International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Electricity Theft Detection in Power Grids with Deep Learning & Random Forests

Tejavath Balakrishna
Computer Science Engineering
(JNTUH)
Sphoorthy Engineering College
Hyderabad, India
balakrishna@sphoorthyEngg.ac.in

Dr.Subba rao Kolavennu
Computer Science and Engineering
(JNTUH)
Sphoorthy Engineering College
Hyderabad, India
profrao99@gmail.com

Maddirala Vishnu Vardhan Reddy
Computer Science Engineering
(JNTUH)
Sphoorthy Engineering College
Hyderabad,India
20N85A0506vishnu@gmail.com

Pasham.Rajesh Goud
Computer Science Engineering
(JNTUH)
Sphoorthy Engineering College
Hyderabad, India
19N81A05A5prg@gmail.com

Bhosle Rohith
Computer Science Engineering
(JNTUH)
Sphoorthy Engineering College
Hyderabad,India
19n81a0593bosley@gmail.com

1.ABSTRACT

Electricity theft is a significant contributor to nontechnical losses in distribution networks, causing considerable harm to power grids. This, in turn, affects the quality of power supply and reduces operating profits for utility companies. To address the challenges of inefficient electricity inspection and irregular power consumption, a new approach called the hybrid CNN-RF model is introduced in this research paper. The model combines a convolutional neural network (CNN) and a random forest (RF) to automatically detect instances of electricity theft. The CNN is designed to learn patterns and features from a large volume of smart meter data, capturing variations across different hours and days through convolution and downsampling operations. To prevent overfitting, a dropout layer is added, and the backpropagation algorithm is utilized to update network parameters during training. The obtained features are then used to train the RF, which determines whether a consumer is stealing electricity. The hybrid model incorporates the grid search algorithm to identify optimal parameters for the RF. Real energy consumption data is employed in experiments, demonstrating that the proposed detection model outperforms other methods in terms of accuracy and efficiency.

Keywords- Conventional Neural Network,Random Forest,Grid Search Algorithm.

2.INTRODUCTION-

Electricity theft is a significant issue that leads to energy losses in power transmission and distribution systems. It affects power companies financially and poses safety risks to the public. Detecting electricity theft accurately is crucial for maintaining a stable and safe power grid.

With the advent of advanced metering infrastructure (AMI) in smart grids, power utilities can collect a vast amount of electricity consumption data from smart meters. However, this also opens up opportunities for new types of electricity theft attacks, including digital tools and cyber attacks. Traditional methods of detecting electricity theft, such as manual examination and verification of meters, are time-consuming and costly.Various approaches have been proposed to address this problem. State-based detection relies on specialized devices and real-time system measurements, but these may not always be feasible. Game-theory-based schemes involve modeling the interactions between utility companies and thieves, but formulating utility functions for all stakeholders remains a challenge. Artificial intelligence-based methods, including machine learning and deep learning, have shown promise but still have room for improvement.Deep learning techniques, such as convolutional neural networks (CNNs), have been explored for feature extraction from smart meter data. However, the performance of these detectors needs further evaluation,especially compared to traditional architectures. Combining CNNs with random forest (RF) classifiers has shown potential as a more effective approach for electricity theft detection. The CNN captures relevant features from smart meter data, while the RF classifier enhances the detection process by leveraging bagging and random feature selection techniques.This novel CNN-RF model has been trained and tested using real data from electricity utility customers in Ireland and London, demonstrating its effectiveness in detecting electricity theft.

3. LITERATURE SURVEY-

The research paper focuses on addressing the issue of electricity theft in power grids, which results in significant financial losses and disruptions in power supply. To tackle this problem effectively, the author proposes a novel approach that combines Convolutional Neural Networks (CNN) and Random Forest algorithms.

By leveraging the power of CNNs, the model can learn and extract relevant features from the power consumption data. These features capture patterns and abnormalities that indicate potential instances of theft. The CNN is trained to distinguish between periods of normal energy usage and those associated with theft, assigning a label of 0 or 1 accordingly.To further enhance the detection accuracy, the author integrates the CNN with Random Forest, a versatile and robust machine learning algorithm. The Random Forest algorithm utilizes an ensemble of decision trees, which collectively make predictions based on the extracted features from the CNN. This combination improves the overall prediction accuracy compared to using either algorithm individually.The proposed model goes through a training phase using labeled data, where patterns of energy theft and normal energy usage are learned. The model is then evaluated and tested on real-world energy consumption data to assess its effectiveness.

The results demonstrate that the combined CNN-Random Forest model outperforms traditional algorithms in accurately identifying instances of electricity theft. This innovative approach offers a more efficient and reliable method for utility companies to detect and mitigate energy theft, thereby minimizing financial losses and ensuring a stable power supply.

4. IMPLEMENTATION-

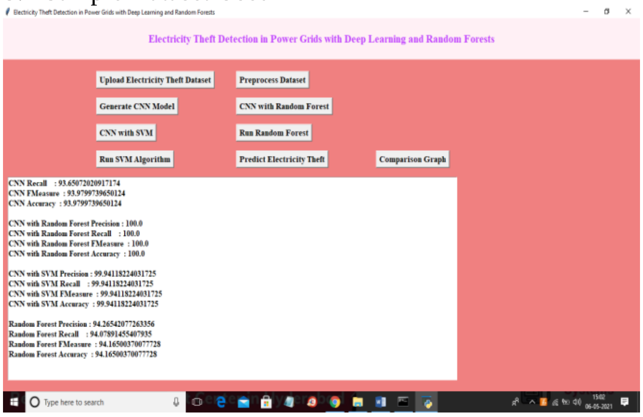
Here's a high-level overview of the implementation process:

1. Data Collection: Gather a dataset of power consumption records from smart meters or other sources. Include both normal energy usage samples and instances of known electricity theft.
2. Data Preprocessing: Clean and preprocess the collected data. This may involve handling missing values, removing outliers, and normalizing the data.
3. Feature Extraction: Utilize deep learning techniques, such as Convolutional Neural Networks (CNN), to extract meaningful features from the power consumption data. Train the CNN model to learn patterns and characteristics associated with electricity theft.
4. Labeling and Dataset Preparation: Assign labels to the dataset indicating whether each instance represents normal energy usage or electricity theft. Split the dataset into
5. Random Forest Training: Build a Random Forest model using the extracted features as input and the labeled dataset for training. Fine-tune the Random Forest model using appropriate hyperparameter tuning techniques.
6. Model Evaluation: Evaluate the performance of the combined CNN-Random Forest model using appropriate evaluation metrics such as accuracy, precision, recall, and F1 score. Use the testing dataset to assess the model's ability to detect electricity theft accurately.
7. Model Deployment: Once the model demonstrates satisfactory performance, integrate it into a software system or application that can process real-time or near real-time power consumption data. This system should provide timely alerts or notifications whenever instances of electricity theft are detected.

It's important to note that implementing the project requires proficiency in programming languages such as Python and familiarity with deep learning frameworks like TensorFlow or PyTorch for CNN implementation. Additionally, you will need knowledge of machine learning libraries like scikit-learn for Random Forest implementation.

5. Results-

5.1 Sample Dataset Used

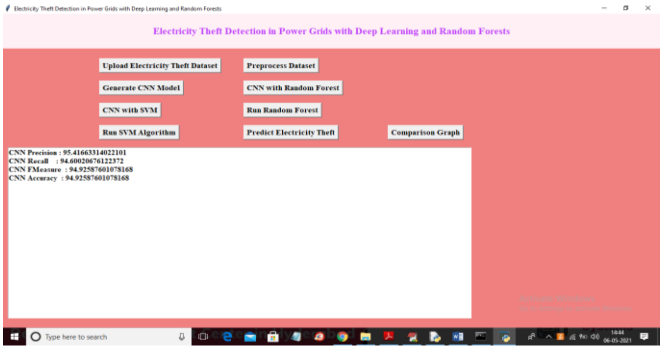

Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
Upload Electricity Theft Dataset
Preprocess Dataset
Generate CNN Model
CNN with Random Forest
CNN with SVM
Run Random Forest
Run SVM Algorithm
Predict Electricity Theft
Comparison Graph
Type here to search
15:02
06-05-2021

```
CNN Recall    : 93.65072020917174
CNN FMeasure  : 93.9799739650124
CNN Accuracy  : 93.9799739650124

CNN with Random Forest Precision : 100.0
CNN with Random Forest Recall    : 100.0
CNN with Random Forest FMeasure  : 100.0
CNN with Random Forest Accuracy  : 100.0

CNN with SVM Precision : 99.94118224031725
CNN with SVM Recall    : 99.94118224031725
CNN with SVM FMeasure  : 99.94118224031725
CNN with SVM Accuracy  : 99.94118224031725

Random Forest Precision : 94.26542077263356
Random Forest Recall    : 94.07891455407935
Random Forest FMeasure  : 94.16500370077728
Random Forest Accuracy  : 94.16500370077728
```

training and testing sets.

learning,so we need to convert to numeric by assigning integer ID so click on 'Preprocess Dataset' button to clean data

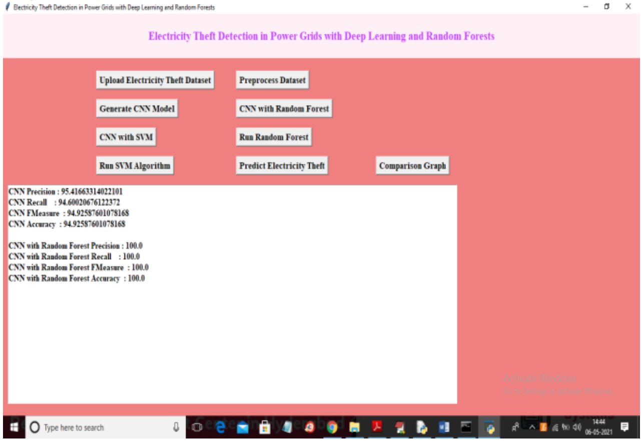
5.3 Train CNN with dataset



```
CNN Precision : 95.41663314022101
CNN Recall    : 94.60020676122372
CNN FMeasure  : 94.92587601078168
CNN Accuracy  : 94.92587601078168
```

Fig3.In above screen with normal CNN we got 94% accuracy and now click on 'CNN with Random Forest' button to train CNN with RF

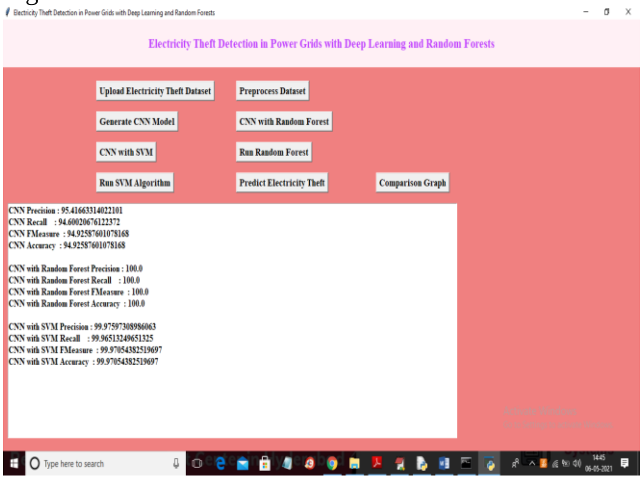
Fig 5.4 Train CNN with Random Forest.



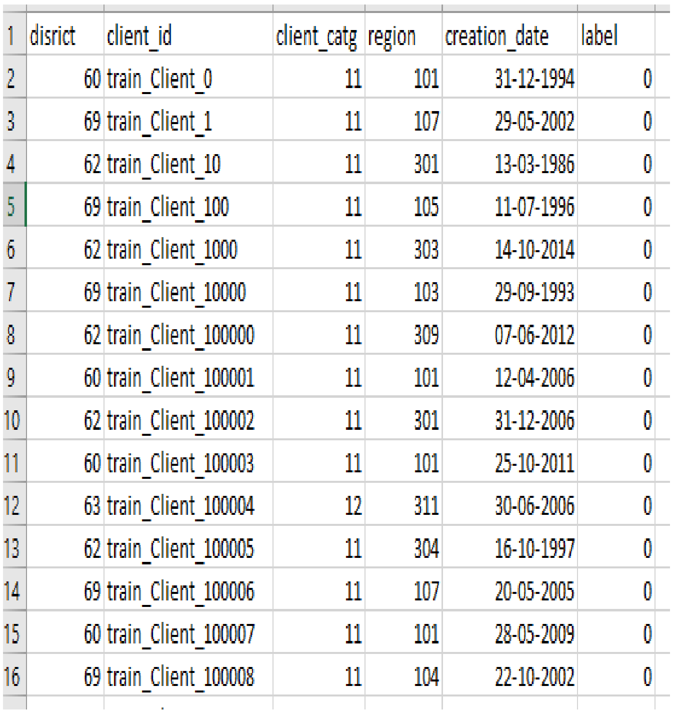
```
CNN Precision : 95.41663314022101
CNN Recall    : 94.60020676122372
CNN FMeasure  : 94.92587601078168
CNN Accuracy  : 94.92587601078168

CNN with Random Forest Precision : 100.0
CNN with Random Forest Recall    : 100.0
CNN with Random Forest FMeasure  : 100.0
CNN with Random Forest Accuracy  : 100.0
```

Fig4. In above screen with CNN-RF we got 100% accuracy and now click on 'CNN with SVM' button to train dataset with CNN and SVM

Fig 5.5 Train CNN with SVM



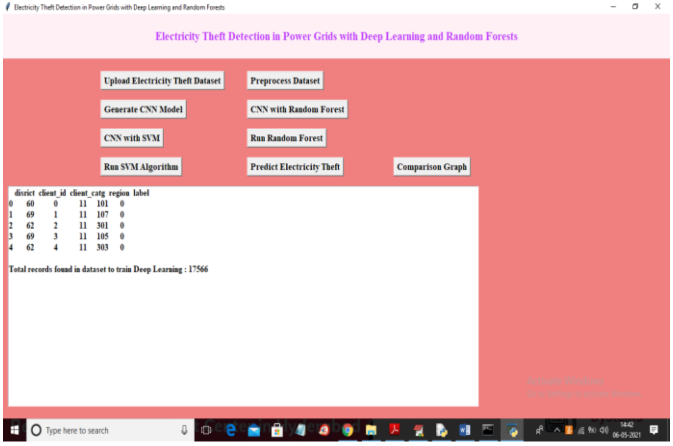
```
CNN Precision : 95.41663314022101
CNN Recall    : 94.60020676122372
CNN FMeasure  : 94.92587601078168
CNN Accuracy  : 94.92587601078168

CNN with Random Forest Precision : 100.0
CNN with Random Forest Recall    : 100.0
CNN with Random Forest FMeasure  : 100.0
CNN with Random Forest Accuracy  : 100.0

CNN with SVM Precision : 99.97597308986063
CNN with SVM Recall    : 99.96513249651325
CNN with SVM FMeasure  : 99.97054382519697
CNN with SVM Accuracy  : 99.97054382519697
```



1	disrict	client_id	client_catg	region	creation_date	label
2	60	train_Client_0	11	101	31-12-1994	0
3	69	train_Client_1	11	107	29-05-2002	0
4	62	train_Client_10	11	301	13-03-1986	0
5	69	train_Client_100	11	105	11-07-1996	0
6	62	train_Client_1000	11	303	14-10-2014	0
7	69	train_Client_10000	11	103	29-09-1993	0
8	62	train_Client_100000	11	309	07-06-2012	0
9	60	train_Client_100001	11	101	12-04-2006	0
10	62	train_Client_100002	11	301	31-12-2006	0
11	60	train_Client_100003	11	101	25-10-2011	0
12	63	train_Client_100004	12	311	30-06-2006	0
13	62	train_Client_100005	11	304	16-10-1997	0
14	69	train_Client_100006	11	107	20-05-2005	0
15	60	train_Client_100007	11	101	28-05-2009	0
16	69	train_Client_100008	11	104	22-10-2002	0

Fig. 1 In the above dataset first row contains column names and remaining rows contains dataset values and in last column we can see values as 0 or 1 which means normal or energy theft.

5.2 Pre-Processing.



```
   disrict  client_id  client_catg  region  label
0       60          0           11     101      0
1       69          1           11     107      0
2       62          2           11     301      0
3       69          3           11     105      0
4       62          4           11     303      0

Total records found in dataset to train Deep Learning : 17566
```

Fig.2 In above screen dataset loaded and displaying few records from dataset and this records contains non-numeric values and this values will not accept by machine

Fig 5.6 Train Random forest without SVM

Fig 6. In above screen with alone Random Forest we got 94% accuracy and now click on 'Run SVM Algorithm' button to train alone SVM with above dataset

Fig 5.7 Train SVM without CNN

Fig 5. In above screen with CNN-SVM we got 99% accuracy and now click on 'Run Random Forest' button to train alone RF on dataset

5.9 Comparison Graph

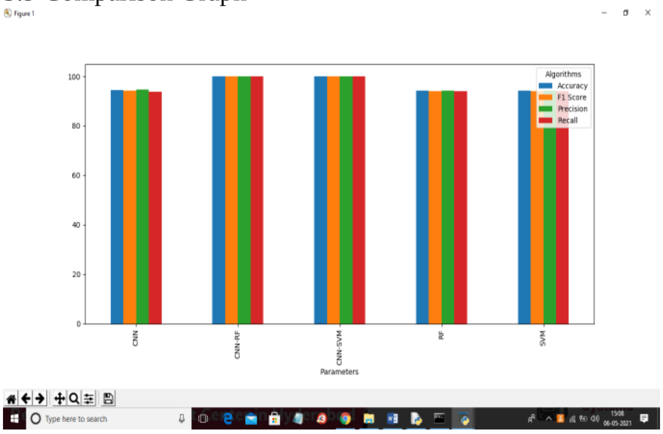


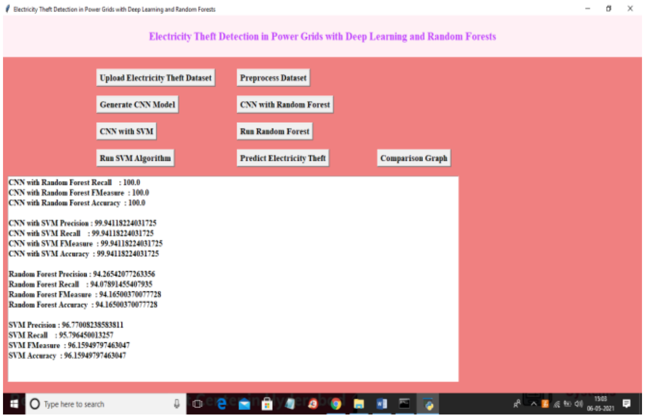
Fig 9 In above graph x-axis represents algorithm names and y-axis represents precision, recall, FSCORE and Accuracy for each algorithm and in all algorithms CNN-RF is giving 100% accuracy.

6. CONCLUSION-

In this research paper, a new approach called CNN-RF model is introduced to effectively detect instances of electricity theft. The model combines the power of Convolutional Neural Networks (CNN) as a feature extractor and Random Forest (RF) as the classification algorithm. To address the challenge of overfitting, a fully connected layer with a dropout rate of 0.4 is included during the model training phase. Furthermore, the SMOT algorithm is employed to handle the issue of imbalanced data.

To evaluate the performance of the CNN-RF model, several other machine learning and deep learning methods such as SVM, RF, GBDT, and LR were compared using SEAI and LCL datasets as benchmarks. The results clearly demonstrate that the proposed CNN-RF model shows great potential for accurately classifying instances of electricity theft. One notable advantage of the model is its ability to automatically extract meaningful features from smart meter data, eliminating the need for laborious and time-consuming manual feature engineering. Another advantage is that the hybrid model combines the strengths of both RF and CNN, which are widely recognized as successful classifiers in the field of electricity theft detection.

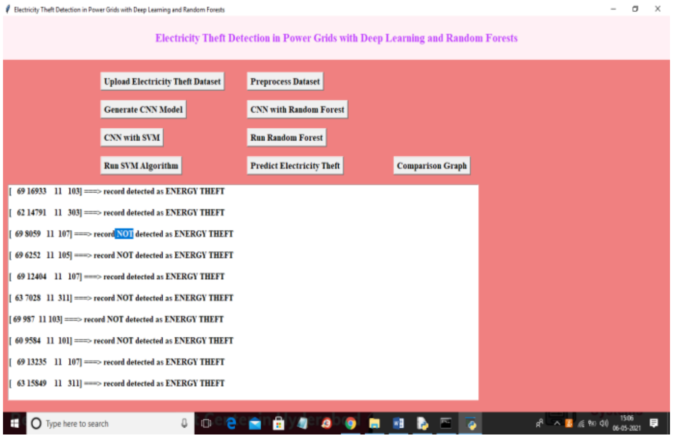
As the detection of electricity theft can impact consumer privacy, future research should focus on understanding how


Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
Upload Electricity Theft Dataset
Preprocess Dataset
Generate CNN Model
CNN with Random Forest
CNN with SVM
Run Random Forest
Run SVM Algorithm
Predict Electricity Theft
Comparison Graph
Type here to search

```
CNN with Random Forest Recall   : 100.0
CNN with Random Forest FMeasure  : 100.0
CNN with Random Forest Accuracy  : 100.0

CNN with SVM Precision : 99.94118224031725
CNN with SVM Recall   : 99.94118224031725
CNN with SVM FMeasure  : 99.94118224031725
CNN with SVM Accuracy  : 99.94118224031725

Random Forest Precision : 94.26542077263356
Random Forest Recall   : 94.07891455407935
Random Forest FMeasure  : 94.16500370077728
Random Forest Accuracy  : 94.16500370077728

SVM Precision : 96.77008238583811
SVM Recall   : 95.796450013257
SVM FMeasure  : 96.15949797463047
SVM Accuracy  : 96.15949797463047
```

Fig 7. In above screen with alone SVM we got 96% accuracy and now click on 'Predict Electricity Theft' button to upload test data

Fig 5.8 Load Test Data and get Prediction Result.


Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
Upload Electricity Theft Dataset
Preprocess Dataset
Generate CNN Model
CNN with Random Forest
CNN with SVM
Run Random Forest
Run SVM Algorithm
Predict Electricity Theft
Comparison Graph
Type here to search

```
[  69 16933   11   103] ===> record detected as ENERGY THEFT

[  62 14791   11   303] ===> record detected as ENERGY THEFT

[ 69 8059   11  107] ===> record NOT detected as ENERGY THEFT

[ 69 6252   11  105] ===> record NOT detected as ENERGY THEFT

[  69 12404   11   107] ===> record detected as ENERGY THEFT

[ 63 7028   11  311] ===> record NOT detected as ENERGY THEFT

[ 69 987  11 103] ===> record NOT detected as ENERGY THEFT

[ 60 9584   11  101] ===> record NOT detected as ENERGY THEFT

[  69 13235   11   107] ===> record detected as ENERGY THEFT

[  63 15849   11   311] ===> record detected as ENERGY THEFT
```

Fig8. In the above screen in square brackets we can see test data and after square brackets we can see prediction results as 'record detected as ENERGY THEFT' or 'record NOT detected as ENERGY THEFT'. Now click on 'Comparison Graph' button to get below graph

7. ACKNOWLEDGEMENT-

This work has been carried out as part of our academic project to be submitted to the university. In this project work, we got the guidance and all inputs from our internal guide Mr. Tejavath BalaKrishna and we are thankful to our guide for his constant support and encouragement without which, the paper could not be completed.

8. REFERENCES-

[1] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Electricity theft: overview, issues, prevention and a

the level of detail and time span of smart meter data can affect privacy concerns. Additionally, it would be valuable to explore the application of the hybrid CNN-RF model in other areas such as load forecasting. This would allow for further investigation into its potential and effectiveness in different domains.

smart meter based approach to control theft," *Energy Policy*, vol. 39, no. 2, pp. 1007–1015, 2011.View at: PublisherSite | Google Scholar

[2] J. P. Navani, N. K. Sharma, and S. Sapra, "Technical and non-technical losses in the power system and its economic consequence in the Indian economy," *International Journal of Electronics and Computer Science Engineering*, vol. 1, no. 2, pp. 757–761, 2012.View at: Google Scholar

[3] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A multi-sensor energy theft detection framework for advanced metering infrastructures," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1319–1330, 2013.View at: Publisher Site | Google Scholar

[4] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Security & Privacy Magazine*, vol. 7, no. 3, pp. 75–77, 2009.View at: Publisher Site | Google Scholar

[5] T. B. Smith, "Electricity theft: a comparative analysis," *Energy Policy*, vol. 32, no. 1, pp. 2067–2076, 2004.View at: Publisher Site | Google Scholar

[6] J. I. Guerrero, C. León, I. Monedero, F. Biscarri, and J. Biscarri, "Improving knowledge-based systems with statistical techniques, text mining, and neural networks for non-technical loss detection," *Knowledge-Based Systems*, vol. 71, no. 4, pp. 376–388, 2014.View at: Publisher Site | Google Scholar

[7] C. C. O. Ramos, A. N. Souza, G. Chiachia, A. X. Falcão, and J. P. Papa, "A novel algorithm for feature selection using harmony search and its application for non-technical losses detection," *Computers & Electrical Engineering*, vol. 37, no. 6, pp. 886–894, 2011.View at: Publisher Site | Google Scholar

[8] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, "The challenge of non-technical loss detection using artificial intelligence: a survey ficial intelligence: a survey," *International Journal of Computational Intelligence Systems*, vol. 10, no. 1, pp. 760–775, 2017.View at: Publisher Site | Google Scholar

[9] S.-C. Huang, Y.-L. Lo, and C.-N. Lu, "Non-technical loss detection using state estimation and analysis of variance," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2959–2966, 2013.View at: Publisher Site | Google Scholar

[10] O. Rahmati, H. R. Pourghasemi, and A. M. Melesse, "Application of GIS-based data water potential mapping: a case study at Mehran region, Iran," *CATENA*,vol.360–372,2016.View at: Publisher Site | Google