. *Energies* **2018**, *11*, 1255. [\[CrossRef\]](#)
26. Bolton, R.J.; Hand, D.J. Statistical fraud detection: A review. *Stat. Sci.* **2002**, *17*, 235–249.
27. Kou, Y.; Lu, C.-T.; Sirwongwattana, S.; Huang, Y.-P. Survey of fraud detection techniques. In Proceedings of the IEEE International Conference on Networking, Sensing and Control, Taipei, Taiwan, 21–23 March 2004; IEEE: Piscataway, NJ, USA, 2004; Volume 2, pp. 749–754. [\[CrossRef\]](#)
28. Nizar, A.; Dong, Z.; Jalaluddin, M.; Raffles, M. Load profiling method in detecting non-technical loss activities in a power utility. In Proceedings of the 2006 IEEE International Power and Energy Conference, Putra Jaya, Malaysia, 28–29 November 2006; IEEE: Piscataway, NJ, USA, 2006; pp. 82–87. [\[CrossRef\]](#)
29. Nizar, A.H.; Dong, Z.Y.; Zhang, P. Detection rules for non technical losses analysis in power utilities. In Proceedings of the 2008 IEEE Power and Energy Society General Meeting—Conversion and Delivery of Electrical Energy in the 21st Century, Pittsburgh, PA, USA, 20–24 July 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 1–8. [\[CrossRef\]](#)
30. Coma-Puig, B.; Carmona, J. Bridging the Gap between Energy Consumption and Distribution through Non-Technical Loss Detection. *Energies* **2019**, *12*, 1748. [\[CrossRef\]](#)
31. Hodge, V.; Austin, J. A survey of outlier detection methodologies. *Artif. Intell. Rev.* **2004**, *22*, 85–126. [\[CrossRef\]](#)
32. Jokar, P.; Arianpoo, N.; Leung, V.C. Electricity theft detection in AMI using customers’ consumption patterns. *IEEE Trans. Smart Grid* **2015**, *7*, 216–226. [\[CrossRef\]](#)
33. Nagi, J.; Mohammad, A.; Yap, K.S.; Tiong, S.K.; Ahmed, S.K. Non-technical loss analysis for detection of electricity theft using support vector machines. In Proceedings of the 2008 IEEE 2nd International Power and Energy Conference, Johor Bahru, Malaysia, 1–3 December 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 907–912. [\[CrossRef\]](#)
34. Di Martino, M.; Decia, F.; Molinelli, J.; Fernández, A. Improving Electric Fraud Detection using Class Imbalance Strategies. In Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM), Vilamoura, Portugal, 6–8 February 2012; pp. 135–141.
35. Nagi, J.; Yap, K.S.; Tiong, S.K.; Ahmed, S.K.; Mohamad, M. Nontechnical loss detection for metered customers in power utility using support vector machines. *IEEE Trans. Power Deliv.* **2009**, *25*, 1162–1171. [\[CrossRef\]](#)
36. Nagi, J.; Yap, K.S.; Tiong, S.K.; Ahmed, S.K.; Mohammad, A. Detection of abnormalities and electricity theft using genetic support vector machines. In Proceedings of the TENCON 2008–2008 IEEE Region 10 Conference, Hyderabad, India, 19–21 November 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 1–6. [\[CrossRef\]](#)
37. Nizar, A.; Dong, Z. Identification and detection of electricity customer behaviour irregularities. In Proceedings of the 2009 IEEE/PES Power Systems Conference and Exposition, Seattle, WA, USA, 15–18 March 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1–10. [\[CrossRef\]](#)
38. Nizar, A.; Dong, Z.; Wang, Y. Power utility nontechnical loss analysis with extreme learning machine method. *IEEE Trans. Power Syst.* **2008**, *23*, 946–955. [\[CrossRef\]](#)
39. Nizar, A.; Dong, Z.; Zhao, J.; Zhang, P. A data mining based NTL analysis method. In Proceedings of the 2007 IEEE Power Engineering Society General Meeting, Tampa, FL, USA, 24–28 June 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 1–8. [\[CrossRef\]](#)
40. Lighari, S.N.; Jensen, B.B.; Shaikh, A.A. Attacks and their defenses for advanced metering infrastructure. In Proceedings of the 2014 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), St. Petersburg, Russia, 6–8 October 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 148–151. [\[CrossRef\]](#)
41. Angelos, E.W.S.; Saavedra, O.R.; Cortés, O.A.C.; de Souza, A.N. Detection and identification of abnormalities in customer consumptions in power distribution systems. *IEEE Trans. Power Deliv.* **2011**, *26*, 2436–2442. [\[CrossRef\]](#)

42. Jiang, R.; Tagaris, H.; Lachsz, A.; Jeffrey, M. Wavelet based feature extraction and multiple classifiers for electricity fraud detection. In Proceedings of the IEEE/PES Transmission and Distribution Conference and Exhibition, Yokohama, Japan, 6–10 October 2002; IEEE: Piscataway, NJ, USA, 2002; Volume 3, pp. 2251–2256. [\[CrossRef\]](#)
43. Chicco, G.; Napoli, R.; Postolache, P.; Scutariu, M.; Toader, C. Customer characterization options for improving the tariff offer. *IEEE Trans. Power Syst.* **2003**, *18*, 381–387. [\[CrossRef\]](#)
44. Gerbec, D.; Gašperič, S.; Šmon, I.; Gubina, F. Determining the load profiles of consumers based on fuzzy logic and probability neural networks. *IEE Proc.-Gener. Transm. Distrib.* **2004**, *151*, 395–400. [\[CrossRef\]](#)
45. Pitt, B.; Kirschen, D.S. Application of data mining techniques to load profiling. In Proceedings of the 21st International Conference on Power Industry Computer Applications. Connecting Utilities. PICA 99. To the Millennium and Beyond (Cat. No. 99CH36351), Santa Clara, CA, USA, 21 May 1999; IEEE: Piscataway, NJ, USA, 1999; pp. 131–136. [\[CrossRef\]](#)
46. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
47. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Mohamed, A.-R.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Kingsbury, B. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal. Process. Mag.* **2012**, *82*–97. [\[CrossRef\]](#)
48. Tian, C.; Ma, J.; Zhang, C.; Zhan, P. A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network. *Energies* **2018**, *11*, 3493. [\[CrossRef\]](#)
49. Yan, K.; Wang, X.; Du, Y.; Jin, N.; Huang, H.; Zhou, H. Multi-Step short-term power consumption forecasting with a hybrid deep learning strategy. *Energies* **2018**, *11*, 3089. [\[CrossRef\]](#)
50. Kuo, P.-H.; Huang, C.-J. An electricity price forecasting model by hybrid structured deep neural networks. *Sustainability* **2018**, *10*, 1280. [\[CrossRef\]](#)
51. Kim, T.-Y.; Cho, S.-B. Predicting the Household Power Consumption Using CNN-LSTM Hybrid Networks. In Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, Madrid, Spain, 21–23 November 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 481–490. [\[CrossRef\]](#)
52. Yang, R.; Zhang, C.; Gao, R.; Zhang, L. A novel feature extraction method with feature selection to identify Golgi-resident protein types from imbalanced data. *Int. J. Mol. Sci.* **2016**, *17*, 218. [\[CrossRef\]](#) [\[PubMed\]](#)
53. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [\[PubMed\]](#)
54. Nagi, J.; Yap, K.S.; Nagi, F.; Tiong, S.K.; Koh, S.; Ahmed, S.K. NTL detection of electricity theft and abnormalities for large power consumers in TNB Malaysia. In Proceedings of the 2010 IEEE Student Conference on Research and Development (SCORED), Putrajaya, Malaysia, 13–14 December 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 202–206. [\[CrossRef\]](#)
55. Muniz, C.; Figueiredo, K.; Vellasco, M.; Chavez, G.; Pacheco, M. Irregularity detection on low tension electric installations by neural network ensembles. In Proceedings of the 2009 International Joint Conference on Neural Networks, Atlanta, GA, USA, 14–19 June 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 2176–2182. [\[CrossRef\]](#)
56. Muniz, C.; Vellasco, M.M.B.R.; Tanscheit, R.; Figueiredo, K. A Neuro-fuzzy System for Fraud Detection in Electricity Distribution. In Proceedings of the IFSA/EUSFLAT Conference, Lisbon, Portugal, 20–24 July 2009; pp. 1096–1101.
57. Bhat, R.R.; Trevizan, R.D.; Sengupta, R.; Li, X.; Bretas, A. Identifying nontechnical power loss via spatial and temporal deep learning. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 272–279. [\[CrossRef\]](#)

