*forecasting*

MDPI

*Article*

# Predictive Data Analytics for Electricity Fraud Detection Using Tuned CNN Ensembler in Smart Grid

**Nasir Ayub [1], Usman Ali [2], Kainat Mustafa [3], Syed Muhammad Mohsin [4,5] and Sheraz Aslam [6,*]**

[1] Faculty of Computing, Department of Software Engineering, Capital University of Science and Technology, Islamabad 44000, Pakistan
[2] Department of Computing, Riphah International University, Faisalabad 45320, Pakistan
[3] Department of Computer Science, Virtual University of Pakistan, Lahore 55150, Pakistan
[4] Department of Computer Science, COMSATS University Islamabad, Islamabad 45550, Pakistan
[5] College of Intellectual Novitiates (COIN), Virtual University of Pakistan, Lahore 55150, Pakistan
[6] Department of Electrical Engineering, Computer Engineering, and Informatics, Cyprus University of Technology, Limassol 3036, Cyprus
* Correspondence: sheraz.aslam@cut.ac.cy

**Abstract:** In the smart grid (SG), user consumption data are increasing very rapidly. Some users consume electricity legally, while others steal it. Electricity theft causes significant damage to power grids, affects power supply efficiency, and reduces utility revenues. This study helps utilities reduce the problems of electricity theft, inefficient electricity monitoring, and abnormal electricity consumption in smart grids. To this end, an electricity theft dataset from the state grid corporation of China (SGCC) is employed and this study develops a novel model, a mixture of convolutional neural network and gated recurrent unit (CNN-GRU), for automatic power theft detection. Moreover, the hyperparameters of the proposed model are tuned using a meta-heuristic method, the cuckoo search (CS) algorithm. The class imbalance problem is solved using the synthetic minority oversampling technique (SMOTE). The clean data are trained and then tested with the proposed classification. Extensive simulations are performed based on real energy consumption data. The simulated results show that the proposed theft detection model (CNN-GRU-CS) solved the theft classification problem better than other approaches in terms of effectiveness and accuracy by 10% on average. The calculated accuracy of the proposed method is 92% and the precision is 94%.

**Keywords:** data analytics; deep learning; electricity fraud detection; optimization; smart grid



**Citation:** Ayub, N.; Ali, U.; Mustafa, K.; Mohsin, S.M.; Aslam, S. Predictive Data Analytics for Electricity Fraud Detection Using Tuned CNN Ensembler in Smart Grid. *Forecasting* **2022**, *4*, 936–948. https://doi.org/10.3390/forecast4040051

Academic Editor: Ted Soubdhan

Received: 21 October 2022
Accepted: 18 November 2022
Published: 21 November 2022



## 1. Introduction

In light of the expensive cost of attaining energy, just as the constrained measure of energy resources, operative, productive, and efficient utilization of power resources is a significant part of financial and social improvement for any state. The Smart Grid (SG) became the only solution to monitor future energy generation and consumption for its efficient utilization [1]. The SG system can be portrayed as a whole electrical network comprising the infrastructure of the power system foundation and computer systems to oversee and screen the power use, alongside an insightful checking system to overview the usage patterns and method of activity of every consumer associated with the system [1–3]. The SG facilitates utility companies and clients to foresee and control energy use by coordinating present-day advanced digital components with the current power system. Through this system, the collective digital devices convey usage readings with the help of the internet to operational centers, and the power transmission organization plays out the billing procedure relying upon these readings. Simultaneously, the operation centers gather client readings from nearly clients' periodical updates through a remote system. The principle target is to minimize loss because of energy wastage and give reasonable, secure, and cheap electricity supplies [4]. The reporting of usage is performed by a Smart

Meter (SM) device. SM is the traditional meter's digital edition. The high-speed processor, nonvolatile capacity, communication services, and ability to keep up across-the-board client energy generation make SMs a significant piece of SG systems.

Today, loss of electrical power has become the most prominent issue influencing both traditional power grids and SGs. The difference between generated and delivered consumer energy is acknowledged as power loss. Advanced SMs get data from users' load devices and calculate energy utilization in hourly periods. The SM gives service provider companies and administrators detailed information to improve monitoring and billing services. It also delivers bi-directional communications between service provider corporations and users [5]. It is also possible with SMs to limit the peak usage amount of electrical energy, which can be used to terminate and remotely reconnect the electrical supply [6].

Specifically, two categories, non-technical loss (NTL) and loss from technical issues (TL), are the major electricity losses. TL happened due to power lines affected by joules and the transformer loss through the transportation of electrical power [7]. It is complex to calculate the TL because finding the point of loss is difficult. It is impossible to stop the TL completely, but it can be minimized with the help of some enhanced techniques throughout the system. The difference between a total loss and TL is described as NTL [8]. The main reasons behind NTL are faulty energy meters, unpaid bills, fraud, and electricity theft [9]. NTL affects both developing and developed nations' economies. For example, every year, the power sector of Brazil and India lost 3 billion US dollars and 44.5 billion US dollars, respectively, due to NTL [10]. Energy theft is a big challenge in the way of strengthening the economy of these countries. Real-time electricity theft recognition is the only solution to this issue.

Electricity theft is one of the many forms of NTL that savagely undermines the revenue of organizations in the energy sector, leading to a huge loss of power assets and harming the economies of all countries. Several techniques have been created to address the issue of power theft. Game theory-based, classification-based, and state-based detection were the three categories into which the researchers divided electricity theft detection (ETD) [11]. In previous studies [12–14], updated devices and sensors gave higher accuracy results for state-based detection. Utilizing the analyzing solutions is primarily constrained by vulnerability, the higher cost of equipment devices, and device upkeep. The authors of [15] described the game theory-based detection framework as the best option for ETD. The main issue of this process is calculating the utility charges among service providers, users, and thieves. Several authors used machine learning techniques for classification-based ETD systems. They analyzed the consumption patterns of electricity using machine learning techniques and established them for classification models such as neural networks [16], decision support systems, decision trees, and support vector machine (SVM) [17].

Deep learning techniques performed well throughout the machine learning-based classification techniques and wildly succeeded in computer vision and image classification areas [18]. It can handle large amounts of data, has better feature selection capability, and has an efficient classification process. These methods are used to build models with SM data from SG. The author in [19] proposed a deep convolutional neural network (CNN) for ETD in SG environments. A combination of CNN and LSTM techniques was used for load forecasting through SG [20]. The hybrid of CNN-LSTM is also proposed for household energy consumption [21] and price forecasting [22]. In these studies, authors used different deep learning and machine learning techniques to perform classification using electrical datasets in SG. For ETD, this paper developed a hybrid model using a CNN and a gated recurrent unit (GRU). Additionally, the CNN-GRU parameters are set using the grey wolf optimization (GWO) method. The main contribution of this research work is described below:

1. Feature correlation and class separation problem is mitigated by oversampling the classes with a smaller observation sample using the SMOTE balancing algorithm.
2. For better training and classification, the CNN is combined with GRU and then tuned with the CS meta-heuristic technique.

3. To detect theft users with no time, computational complexity is reduced by using the optimal parameters of CNN-GRU.
4. Accurate classification is evaluated using performance evaluation metrics and gain classification performance is state of the art.

The remainder of the manuscript is organized as follows. Section 2 gives a detailed discussion of materials and methods. The proposed CNN-GRU-CS classification model is described in Section 3. Section 4 explains simulation results along with relevant detailed discussion. Finally, Section 5 concludes this study.

## 2. Materials and Methods

This paper uses users' energy consumption patterns to detect electricity theft. CNN-GRU-based deep learning technique classifier is used for this purpose. Further, CS is used to optimize the hyperparameters of CNN-GRU. CNN-GRU classification technique is trained by power consumption-related datasets, which consist of both standard and fraudulent users. To train the model, the dataset is prepared by preparing the data with a preprocessing technique. Following the model's introduction, data are fed into the heuristic algorithm CS for tweaking the CNN-GRU classifier's hyperparameters. Finally, the classifier findings are validated using test data. Figure 1 depicts the suggested model.

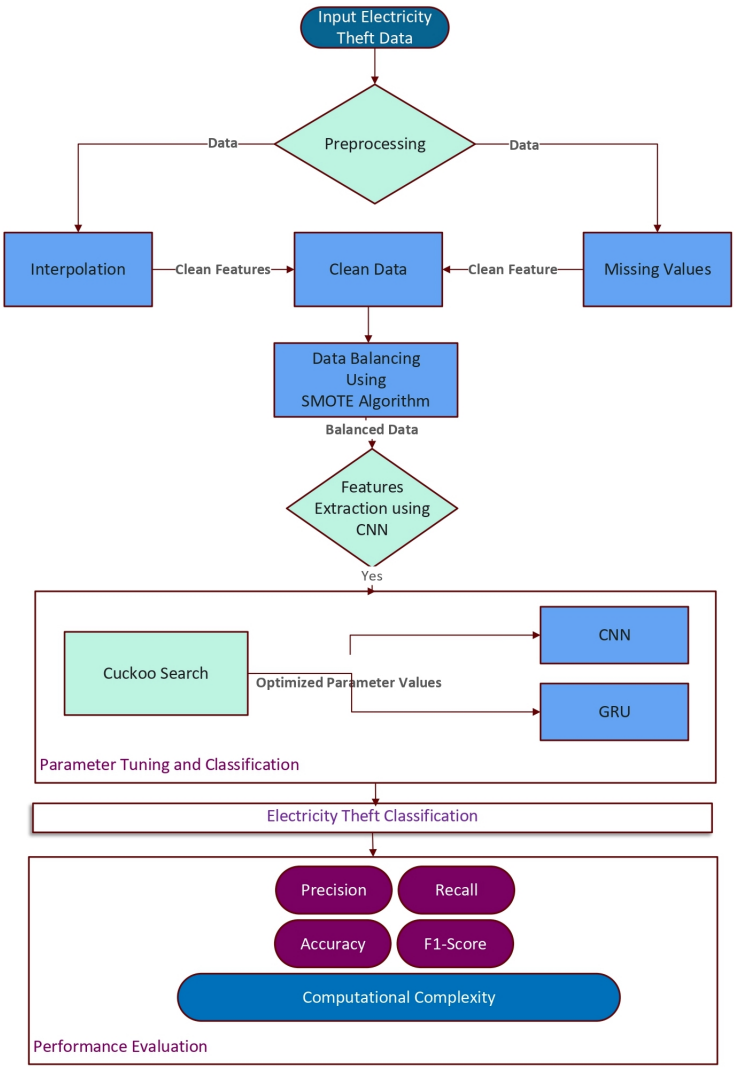


**Figure 1.** Detailed flowchart of System Model.

### *2.1. Electricity Theft Dataset*

This analysis is based on a compilation of authentic customer power use data made available by China's State Grid Corporation [23,24]. In Table 1, dataset information is presented. The dataset consists of the 9655 consumer's power consumption patterns over a 12-month duration. The first observation revealed that regular and fraudulent users have different power consumption patterns, as shown in Figure 2.

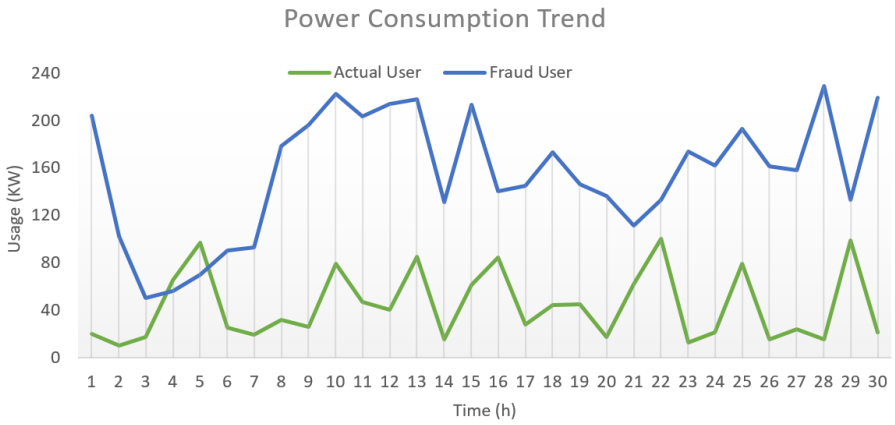


**Figure 2.** Consumption Patterns for Normal and Theft Users.

It shows that both users' habitual consumption patterns and trends are standard and fraudulent. It indicates that there is more fluctuation in the pattern of fraud users than regular users.

**Table 1.** Dataset description.

| Indicator | Value |
|---|---|
| Dataset Period | 2 February 2021 to 2 February 2022 |
| Normal users | 8562 |
| Fraudulent users | 1394 |
| Total users | 9956 |

The data on electricity consumption is commonly obtained through SMs or different sensors at the consumer end. Then, data are accumulated using a data communication system to any central location. Throughout this scenario, there may be a failure in the smart meter, sensors malfunctioning, or faults in transmission lines and storage servers. Due to these reasons, the dataset may have missing or invalid values. Several missing values were found in this dataset, but when these values are discarded, then the dataset could be shrinking. So, this study avoids downsizing using a preprocessing algorithm. The missing values of the dataset are filled using this proposed preprocessing algorithm. The electrical theft dataset is an example of an unbalanced dataset since it comprises two classes, with cases of one class being less than the others. Figure 3 depicts the class distribution.

Figure 3 shows that the number of theft consumers is lower than that of regular clients. This is referred to as a class imbalance problem. The outcomes of training a model with such an unbalanced dataset are presented in the results section. The classification model can only properly classify the more significant class. The imbalance problem in this class was solved by producing synthetic data to extend the minority class. The suggested categorization model was created utilizing balanced data at the moment.

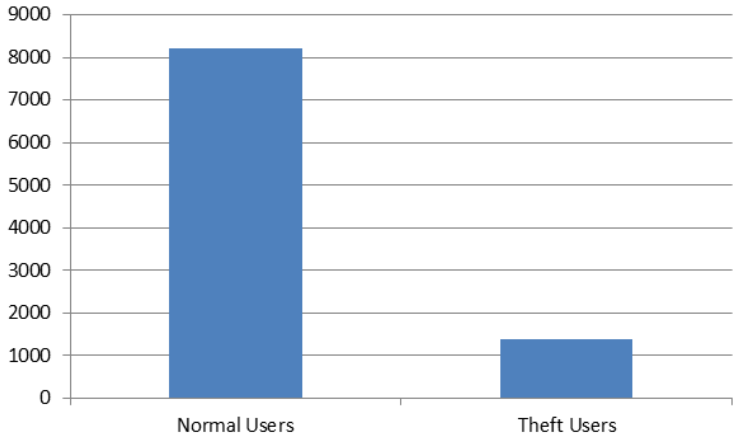


**Figure 3.** Distribution pattern of Normal and Fraudulent customers.

### *2.2. Data Preprocessing*

This part is broken into two sections. To tackle the imbalance problem, the preprocessing procedure is utilized first to calculate the missing values. Then, the missing values are changed with synthetic values in the second phase [25].

#### 2.2.1. Missing Values Computation

The missing values are calculated with the proposed algorithm. The method used the local average power usage values to compute missing data. If there is any missing value at position x, then it will be calculated as in Equation (1):

$$f(X)_i = \sum_{n=i-5}^{i+5} \ldots G_n H_n x_i \text{ Average }_{\text{local}} \quad (1)$$

The local average is presented in Equation (2) as follows:

$$\text{Average }_{\text{local}} = \frac{1}{10} X \sum_{1-5}^{1+5} f(x_i) \quad (2)$$

The binary values of the $H_n$ parameter are based on the entry $n$ threshold value. The thresholding procedure is implemented as follows in Equation (3):

$$H_n = \begin{cases} 1, & \text{if } x_n \geq \text{ Average} \\ 0, & \text{if } x_n < \text{ Average} \\ 0 \text{ local} & \end{cases} \quad (3)$$

The local values are discovered that have the same occurrence probability, and the selected value of $G_n$ is 0.10. One specific instance with a persistent NaN entry should have been handled. Before preprocessing, such situations are addressed by embedding the row average in those items.

#### 2.2.2. Creating Synthetic Data Points

As shown in Figure 3, the number of fraudulent users in the dataset is lower than the number of non-fraudulent users. If the classifier model is created through this dataset, it could have a bias in the direction of the majority class. The model may be accurate, but the minority class could be misclassified. The class imbalance problem may be solved through cost mechanisms and parameter estimation. For this project, a sampling-based technique is employed. This approach solves the imbalanced problem with the under-sampling and oversampling methods. The under-sampling values of the majority class are reduced by randomly discarding the instances to make balance. It shrinks the dataset

size, which is also helpful in reducing computational time. Nonetheless, random values can lose important information, and the remaining portion of the data may be poorly represented after sampling. The produced results may not be highly accurate using this method. However, the oversampling technique increases the instances of the minority with replication of the minority class. Because of the data point replication, no information is lost through this approach, but overfitting may occur in this developed model. Overfitting can be prevented by producing synthetic values rather than duplicating instances.

In this study, the synthetic minority over-sampling method (SMOTE) is utilized to generate synthetic data from minority cases. In features space, SMOTE provides synthetic data over line segments neighboring all or any of the minority class's nearest k neighbors. If the minority class instance is (x1, x2) and the nearest neighbor is $(x_1', x_2') \times 2)$, the data point is synthesized as in Equation (4).

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

where D = $(x_1', x_1), (x_2, x_2')$ and random (0, 1) provides a random number between 0 and 1.

The minority class instances increased using the SMOTE algorithm to make a balance between both classes, i.e., normal and theft users. Figure 4 represents the distribution of both classes after SMOTE algorithm utilization.

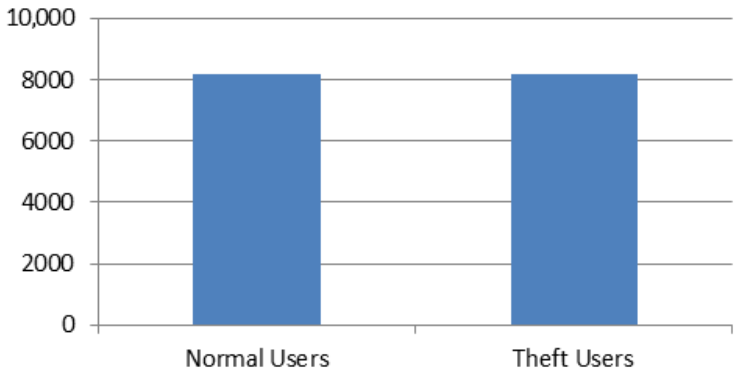


**Figure 4.** Normal and theft users' distribution after SMOTE balancing.

## 3. Proposed CNN-GRU-CS Classification Model

Figure 5 shows the suggested response to electricity fraud detection. The suggested system model primarily consists of five components: preprocessing data, creating data-points using SMOTE, fraud class classification, and validation of data. Initially, this study takes the dataset of electricity theft from SGCC [26]. Electricity data are first preprocessed using normalization, the three-sigma rule, and interpolation of missing values techniques. The following model gets the preprocessed data and uses them to balance the data. The data are balanced using the SMOTE approach. Third, significant features are retrieved from time series data using CNN, and finally, CNN-GRU-CS is provided the significant features for classification. A comparison is performed to compare the performance of proposed model using several performance measures, including accuracy, recall, F1-score, ROC-AUC, and PR-AUC.

### *3.1. CNN*

The CNN technique is introduced by [27] and is part of the neural network family. CNN is trained to identify objects along with their classes in image classification. The only difference between CNN and traditional neural networks is that it can extract the best features with layers. CNN design typically includes many convolutional and pooling layers. Following the CNN layers, there is at least one completely linked layer. CNN's basic building block is the convolutional layer. This layer comprises many different learnable filters or kernels. Convolutional activity is carried out by distributing the kernel throughout the whole input, producing a feature map. Additional filters generate varied feature maps

to play out multiple convolutions. The output of the convolutional layer is formed by integrating these feature maps.

Nonlinearity in the model is represented by activation functions such as ReLU, sigmoid, tanh, and linear. Rectifier linear unit (ReLU) has a better ability to efficiently train the model and also gives assurance optimization of nearly global weights. After the convolutional layer, the pooling layer appears. Through this layer, the overfitting problem and training time can be reduced with the downsampling of every feature map. This paper uses a max-pooling layer in which only maximum values are selected from every feature map. In CNN, fully connected layers take input from one layer and connect it to every other activation unit. The pooling and convolution layers extract low-level properties such as lines and edges. These low-level properties are used for categorization in a fully connected layer. The SoftMax function is commonly used as an activation function in the final classification layer; this function also assigns a probability value to each class.

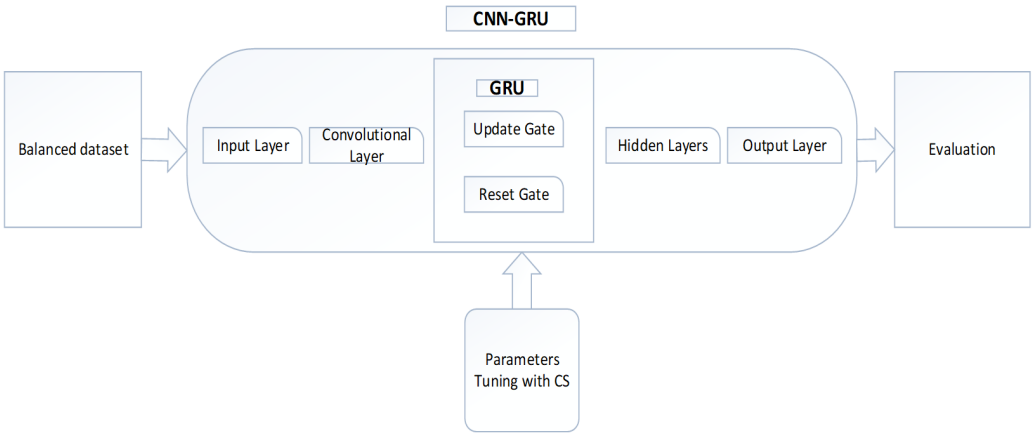


**Figure 5.** CNN-GRU-CS classification model.

### *3.2. GRU*

An artificial neural system appropriate for processing and interpreting time-sequential data is known as a recurrent neural network (RNN). This technique performs well while the output is adjacent to its associate inputs. However, when the interval of time and number of weights increase, input leaves a minor effect on output values because of a problem called the gradient vanishing problem. A particular type of RNN is presented called Gated recurrent unit (GRU) to solve this problem. The update gate manages data that enter the memory, while the reset gate holds data that must flow via output. These two gates are the foundation of the GRU. These gates are the two vectors that decide what information is used to train the model and what kind of irrelevant information must be discarded to increase accuracy.

### *3.3. CS Algorithm*

Cuckoo species that deposit their eggs in the nests of host birds of other species are known as obligate brood parasites, where the concept for cuckoo search originated. It is conceivable for certain host birds and intruding cuckoos to engage in direct conflict. If the host bird discovers the eggs are not its own, it could either throw the foreign eggs away or abandon the nest and build a new one somewhere else. Female parasitic cuckoos of some cuckoo species, such as the New World brood-parasitic Tapera, are frequently quite good at replicating the colors and patterns of the eggs of a small number of host species. CS, a tool that has the potential to address many optimization problems, idealized this type of breeding behavior. Figure 5 also describes CNN-GRU-CS' internal structure [28].

The qualities mentioned earlier allow CS to be represented as three idealized rules:

- Each cuckoo lays one egg at a time, and it is placed in a nest that is picked at random.

- The best nests and greatest eggs (solutions) will be passed down to the following generations.
- The host bird finds the alien egg with a probability of pa[0,1], and the number of available host nests is fixed. The nest is abandoned and a new one is constructed in a different location if the alien egg is found.

The Lévy flights can be integrated using the CS method. Using the D-dimensional vector $x_i = (x_{i1}, x_{i2}, \ldots, x_{id})$ in the ith nest's location is specified, and a Lévy flight is carried out in Equation (5):

$$X_i^{t+1} = x_i^t + a \otimes \text{levy}(\lambda)(i = 1, 2, \ldots, n) \tag{5}$$

$$a = a_0 \otimes \left(x_j^t - x_i^t\right) \pm \tag{6}$$

where $\alpha > 0$ is the step size used to regulate the range of the random search, which should be connected to the scales of the problem of interest, and step size information may be derived by Equation (6). Entry-wise multiplicands are meant by the term "product". Two randomly chosen solutions are $x_{ti}$ and $x_{tj}$. Partial solutions are eliminated, followed by a new solution with the same number of cuckoos. A straightforward power-law equation may represent levy() with the random walk as in Equation (7).

$$\text{lev}\, y(\beta) \sim \mu = t^{-1-\beta}, 0 < \beta \leq 2 \tag{7}$$

where $t$ and $\mu$ are two normal-distributed random values and frequently take the fixed value of 1.5.

$$\text{levy}(\beta) \sim \frac{\phi \times \mu}{|v|^{1/\beta}} \tag{8}$$

$$\phi = \left[\frac{\Gamma(1+\beta) \times \sin\left(\frac{\pi \times \beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\frac{\beta-1}{2}}}\right]^{1/\beta} \tag{9}$$

where the gamma function is described in Equations (8) and (9). A normal distribution with a mean of 0 and a standard deviation of 1 has an infinite variance with an infinite mean. The random numbers $\mu$ and v are taken randomly from this distribution. A cuckoo's repeated leaps and steps create a random walk process that follows a power-law step length distribution with a long tail. Equation (10) produces the new solution $X_i$ in the Lévy flights' random walk component [28].

$$X_{g+1,i} = X_{g,i} + \alpha_0 \frac{\phi \times \mu}{|v|^{1/\beta}} \left(X_{g,i-} X_{g,best}\right) \tag{10}$$

where 0 is a scaling factor and $X_g$ is the best answer. The Lévy distribution is a process of random walking; Lévy flights might abruptly acquire a comparatively greater step size following a succession of smaller steps. At the beginning of the process, the Lévy distribution is used, which aids in breaking out of the local optimum.

$$X_{g+1,i} = X_{g,i} + \alpha_0 \frac{\phi \times \mu}{|v|^{1/\beta}} \left(X_{g,i-} X_{g,best}\right) \tag{11}$$

In Equation (11), $X_t^m$ and $X_t^n$ are the $t_{th}$ generation's random solutions. Between 0 and 1, $r$ creates a random number. The CS algorithm flowchart is shown in Figure 6.

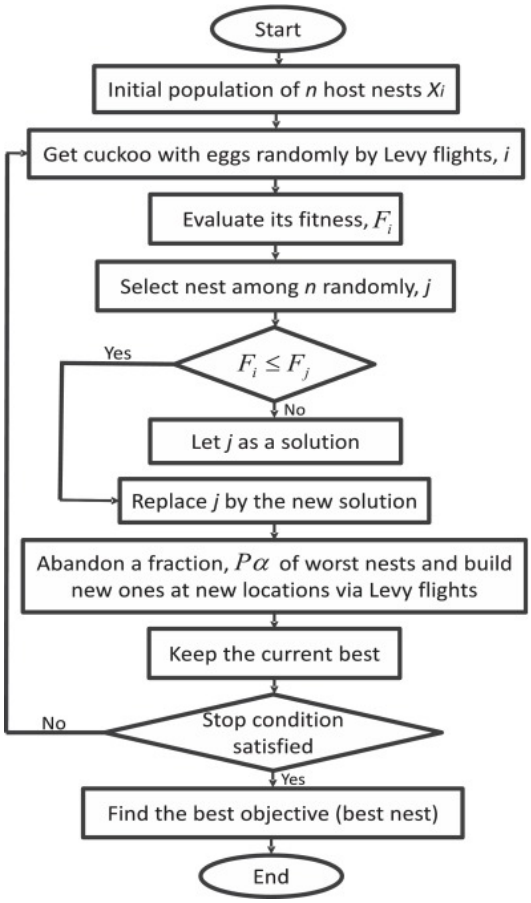


**Figure 6.** Steps of CS Algorithm.

## 4. Results and Discussion

The effectiveness of the suggested CNN-GRU model is validated by comparing several benchmark methods. To assess performance, this study employs a variety of performance measures.

*Experimental Results*

For this study, a CNN-GRU combination is proposed. Furthermore, the parameters of this combination are tweaked using the Cuckoo Search method. The suggested approach is also compared to several benchmark techniques.

The autocorrelation between regular and theft users is presented in Figures 7 and 8, respectively. It can be seen that the data distribution of normal users has a specific usage flow between the orders. Only some consecutive days in a month have high consumption. In the same way, the theft autocorrelation is presented in Figure 8. The distribution of theft has many variations on normal days also. The consumption pattern has too many deviations from the normal variation.

In Figure 9, initially at epoch 0, the train and test accuracy is very low. As the iteration increases, the model gets the best training with the help of optimized parameters. The difference between the train/test is reduced and attains an accuracy of 93%. The proposed algorithm is training by covering all the data points with very minimum error, as seen in Figure 9. After good training, the trained model is tested on unseen data as seen in Figure 10.

Figure 10 illustrates how the model achieves a decent training level as its loss decreases as accuracy rises.

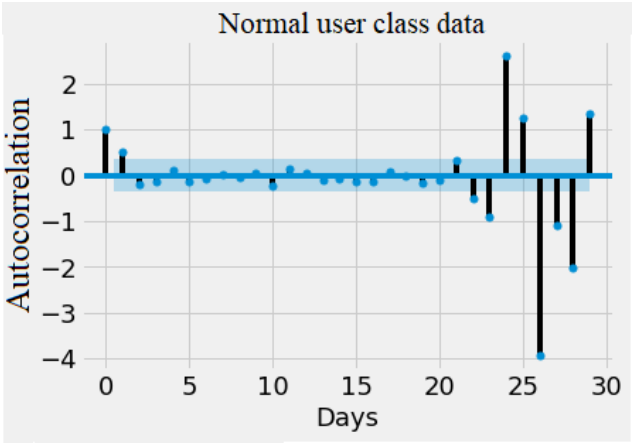


**Figure 7.** Data Distribution of normal users.

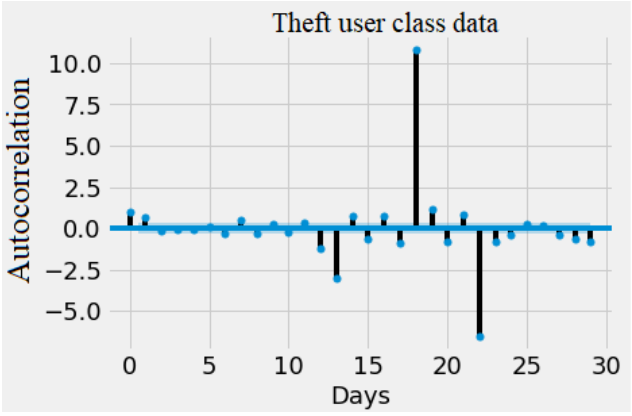


**Figure 8.** Data distribution of theft users.

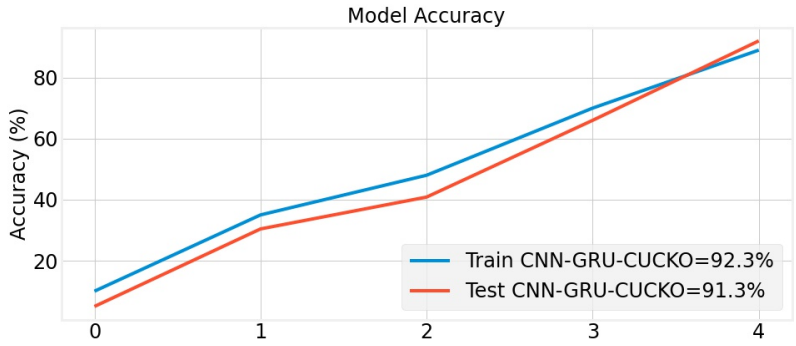


**Figure 9.** Proposed model accuracy vs. epoch.

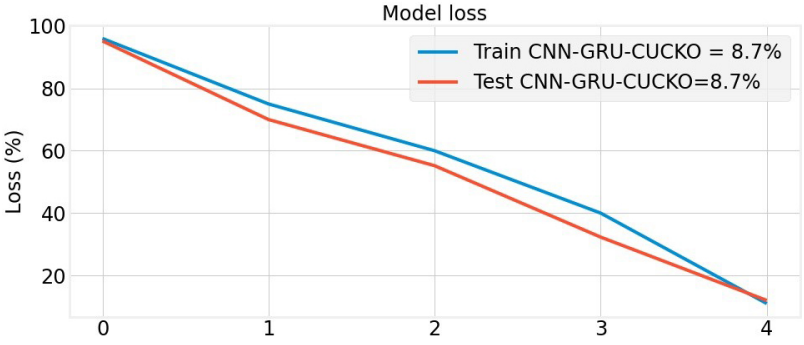


**Figure 10.** Proposed model loss vs. epoch.

False-positive rates and true-positive rates of the proposed technique and conventional techniques are also displayed in Figure 11. Figure 11 shows the ROC curve of the proposed method and state-of-the-art. The proposed method betters classify the data as true and false positive values.

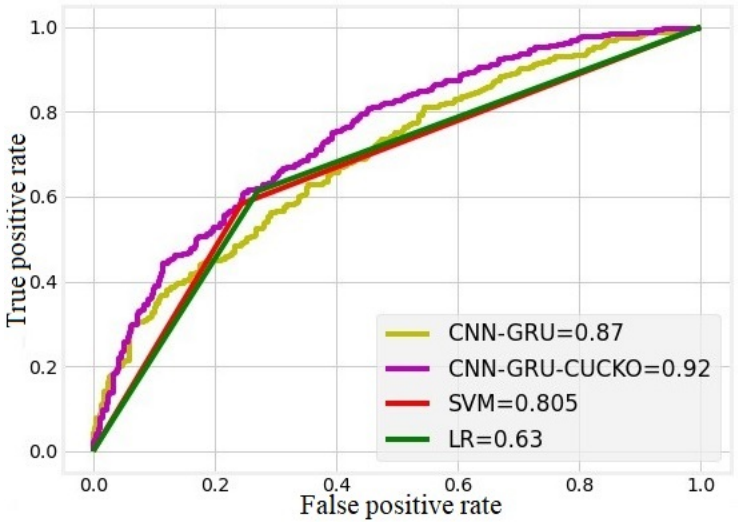


**Figure 11.** ROC curve of proposed and existing methods.

After successful training and testing, the proposed method is validated using performance evaluation metrics as shown in Figure 12. The figure further describes that as the spike of error rate is decreasing, it gives us an achievement of an increase in the values of performance evaluation metrics and true positive rate. The ROC curve shows that the proposed algorithm better classifies the normal and theft user by gaining more true positive rates.

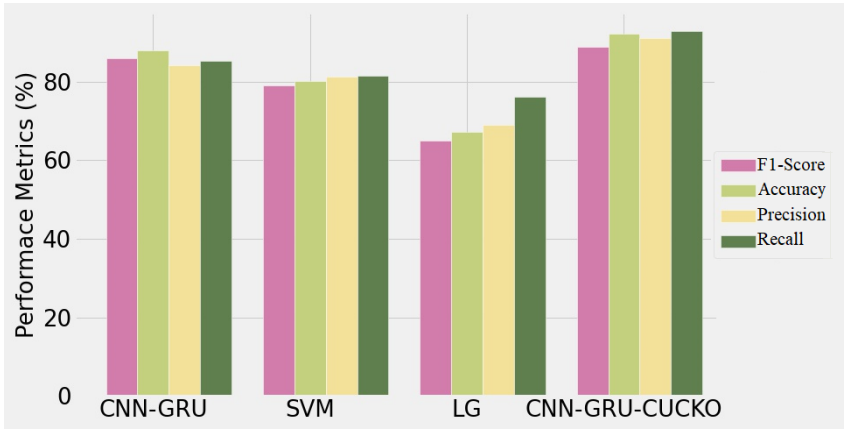


**Figure 12.** Performance evaluation results proposed vs. existing.

The performance evaluation of the suggested model and existing methods is shown in Figure 12. It can be observed that the suggested model's accuracy rate and f1-score (CNN-GRU-CS) are higher, demonstrating the proposed strategy's effectiveness.

## 5. Conclusions

Using historical power data of 10,000 users, this study proposes a strong CNN-GRU model for identifying electrical theft. In this study, a data preprocessing technique is employed to fill in multiple missing data points instead of deleting the missing data points. Furthermore, it was observed in available data that there are fewer stolen users in the sample data than ordinary users. Because of this class imbalance, the model's efficiency in

categorizing thieving users is insufficient. After that, SMOTE is used to generate new data points to correct the class imbalance problem. The proposed model's hyperparameters are tuned using the Cuckoo Search (CS) algorithm. The model performance increased after using the additional dataset in the training phase. Overall, 92.3% classification accuracy is achieved with the proposed electricity theft detection model (CNN-GRU-CS), which is higher than the simulation results of other techniques, namely SVM, LG, and CNN-GRU using the same dataset.

In the future, other optimization algorithms will be used with the proposed method to reduce the time complexity and gain an effective ROC curve.

**Author Contributions:** Conceptualization, N.A.; Methodology, U.A., K.M. and S.M.M.; Software, N.A. and S.M.M.; Validation, S.M.M. and S.A.; Formal analysis, N.A., U.A., K.M. and S.M.M.; Investigation, U.A., K.M. and S.M.M.; Resources, U.A. and S.A.; Data curation, U.A., K.M. and S.M.M.; Writing—original draft, N.A., U.A. and K.M.; Writing—review & editing, K.M. and S.A.; Visualization, S.M.M. and S.A.; Supervision, S.A.; Project administration, N.A. and S.M.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study does not receive any external funding.

**Data Availability Statement:** This analysis is based on a compilation of authentic customer power use data made available by China's State Grid Corporation [23,24].

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| NTL | Non Technical Loss |
| CNN | Convolutional Neural Network |
| CS | Cuckoo Search |
| SG | Smart Grid |
| TL | Technical Loss |
| SVM | Support Vector Machine |
| SM | Smart Meter |
| TL | Technical Loss |
| ETD | Electricity Theft Detection |
| LSTM | Long Short Term Memory |
| GWO | Grey Wolf Optimization |
| GRU | Gated Recurrent Unit |
| SMOTE | Synthetic Minority Over-sampling Technique |
| GWO | Grey Wolf Optimization |
| SGCC | State Grid Corporation of China |

## References

1. Aslam, S.; Herodotou, H.; Mohsin, S.M.; Javaid, N.; Ashraf, N.; Aslam, S. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renew. Sustain. Energy Rev.* **2021**, *144*, 110992. [CrossRef]
2. Fan, D.; Ren, Y.; Feng, Q.; Liu, Y.; Wang, Z.; Lin, J. Restoration of smart grids: Current status, challenges, and opportunities. *Renew. Sustain. Energy Rev.* **2021**, *143*, 110909. [CrossRef]
3. Aurangzeb, K.; Aslam, S.; Haider, S.I.; Mohsin, S.M.; Islam, S.u.; Khattak, H.A.; Shah, S. Energy forecasting using multiheaded convolutional neural networks in efficient renewable energy resources equipped with energy storage system. *Trans. Emerg. Telecommun. Technol.* **2022**, *33*, e3837. [CrossRef]
4. Knayer, T.; Kryvinska, N. An analysis of smart meter technologies for efficient energy management in households and organizations. *Energy Rep.* **2022**, *8*, 4022–4040. [CrossRef]
5. van Dinther, C.; Lau, M.; Terzidis, O. Case Studies in the Smart Grid Sector. In *Smart Grid Economics and Management*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 223–229.
6. Suriyan, K.; Ramalingam, N.; Jayaraman, M.K.; Gunasekaran, R. Recent developments of smart energy networks and challenges. In *Smart Energy and Electric Power Systems*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 37–47. [CrossRef]
7. Chuwa, M.G.; Wang, F. A review of non-technical loss attack models and detection methods in the smart grid. *Electr. Power Syst. Res.* **2021**, *199*, 107415. [CrossRef]

8. Ponnusamy, V.K.; Kasinathan, P.; Madurai Elavarasan, R.; Ramanathan, V.; Anandan, R.K.; Subramaniam, U.; Ghosh, A.; Hossain, E. A Comprehensive Review on Sustainable Aspects of Big Data Analytics for the Smart Grid. *Sustainability* **2021**, *13*, 13322. [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. Firoozi, H.; Mashhadi, H.R. Non-technical loss detection in limited-data low-voltage distribution feeders. *Int. J. Electr. Power Energy Syst.* **2022**, *135*, 107523. [CrossRef]
11. Ahmed, M.; Khan, A.; Ahmed, M.; Tahir, M.; Jeon, G.; Fortino, G.; Piccialli, F. Energy Theft Detection in Smart Grids: Taxonomy, Comparative Analysis, Challenges, and Future Research Directions. *IEEE/CAA J. Autom. Sin.* **2022**, *9*, 578–600. [CrossRef]
12. Akram, R.; Ayub, N.; Khan, I.; Albogamy, F.R.; Rukh, G.; Khan, S.; Shiraz, M.; Rizwan, K. Towards Big Data Electricity Theft Detection Based on Improved RUSBoost Classifiers in Smart Grid. *Energies* **2021**, *14*, 8029. [CrossRef]
13. Javaid, N.; Jan, N.; Javed, M.U. An adaptive synthesis to handle imbalanced big data with deep siamese network for electricity theft detection in smart grids. *J. Parallel Distrib. Comput.* **2021**, *153*, 44–52. [CrossRef]
14. Javaid, N. A PLSTM, AlexNet and ESNN Based Ensemble Learning Model for Detecting Electricity Theft in Smart Grids. *IEEE Access* **2021**, *9*, 162935–162950. [CrossRef]
15. Ahir, R.K.; Chakraborty, B. Pattern-based and context-aware electricity theft detection in smart grid. *Sustain. Energy Grids Netw.* **2022**, *32*, 100833. [CrossRef]
16. Arif, A.; Alghamdi, T.A.; Khan, Z.A.; Javaid, N. Towards efficient energy utilization using big data analytics in smart cities for electricity theft detection. *Big Data Res.* **2022**, *27*, 100285. [CrossRef]
17. Bochie, K.; Gilbert, M.S.; Gantert, L.; Barbosa, M.S.; Medeiros, D.S.; Campista, M.E.M. A survey on deep learning for challenged networks: Applications and trends. *J. Netw. Comput. Appl.* **2021**, *194*, 103213. [CrossRef]
18. 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. [CrossRef]
19. Yao, R.; Wang, N.; Liu, Z.; Chen, P.; Sheng, X. Intrusion detection system in the advanced metering infrastructure: A cross-layer feature-fusion CNN-LSTM-based approach. *Sensors* **2021**, *21*, 626. [CrossRef] [PubMed]
20. Aslam, S.; Ayub, N.; Farooq, U.; Alvi, M.J.; Albogamy, F.R.; Rukh, G.; Haider, S.I.; Azar, A.T.; Bukhsh, R. Towards electric price and load forecasting using cnn-based ensembler in smart grid. *Sustainability* **2021**, *13*, 12653. [CrossRef]
21. Khan, S.; Aslam, S.; Mustafa, I.; Aslam, S. Short-Term Electricity Price Forecasting by Employing Ensemble Empirical Mode Decomposition and Extreme Learning Machine. *Forecasting* **2021**, *3*, 28. [CrossRef]
22. Irfan, M.; Raza, A.; Althobiani, F.; Ayub, N.; Idrees, M.; Ali, Z.; Rizwan, K.; Alwadie, A.S.; Ghonaim, S.M.; Abdushkour, H.; et al. Week Ahead Electricity Power and Price Forecasting Using Improved DenseNet-121 Method. *Comput. Mater. Contin.* **2022**, *72*, 4249–4265. [CrossRef]
23. SGCC. SGCC Electricity Theft Dataset. Available online: https://github.com/henryRDlab/ElectricityTheftDetection (accessed on 18 October 2022).
24. 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.* **2017**, *14*, 1606–1615. [CrossRef]
25. Raja, P.; Thangavel, K. Missing value imputation using unsupervised machine learning techniques. *Soft Comput.* **2020**, *24*, 4361–4392. [CrossRef]
26. Wu, L.; Kong, C.; Hao, X.; Chen, W. A short-term load forecasting method based on GRU-CNN hybrid neural network model. *Math. Probl. Eng.* **2020**, *2020*, 1428104. [CrossRef]
27. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* 2021, *early access*. [CrossRef] [PubMed]
28. Mareli, M.; Twala, B. An adaptive Cuckoo search algorithm for optimisation. *Appl. Comput. Inform.* **2018**, *14*, 107–115. [CrossRef]