electronics

MDPI

Article

BiGRU-CNN Neural Network Applied to Electric Energy Theft Detection

Lucas Duarte Soares [1], Altamira de Souza Queiroz [2], Gloria P. López [3], Edgar M. Carreño-Franco [1], Jesús M. López-Lezama [4,*] and Nicolás Muñoz-Galeano [4]

[1] Department of Electrical Engineering CECE-UNIOESTE, Foz du Iguaçu 85870-650, Brazil; lucas.2012@alunos.utfpr.edu.br (L.D.S.); edgar.franco@unioeste.br (E.M.C.-F.)
[2] Computer Science Department, Estácio Univerity, Belo Horizonte 14096-160, Brazil; altamira.queiroz@estacio.br
[3] Academic Department of Computational Science—UTFPR, Santa Helena 85892-000, Brazil; gloriap@utfpr.edu.br
[4] Research Group on Efficient Energy Management (GIMEL), Department of Electrical Engineering, Universidad de Antioquia (UdeA), Medellín 050010, Colombia; nicolas.munoz@udea.edu.co
* Correspondence: jmaria.lopez@udea.edu.co; Tel.: +57-4-2198557

Abstract: This paper presents an assessment of the potential behind the BiGRU-CNN artificial neural network to be used as an electric power theft detection tool. The network is based on different architecture layers of the bidirectional gated recurrent unit and convolutional neural network. The use of such a tool with this classification model can help energy sector companies to make decisions regarding theft detection. The BiGRU-CNN artificial neural network singles out consumer units suspected of fraud for later manual inspections. The proposed artificial neural network was programmed in python, using the *keras* package. The best detection model was that of the BiGRU-CNN artificial neural network when compared to multilayer perceptron, recurrent neural network, gated recurrent unit, and long short-term memory networks. Several tests were carried out using data of an actual electricity supplier, showing the effectiveness of the proposed approach. The metric values assigned to their classifications were 0.929 for accuracy, 0.885 for precision, 0.801 for recall, 0.841 for *F1-Score*, and 0.966 for area under the receiver operating characteristic curve.

Keywords: artificial intelligence; machine learning; recurrent neural networks; time series

Citation: Duarte Soares, L.; de Souza Queiroz, A.; López, G.P.; Carreño-Franco, E.M.; López-Lezama, J.M.; Muñoz-Galeano, N. BiGRU-CNN Neural Network Applied to Electric Energy Theft Detection. *Electronics* **2022**, *11*, 693. https://doi.org/10.3390/electronics11050693

Academic Editor: Cheng-Chi Lee

Received: 25 January 2022
Accepted: 22 February 2022
Published: 24 February 2022

1. Introduction

The economic progress of developing countries directly relates to the use of electricity by manufacturing industries. Therefore, the lack of this essential resource significantly impacts the economy at large [1–3]. There might be numerous reasons behind the shortage of electricity availability; the causes are classified as technical and non-technical losses [4]. Technical losses naturally occur due to irradiation and to electrical energy dissipation during its transmission and distribution, which entails losses in dielectrics and especially in electrical conductors by the Joule effect [5]. Non-technical losses, on the other hand, are defined as any energy consumed or any unbilled service due to the failure of measuring equipment or its fraudulent manipulation. These losses are caused by breakdown or illegal handling at the consumer's premises. Non-technical losses are very difficult to predict [6].

For electricity suppliers, the main cause of non-technical losses is the illegal use of electricity by fraudulent customers [7]. This problem has long been one of the main concerns in the energy system management sector, for it can imbalance demand and supply, causing energy network regulation problems and, consequently, drastic profit losses [8]. Theft detection in electricity networks is thus essential to avoid economic loss and mitigate safety risks. However, conventional methods primarily rely on human verification or specific measuring equipment which are extremely slow, expensive, and

inefficient [9]. A considerable number of fraud detection modeling techniques in electrical energy consumption help overcome these obstacles [10]. There are several fraud detection techniques in electricity networks, where classification-based detection is one of the most used approaches. This type of technique mainly distinguishes abnormal energy use patterns from all normal consumption patterns in a test sample containing both normal class and fraudulent class examples [11]. Some algorithms that perform this technique are: k-nearest neighbors [12,13], Support Vector Machine [14,15], Random Forest [16,17], Gradient Boosting [18,19], and Ensemble Learning [20,21].

Classification algorithms such as K-nearest neighbors, support vector machine, decision tree, and logistic regression have already been established in several applications based on electricity-related problems as well as in other research areas [22–24]. However, most of these are based on artificial resources extraction that requires manual intervention and has low electricity theft detection accuracy [25]. It is important to emphasize that all of the above algorithms disregard the data's sequential nature, assuming that they are time-independent [26]. In the real world, however, the opposite happens, given the electricity consumption dynamic behavior [9]. To address these limitations, [27] proposed the use of a widespread deep recurrent neural network based on the detection of electricity theft that can effectively pinpoint cyber-attacks in smart grids. Applied to energy problems, this model explores the nature of the time series of customers' electricity consumption to implement a recurrent neural network of gated recurrent unit (GRU) architecture, thus improving the detection performance and, consequently, better performance simulations results than those of other classic methods.

Reference [28] added non-dominated sorting genetic algorithm to tune the hyperparameters of the GRU network, which explores the nature of the time series of power consumption readings, thereby improving detection performance above classic algorithms.

Recurring neural networks of GRU architecture can be used with other architectures to form hybrid models of electric power fraud detection. Authors in [29] proposed a deep hybrid neural network model based on the combination of GRU and Convolutional Neural Network (CNN) networks and the Particle Swarm Optimization (PSO) algorithm, where the data used was users' real-time electricity consumption. The selection and extraction of resources are performed using the CNN network, which reduce the dimensionality and redundancy present in the time series. The classification of consumption patterns as normal and fraudulent is done using the GRU network with the PSO algorithm. The simulation results show that the proposed model outperforms existing techniques in terms of energy theft detection. Additionally, the proposed model is more robust and accurate than existing classification methods. Reference [30] first used the bidirectional gated recurrent unit (BiGRU) to classify a consumer as honest or fraudulent, using real-time historical series. The experiments showed that this proposed model surpassed traditional classification techniques.

In view of the possibility of using classification algorithms to detect electricity consumption fraud, the present work aims to improve electricity theft detection with a model based on different layers of artificial neural networks called BiGRU-CNN. Most of the times, the process of fraud detection is carried out manually, and it is necessary for the energy companies' employees to collect information on energy consumption for each user. In many cases, this procedure is not efficient. For this reason, the Artificial Intelligence methods proposed in this work become an important alternative solution to the problem of fraud detection, since they allow an efficient exploration of the large amount of information available in the database of the electric power companies. This constitutes the main contribution of our paper.

More accurate classification models can cut costs and add revenue to energy sector companies. The proposed classification of distinct layers BiGRU-CNN was thus compared with the classic artificial neural networks multilayer perceptron (MLP), recurrent neural network (RNN), long short-term memory (LSTM), and GRU to check whether their energy theft classifications are more accurate or not. In this case, the historical series of electrical

energy demand of several consumers of the respective company were used as feed in the fraud detection neural models.

This paper is structured as follows: Section 2 includes the theoretical framework used in the paper. Section 3 is the proposed methodology that includes the data used in this research work, the data pre-processing, the neural networks used for transforming the time series into a supervised machine learning problem, and the comparison of metrics. Section 4 corresponds to the experiments performed during this research work. Section 5 concludes and highlights the most important aspects of the paper.

2. Theoretical Framework

The recurrent neural network (RNN) is an artificial neural network that uses the connection edge of adjacent temporal nodes and introduces the concept of time in the predictive model, making it suitable for processing time series [31]. However, the conventional RNN architecture is susceptible to interference from adjacent time periods, giving rise to the problem of error flow disappearance [32]. One of the alternatives to overcome this is the use of the GRU architecture neural network, which is basically an improved version of LSTM [33]. Generally, both GRU and LSTM networks are suitable for solving the problem of vanishing gradient through their multiplicative ports. However, GRU networks are more efficient in achieving convergence and updating the internal weights during training, in addition to its internal port structure being more succinct than the LSTM network [32].

A typical unit or cell of the GRU architecture network can be constructed from two ports called the reset gate and update gate [34]. The first port (reset gate) filters previously irrelevant information on hidden layers [31]; the lower its value, the greater the amount of information ignored [33]. On the other hand, the second port (update gate) determines the amount of information to be transferred to the output layer [31]; the higher its value, the more information contained in the previous state is used [33].

Figure 1, adapted from [35], shows the structural diagram of a GRU neural network cell governed by Equations (1)–(4), where z_t is the update gate, ρ is the activation sigmoid function, w represent the weights for each input, r_t is the reset gate, $\tilde{h}_t$ is the candidate hidden state of the current hidden node, h_t is the hidden current state, x_t is the current input of the artificial neural network, h_{t-1} is the hidden state of the previous time instant, and u represent the weights for hidden state of the previous time instant [35].

$$z_t = \sigma(w_{zx}x_t + u_{zh}h_{t-1}) \tag{1}$$

$$r_t = \sigma(w_{rx}x_t + u_{rh}h_{t-1}) \tag{2}$$

$$\tilde{h}_t = tan(w_{hx}x_t + r_t \odot u_{hh}h_{t-1}) \tag{3}$$

$$h_t = (1 - z_t) \odot \tilde{h}_t + z_t \odot h_{t-1} \tag{4}$$

For many sequential modeling tasks, it is interesting to access both past and future information. However, the standard GRU neural network processes temporal sequences chronologically and, therefore, is unable to obtain future context information [36]. The bidirectional GRU (Bi-GRU), on the other hand, can perform this operation, since it consists of two standard GRU that process the input sequence in two divergent directions (chronological and anti-chronological) which are subsequently merged into a single variable [37]. This enables the model to explore past and future information. The latter, in turn, can provide more efficient predictive results [36].

Figure 2, adapted from [35], presents a two-intermediate layer Bi-GRU neural network oriented by Equations (5)–(9), where f is the GRU neural network processing, g is the activation function, $\overrightarrow{h}_t^1$ and $\overrightarrow{h}_t^2$ are the output vectors of the forward layers of the first and second layers of the network at time instant t, respectively; w and b are the weight and bias matrices, respectively. On the other hand, vectors $\overleftarrow{h}_t^1$ and $\overleftarrow{h}_t^2$ represent the concurrent outputs of the first and second backward layer of the network t [35].

$$\overrightarrow{h_t^1} = f(w_{x\overrightarrow{h}_1} x_t + w_{\overrightarrow{h^1}\overrightarrow{h^1}} \overrightarrow{h_{t-1}^1} + b_{\overrightarrow{h^1}}) \tag{5}$$

$$\overleftarrow{h_t^1} = f(w_{x\overleftarrow{h}_1} x_t + w_{\overleftarrow{h^1}\overleftarrow{h^1}} \overleftarrow{h_{t+1}^1} + b_{\overleftarrow{h^1}}) \tag{6}$$

$$\overrightarrow{h_t^2} = f(w_{\overrightarrow{h^1}\overrightarrow{h^2}} \overrightarrow{h_t^1} + w_{\overrightarrow{h^2}\overrightarrow{h^2}} \overrightarrow{h_{t-1}^2} + b_{\overrightarrow{h^2}}) \tag{7}$$

$$\overleftarrow{h_t^2} = f(w_{\overleftarrow{h^1}\overleftarrow{h^2}} \overleftarrow{h_t^1} + w_{\overleftarrow{h^2}\overleftarrow{h^2}} \overleftarrow{h_{t+1}^2} + b_{\overleftarrow{h^2}}) \tag{8}$$

$$y_t = g(w_{\overrightarrow{h^2}y} \overrightarrow{h_t^2} + w_{\overleftarrow{h^2}y} \overleftarrow{h_t^2} + b_y) \tag{9}$$

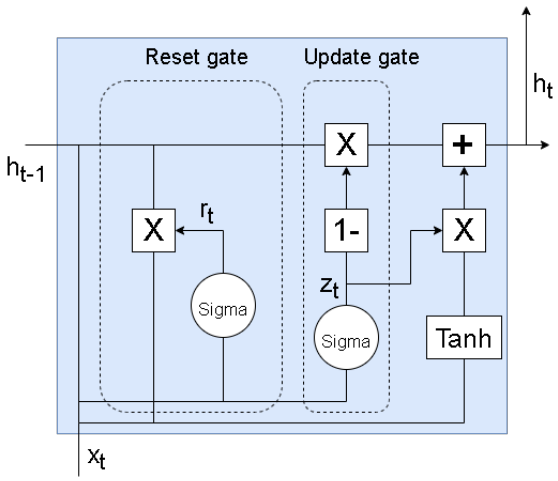

Figure 1. GRU artificial neural network structural unit.

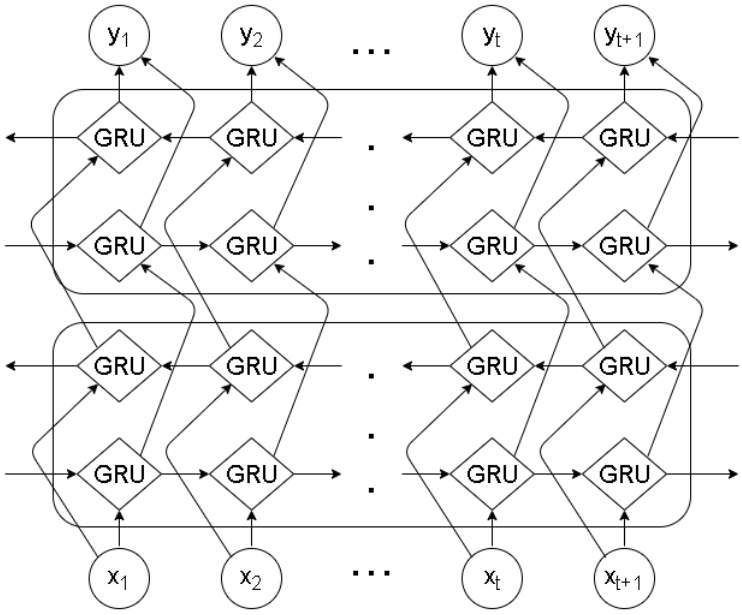

Figure 2. Structural unit of the Bi-GRU neural network.

Unlike recurring networks of GRU and Bi-GRU architecture, CNN is a type of feed-forward network that is not formed by cyclic connections and has no memory as input [34]. Compared to traditional classification methods, CNN can not only map more complex non-linear relationships, but it also has good generalizability [38]. Aside from classifying, CNN networks are widely used for resource extraction through the kernel, which is, in a nutshell, a filter or matrix that slides over the input to perform the convolution operation and produce

a resource map, where different kernels generate different resource maps and all these are merged, thus producing the convolution layer output [39]. CNN architecture neural networks are composed of convolutional layers, pooling layers, and fully connected layers, where convolutional layers and pooling layers are responsible for extracting the electrical energy theft curves characteristics [38]. Figure 3, adapted from [40], presents these layers organized in a generic way to compose the CNN network, and Equations (10) and (11) define their behavior; where x_i is the input of the i-th layer of convolution, y_i is the output of thei-th layer of convolution, $y^{'}$ is the output of the i-th max-pooling layer, f_i is the activation function and, finally, variables b_i and w_i are, respectively, the offset vector and the weights of the i-th convolution layer [38].

$$y_i = f_i(x_i w_i + b_i) \tag{10}$$

$$y^{'} = max(y_{i,j}) \tag{11}$$

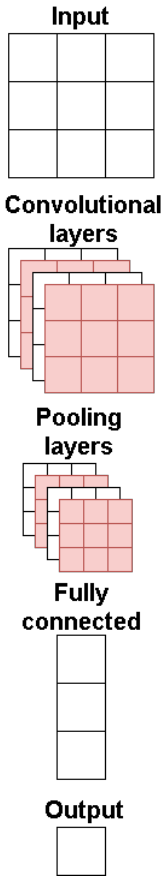

Figure 3. Convolutional neural network structure.

3. Proposed Methodology

The present study proposes a BiGRU-CNN electric energy theft detection model-constructed using a Bi-GRU layer followed by a CNN layer. The input data set for the consumer energy demand historical series was manipulated to feed the Bi-GRU layer that then process it to extract long-term time dependencies. These time-dependent characteristics, which are represented by two hidden state vectors that have past and future information, were introduced into the CNN layer so that significant local relationships are captured through the convolution and pooling layers.

After this procedure, the dataset was structured in several dimensions which were filtered by the flatten layer to become one-dimensional again before being introduced in

the fully connected layer which labels electricity consumers as fraudulent or honest. The structure of the proposed model is shown in Figure 4.

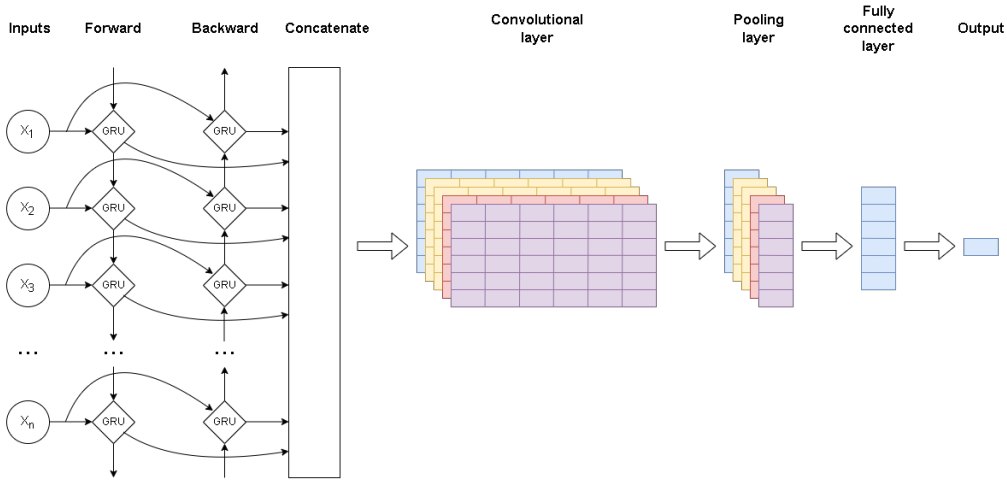

Figure 4. Flowchart of the proposed BiGRU-CNN.

3.1. Data

The database used in this work, also used in the work of [41], was provided by a Colombian electricity supplier which cannot be disclosed due to confidentiality reasons. The data encompass the actual electricity consumption of 462.433 users, where consumption was measured, in kWh, monthly. To complement consumer information, the company also provided a database with manual reviews carried out on all registered customers, as well as anomalies found at the time of such reviews. During the manual inspection, several abnormal consumption patterns were detected, the main causes being clandestine spliced wires, bore meters, a previous connection to the measuring box, and measuring boxes without a security seal. It is worth mentioning that other anomalies were found during checks, although most of them are electric power theft related. Because users' electrical energy consumption pattern is already labeled as fraudulent or normal, neural networks will be trained through supervised learning. The rating provided by the models will be compared with the actual consumer class, making it possible to ascertain greater accuracy and reliability on whether the proposed model is able to correctly label a customer as fraudulent or honest. Evidently, most works on this topic found in the literature lack actual consumer classification given the complex, laborious, expensive, and time-consuming manual fraud checks, as [41] shows. To overcome this obstacle, all consumers are considered honest with fictitious fraud data created to conduct training in a supervised manner. Both ways seemingly create unreliable classification models due to lack of vital data.

3.2. Data Pre-Processing

Initially, the database was cleaned to eliminate incomplete data records and remove irrelevant theft information from the users' consumption curve. This pre-processing step reduced the data from 462.433 to 314.023 users. It is worth mentioning that most of the incomplete data were from "new" customers, people relocating (new rents), or new homes, so it will have been necessary to develop and implement very specific algorithms to fill the missing values. Those customers would present a very atypical load growth during their first years before reaching a steady state. One way to fill the missing data would have been to use existing "similar" information from other customers; however, this will probably lead to populating our data with suppositions. Fortunately, the data set was big enough that even having removing that chunk of data, the statistical impact was very low, so it was decided to simply discard the missing information.

The new slashed database was inserted into a python programming language code on Google Colab. After obtaining the users' electrical energy consumption data by the

program, the MinMaxScaler function of the sklearn pre-processing package was used to normalize the data before feeding it into neural networks. Normalization was needed because energy consumption data varies considerably, potentiality affecting the algorithm's performance during training and thus providing misleading ratings. Equation (12) shows how data normalization was performed by the MinMaxScaler function. In this case, x is the observation of the electrical demand in a time instant.

$$x_{norm} = (x - x_{min})/(x_{max} - x_{min}) \quad (12)$$

The pre-processed dataset contains 314.023 consumers. Of these, 240.774 (76.674%) do not steal electrical energy, they are classified "label 0". The remaining 73.249 consumers (23,326%) do, and they are named "label 1". The pre-processed dataset was fragmented to create the training and test samples. The training sample comprises 80% of all consumers and helps artificial neural networks adjust their respective internal parameters during training. These parameters are evaluated in the test sample, represented by the remaining 20% of total consumers, to check whether they are effective in classifying theft occurrence (label 1) or normal electricity consumption (label 0). The 20–80% ratio for training and test samples is common in similar studies, as seen in [42]. To avoid biased neural networks results, training and test samples have the same proportion of normal and fraudulent consumers as the pre-processed Dataset, i.e., 76,674% and 23,326%. Figure 5 illustrates these sample graphs, as well as their compositions.

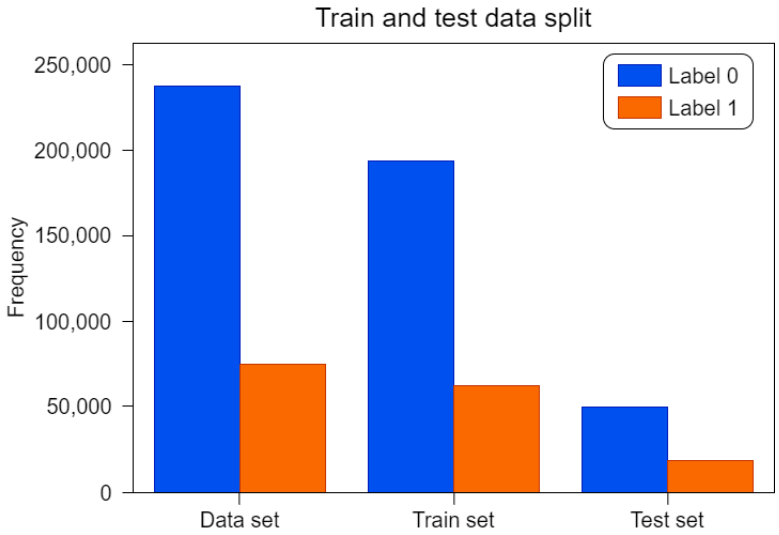

Figure 5. Training and test samples composition.

3.3. Neural Networks

After pre-processing the users' historical electrical energy consumption data, the time series were transformed into a supervised machine learning problem. In other words, a sequence of input and output pairs was created so that a decision could be made and then compared to the correct output. The internal parameters of artificial neural networks are modified during training by the Adam (Adaptive Moment Estimation) algorithm so that the rate of correct network classifications is as high as possible. This algorithm was selected since it has proven to be effective when performing prediction of fraudulent electricity consumption [9]. The ten neurons in each of the two intermediate layers of all predictive models have the rectified linear unit (ReLU) [43], Equation (13), as the activation function, while the only neuron in the output layer has the binary cross-entropy activation function that is responsible for classifying a consumer as fraudulent or honest [44]. Regarding the CNN architecture layers of neural networks, the kernel size quantity was set to 6 and the

number of filters was set to 8. The training of all artificial neural networks was performed in 150 rounds with a batch size of 32.

$$f(x) = max(0, x) \tag{13}$$

3.4. Comparison of Metrics

After network training, neural models were fed by consumer information contained in the test sample to ascertain whether they can provide accurate answers. These predictions were organized in a confusion matrix (Figure 6) to improve the ability to understand each neural network's individual performance regarding accurate classification of fraud or normal electricity consumption.

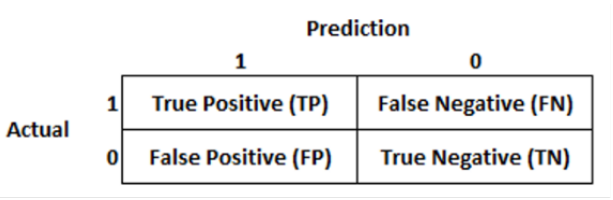

Figure 6. Confusion matrix.

The confusion matrix illustrated in Figure 6 is composed of four classes, where the ordinate axis represents the desired correct response, and the abscissa axis indicates the neural network's forecast. The true positive (*TP*) class encompasses the correct response from the network to the event of interest. In this case, the network is right that a power consumption fraud occurred. On the other hand, the false positive (*FP*) class corresponds to the total of erroneous responses from the network to the event of interest, i.e., the network erroneously predicted a fraud occurrence which was, in fact, normal consumption. The true negative (*TN*) class, on the other hand, comprises the exact classification performed by the neural network regarding the event of no interest, in this case, the network correctly classifies an honest consumer. Finally, the false negative (*FN*) class presents cases in which the network indicated no consumer fraud when, in fact, electricity was stolen.

The confusion matrix comprehensively represents the individual performance of each of the prediction models with regards to fraudulent consumer classification. However, comparing the predictive performances of different models from these matrices is insufficient. Thus, the following metrics are extracted: *Accuracy*, *Precision*, *Recall*, *F1-score*, and ROC AUC. *Accuracy* indicates the number of hits in the neural network, correctly classifying fraudulent and honest consumers. *Precision* is the reason predictions are indeed true when it comes to fraudulent consumers, and all projections cast customers as fraudulent, even when they were not. *Recall*, also known as sensitivity, is the ratio between the assertive forecasts of fraudulent consumers and all consumers who stole electricity.

The weighted average of the precision and recall metric is defined as an *F1-score*. Finally, the ROC AUC is represented by the area under the curve formed by the false positive fraction on the horizontal axis, with the true positive ratio forming the vertical axis. Other than calculating the AUC metric, the ROC curve is also used to define an optimal threshold that can balance the ratio of true positive and false positive. Normally a default threshold is set at 0.5.

All metrics indicate satisfactory results when close to 1 and low predictive results when approaching 0-, which corresponds to the correct classification of fraudulent consumers. Equations (14)–(17) define *Accuracy*, *Precision*, *Recall*, and *F1-score* metrics, respectively. In this case *TP* is true positive, *TN* is true negative, *FP* is false positive, and *FN* is false negative.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{14}$$

$$Precision = TP/(TP + FP) \tag{15}$$

$$Recall = TP/(TP + FN) \tag{16}$$

$$Accuracy = 2/(1/Recall + 1/Precision) \quad (17)$$

4. Tests and Results

After training the neural networks through training samples, the internal parameters were tested to verify their ability to generalize the same results for unprecedented data, which are contained in the test sample. Figures 7–9 depict the confusion matrices for individual results of several artificial neural networks. Additionally, to ease different models' comparison, Table 1 lists metrics based on each matrix's values. Table 1 also indicates the required simulation time used by each network to obtain those metrics.

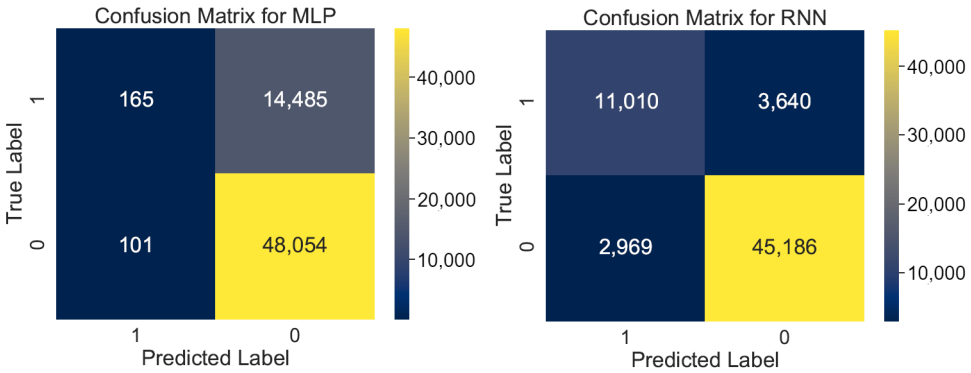

Figure 7. Confusion matrix for MLP and RNN.

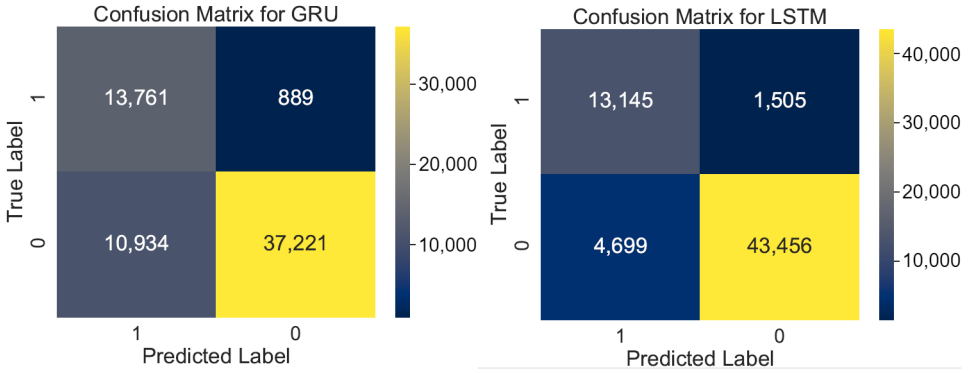

Figure 8. Confusion matrix for GRU and LSTM.

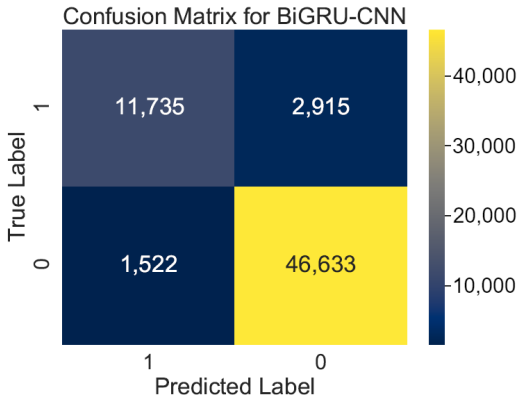

Figure 9. Confusion matrix for GRU and LSTM.

Table 1 shows that the MLP model performed the worst in correctly classifying a consumer as fraudulent or honest when accounting for *Accuracy*, *Recall*, *F1-Score*, and ROC AUC metrics. This is explained by the fact that this network is unable to extract the temporal

dynamic behavior of the energy demand data from different consumers, since its structure does not have information feedback devices. Since it does not process the temporality of the users' energy consumption data, the network has fewer parameters to modify during training and, consequently, lower simulation time than the recurring networks.

Table 1. Comparison of Metrics.

Algorithm	*Accuracy*	*Precision*	*Recall*	*F1-Score*	ROC AUC	Time (min)
MLP	0.767	0.620	0.011	0.022	0.660	45.25
RNN	0.894	0.787	0.751	0.769	0.936	51.17
GRU	0.811	0.557	0.939	0.699	0.912	164.45
LSTM	0.901	0.736	0.897	0.809	0.946	157.47
BiGRU-CNN	0.929	0.885	0.801	0.841	0.966	224.55

The recurring GRU architecture network, on the other hand, had the worst precision metric performance and the best recall metric performance. When the standard GRU was used to create the proposed BiGRU-CNN model, the comparison metrics underwent significant changes. This new network, formed from layers of different architectures with the bidirectional engine, performed best in 4 of the 5 metrics, namely: *Accuracy*, *Precision*, *F1-Score*, and ROC AUC.

These results demonstrate that the procedures performed in the standard GRU have considerably increased its ability to correctly classify a user's consumption as fraudulent or honest. Furthermore, they show their superiority in relation to neural network models. Their simulation time was longer than others, given the larger number of parameters modified during training. Figure 10 shows the ROC curves of all neural models used to calculate the AUC metric in Table 1, as well as the sweet spot that increases true positive while decreasing the false positive ratio. With these coordinates, it is possible to determine the G-mean that is later used to find the optimal threshold, responsible for improving the networks' classification power. The G-mean is given by the square root of the product between the true positive rate (TPR) and the true negative rate (TNR)—the higher its value, the better its predictive ability to classify [17]. Once the optimal threshold is obtained from G-mean, the network can be retrained by setting this new value in the activation function that is responsible for defining whether an energy consumption is fraudulent or not, improving the prediction performance of the network. Table 2 shows the G-mean associated with its ideal threshold.

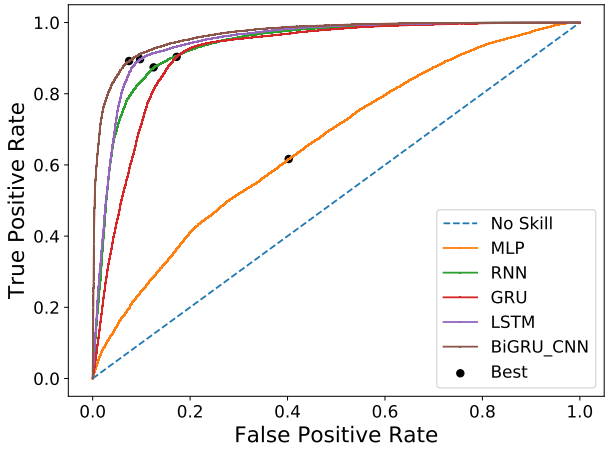

Figure 10. ROC curve of all neural networks.

Analyzing the ROC curves constructed from the classifications of the type of electricity consumption users made by the neural networks, it is apparent that the MLP network

curve is the closest to the curve that indicates an inefficient classification model. Regardless of the chosen threshold, its classification performance will always be the worst when compared to other models capable of processing temporal autocorrelations of electric energy consumption. When observing the recurrent networks, it is easy to see that their performances are similar, and the BiGRU-CNN model superiority is proven by the greater distance to the model without classification ability curve.

Table 2. ROC curve comparison metric of all neural networks.

Algorithm	Best Theshold	G-Mean
MLP	0.236	0.607
RNN	0.186	0.874
GRU	0.605	0.864
LSRM	0.501	0.899
BiGRU-CNN	0.330	0.908

Big data applications will be coming to the power system, bringing large benefits especially on the distribution level; however, the smart metering and communications infrastructure necessary to implement those kinds of algorithms is still far away on the horizon for most utilities across the world. In the meantime, utilities facing the need for automated theft detection today need to rely on their current data, which is basically limited to monthly energy billing information and manual inspections by sample. The tool presented in this paper will help narrow the sample inspections, reducing the overall cost of manual labor and increasing the return of investment on theft detection programs; nonetheless, if there were more data available, for example private information from smart metering infrastructure, and public data such as technical and commercial data from other utilities, local socio-economic data, credit scores, among others, future iterations of this kind of algorithms will benefit the power utility industry overall, and thus the service offered to our customers.

5. Conclusions

The present work proposed the Bigru-CNN model to classify electricity users as fraudulent or honest based on their consumption patterns. This classification is intended to avail energy sector companies making decisions on whether to carry out manual inspections of electricity consuming units.

The experimental results showed that feeding a Bi-GRU layer with the historical series to extract its long-term temporal correlations, and then introducing these time characteristics into a CNN layer so that local trends can be captured, proved to be efficient when comparing *Accuracy*, *Precision*, *F1-Score*, and ROC AUC metrics with MLP, RNN, GRU, and LSTM networks. To ensure that the proposed BiGRU-CNN model is effectively superior to other consumer electricity theft classification models, future work should be carried out altering the hyperparameters of neural networks, as well as the time series of consumers who feed them.

Author Contributions: Conceptualization, L.D.S., A.d.S.Q., E.M.C.-F. and G.P.L.; Data curation, L.D.S., A.d.S.Q., E.M.C.-F. and G.P.L.; Formal analysis, L.D.S., E.M.C.-F., G.P.L., N.M.-G. and J.M.L.-L.; Funding acquisition, J.M.L.-L. and N.M.-G.; Investigation, L.D.S., A.d.S.Q., E.M.C.-F., G.P.L. N.M.-G. and J.M.L.-L.; Methodology, L.D.S., A.d.S.Q., E.M.C.-F., G.P.L., N.M.-G. and J.M.L.-L.; Project administration, E.M.C.-F., G.P.L., N.M.-G. and J.M.L.-L.; Resources, L.D.S., A.d.S.Q., E.M.C.-F., G.P.L., N.M.-G. and J.M.L.-L. Software, L.D.S., E.M.C.-F. and G.P.L.; Supervision, E.M.C.-F., G.P.L., N.M-G. and J.M.L.-L.; Validation, L.D.S., A.d.S.Q., E.M.C.-F. and G.P.L.; Visualization, L.D.S., A.d.S.Q., E.M.C.-F., G.P.L., N.M.-G. and J.M.L.-L.; Writing—original draft, L.D.S., E.M.C.-F. and G.P.L.; Writing—review and editing, E.M.C.-F., G.P.L., N.M.-G. and J.M.L.-L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors gratefully acknowledge the support from the Colombia Scientific Program within the framework of the call Ecosistema Científico (Contract No. FP44842- 218-2018). The authors also want to acknowledge the "estrategia de sostenibilidad" at Universidad de Antioquia in Medellín, Colombia.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sarwar, S.; Chen, W.; Waheed, R. Electricity consumption, oil price and economic growth: Global perspective. *Renew. Sustain. Energy Rev.* **2017**, *76*, 9–18. [CrossRef]
2. López-Lezama, J.M.; Cortina-Gómez, J.; Munoz, N. Assessment of the Electric Grid Interdiction Problem using a nonlinear modeling approach. *Electr. Power Syst. Res.* **2017**, *144*, 243–254. [CrossRef]
3. Agudelo, L.; López-Lezama, J.M.; Muñoz-Galeano, N. Vulnerability assessment of power systems to intentional attacks using a specialized genetic algorithm. *Dyna* **2015**, *82*, 78–84. [CrossRef]
4. Glauner, P.; Meira, J.A.; Valtchev, P.; State, R.; Bettinger, F. The Challenge of Non-Technical Loss Detection Using Artificial Intelligence: A Survey. *Int. J. Comput. Intell. Syst.* **2017**, *10*, 760. [CrossRef]
5. Viegas, J.L.; Esteves, P.R.; Melício, R.; Mendes, V.; Vieira, S.M. Solutions for detection of non-technical losses in the electricity grid: A review. *Renew. Sustain. Energy Rev.* **2017**, *80*, 1256–1268. [CrossRef]
6. Guerrero, J.I.; Monedero, I.; Biscarri, F.; Biscarri, J.; Millán, R.; León, C. Non-Technical Losses Reduction by Improving the Inspections Accuracy in a Power Utility. *IEEE Trans. Power Syst.* **2018**, *33*, 1209–1218. [CrossRef]
7. Ahmad, T. Non-technical loss analysis and prevention using smart meters. *Renew. Sustain. Energy Rev.* **2017**, *72*, 573–589. [CrossRef]
8. Ouyang, Z.; Sun, X.; Yue, D. Hierarchical Time Series Feature Extraction for Power Consumption Anomaly Detection. In *Advanced Computational Methods in Energy, Power, Electric Vehicles, and Their Integration*; Springer: Singapore, 2017; pp. 267–275.
9. Chen, Z.; Meng, D.; Zhang, Y.; Xin, T.; Xiao, D. Electricity Theft Detection Using Deep Bidirectional Recurrent Neural Network. In Proceedings of the 2020 22nd International Conference on Advanced Communication Technology (ICACT), Phoenix Park, Korea, 16–19 February 2020; pp. 401–406. [CrossRef]
10. Ahmad, T.; Chen, H.; Wang, J.; Guo, Y. Review of various modeling techniques for the detection of electricity theft in smart grid environment. *Renew. Sustain. Energy Rev.* **2018**, *82*, 2916–2933. [CrossRef]
11. Jiang, R.; Lu, R.; Wang, Y.; Luo, J.; Shen, C.; Shen, X. Energy-theft detection issues for advanced metering infrastructure in smart grid. *Tsinghua Sci. Technol.* **2014**, *19*, 105–120. [CrossRef]
12. Aziz, S.; Hassan Naqvi, S.Z.; Khan, M.U.; Aslam, T. Electricity Theft Detection using Empirical Mode Decomposition and K-Nearest Neighbors. In Proceedings of the 2020 International Conference on Emerging Trends in Smart Technologies (ICETST), Karachi, Pakistan, 26–27 March 2020; pp. 1–5. [CrossRef]
13. Kong, X.; Zhao, X.; Liu, C.; Li, Q.; Dong, D.; Li, Y. Electricity theft detection in low-voltage stations based on similarity measure and DT-KSVM. *Int. J. Electr. Power Energy Syst.* **2021**, *125*, 106544. [CrossRef]
14. Toma, R.N.; Hasan, M.N.; Nahid, A.A.; Li, B. Electricity Theft Detection to Reduce Non-Technical Loss using Support Vector Machine in Smart Grid. In Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 3–5 May 2019; pp. 1–6. [CrossRef]
15. 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. Informat.* **2016**, *12*, 1005–1016. [CrossRef]
16. Li, S.; Han, Y.; Yao, X.; Yingchen, S.; Wang, J.; Zhao, Q. Electricity Theft Detection in Power Grids with Deep Learning and Random Forests. *J. Electr. Comput. Eng.* **2019**, *2019*, 1–12. [CrossRef]
17. Qu, Z.; Li, H.; Wang, Y.; Zhang, J.; Abu-Siada, A.; Yao, Y. Detection of Electricity Theft Behavior Based on Improved Synthetic Minority Oversampling Technique and Random Forest Classifier. *Energies* **2020**, *13*, 2039. [CrossRef]
18. Punmiya, R.; Choe, S. Energy Theft Detection Using Gradient Boosting Theft Detector With Feature Engineering-Based Preprocessing. *IEEE Trans. Smart Grid* **2019**, *10*, 2326–2329. [CrossRef]
19. Razavi, R.; Gharipour, A.; Fleury, M.; Akpan, I.J. A practical feature-engineering framework for electricity theft detection in smart grids. *Appl. Energy* **2019**, *238*, 481–494. [CrossRef]
20. Gunturi, S.K.; Sarkar, D. Ensemble machine learning models for the detection of energy theft. *Electr. Power Syst. Res.* **2021**, *192*, 106904. [CrossRef]
21. Aslam, Z.; Javaid, N.; Ahmad, A.; Ahmed, A.; Gulfam, S.M. A Combined Deep Learning and Ensemble Learning Methodology to Avoid Electricity Theft in Smart Grids. *Energies* **2020**, *13*, 5599. [CrossRef]
22. Himeur, Y.; Ghanem, K.; Alsalemi, A.; Bensaali, F.; Amira, A. Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. *Appl. Energy* **2021**, *287*, 116601. [CrossRef]
23. Villa-Acevedo, W.M.; López-Lezama, J.M.; Colomé, D.G. Voltage Stability Margin Index Estimation Using a Hybrid Kernel Extreme Learning Machine Approach. *Energies* **2020**, *13*, 857. [CrossRef]
24. Saldarriaga-Zuluaga, S.D.; López-Lezama, J.M.; Muñoz-Galeano, N. Optimal Coordination of Over-Current Relays in Microgrids Using Principal Component Analysis and K-Means. *Appl. Sci.* **2021**, *11*, 7963. [CrossRef]

25. 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. Informat.* **2018**, *14*, 1606–1615. [CrossRef]
26. Pereira, J.; Silveira, M. Unsupervised Anomaly Detection in Energy Time Series Data Using Variational Recurrent Autoencoders with Attention. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 1275–1282. [CrossRef]
27. Nabil, M.; Ismail, M.; Mahmoud, M.; Shahin, M.; Qaraqe, K.; Serpedin, E. Deep Recurrent Electricity Theft Detection in AMI Networks with Random Tuning of Hyper-parameters. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 740–745. [CrossRef]
28. Nabil, M.; Mahmoud, M.; Ismail, M.; Serpedin, E. Deep Recurrent Electricity Theft Detection in AMI Networks with Evolutionary Hyper-Parameter Tuning. In Proceedings of the 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 14–17 July 2019; pp. 1002–1008. [CrossRef]
29. Ullah, A.; Javaid, N.; Samuel, O.; Imran, M.; Shoaib, M. CNN and GRU based Deep Neural Network for Electricity Theft Detection to Secure Smart Grid. In Proceedings of the 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 15–19 June 2020; pp. 1598–1602. [CrossRef]
30. Gul, H.; Javaid, N.; Ullah, I.; Qamar, A.M.; Afzal, M.K.; Joshi, G.P. Detection of Non-Technical Losses Using SOSTLink and Bidirectional Gated Recurrent Unit to Secure Smart Meters. *Appl. Sci.* **2020**, *10*, 3151. [CrossRef]
31. Fengming, Z.; Shufang, L.; Zhimin, G.; Bo, W.; Shiming, T.; Mingming, P. Anomaly detection in smart grid based on encoder-decoder framework with recurrent neural network. *J. China Univ. Posts Telecommun.* **2017**, *24*, 67–73. [CrossRef]
32. Xie, X.; Wang, B.; Wan, T.; Tang, W. Multivariate Abnormal Detection for Industrial Control Systems Using 1D CNN and GRU. *IEEE Access* **2020**, *8*, 88348–88359. [CrossRef]
33. Huo, X.; Wu, K.; Miao, W.; Wang, L.; He, H.; Su, D. Research on Network Traffic Anomaly Detection of Source-Network-Load Industrial Control System Based on GRU-OCSVM. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *300*, 042043. [CrossRef]
34. Lee, K.; Kim, J.K.; Kim, J.; Hur, K.; Kim, H. CNN and GRU combination scheme for Bearing Anomaly Detection in Rotating Machinery Health Monitoring. In Proceedings of the 2018 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII), Jeju, Korea, 23–27 July 2018; pp. 102–105. [CrossRef]
35. Li, P.; Luo, A.; Liu, J.; Wang, Y.; Zhu, J.; Deng, Y.; Zhang, J. Bidirectional Gated Recurrent Unit Neural Network for Chinese Address Element Segmentation. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 635. [CrossRef]
36. Chen, J.X.; Jiang, D.M.; Zhang, Y.N. A Hierarchical Bidirectional GRU Model With Attention for EEG-Based Emotion Classification. *IEEE Access* **2019**, *7*, 118530–118540. [CrossRef]
37. Tao, Q.; Liu, F.; Li, Y.; Sidorov, D. Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU. *IEEE Access* **2019**, *7*, 76690–76698. [CrossRef]
38. Gong, X.; Tang, B.; Zhu, R.; Liao, W.; Song, L. Data Augmentation for Electricity Theft Detection Using Conditional Variational Auto-Encoder. *Energies* **2020**, *13*, 4291. [CrossRef]
39. Madhure, R.U.; Raman, R.; Singh, S.K. CNN-LSTM based Electricity Theft Detector in Advanced Metering Infrastructure. In Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 1–3 July 2020; pp. 1–6. [CrossRef]
40. Kuihua, W.; Jian, W.; Liang, F.; Bo, Y.; Rong, L.; Shenquan, Y.; Ren, Z. An attention-based CNN-LSTM-BiLSTM model for short-term electric load forecasting in integrated energy system. *Int. Trans. Electr. Energy Syst.* **2021**, *31*, e12637. [CrossRef]
41. Queiroz, A.d.S. Algoritmos de Inteligência Computacional Utilizados na Detecção de Fraudes nas Redes de Distribuição de Energia eléTrica. Ph.D. Thesis, Universidade Estadual do Oeste do Paraná, Cascavel, Brasil, 2016.
42. Aldegheishem, A.; Anwar, M.; Javaid, N.; Alrajeh, N.; Shafiq, M.; Ahmed, H. Towards Sustainable Energy Efficiency With Intelligent Electricity Theft Detection in Smart Grids Emphasising Enhanced Neural Networks. *IEEE Access* **2021**, *9*, 25036–25061. [CrossRef]
43. Rouzbahani, H.M.; Karimipour, H.; Lei, L. An Ensemble Deep Convolutional Neural Network Model for Electricity Theft Detection in Smart Grids. In Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 11–14 October 2020; pp. 3637–3642. [CrossRef]
44. Pereira, J.; Saraiva, F. Convolutional neural network applied to detect electricity theft: A comparative study on unbalanced data handling techniques. *Int. J. Electr. Power Energy Syst.* **2021**, *131*, 107085. [CrossRef]