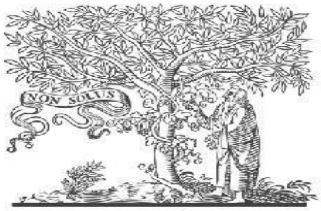
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 21st Feb 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 03

10.48047/IJIEMR/V12/ISSUE 03/20

Title **DEEP LEARNING-BASED ENERGY THEFT DETECTION IN SMART GRIDS**

Volume 12, ISSUE 03, Pages: 148-155

Paper Authors
D Kalpana, Dr A Manjula, CH Sowmya, G Nikitha

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Deep Learning-Based Energy Theft Detection in Smart Grids

D Kalpana[1], Associate Professor, Department of Computer Science and Engineering, Jyothishmathi Institute of Technology and Science, Telangana.
Dr A Manjula[2], Associate Professor, Department of Computer Science and Engineering, Jyothishmathi Institute of Technology and Science, Telangana.
manjula3030@gmail.com
CH Sowmya[3], Student, Computer Science and Engineering, Jyothishmathi Institute of Technological Sciences, Karimnagar, Telangana.
G Nikitha[4], Student, Computer Science and Engineering, Jyothishmathi Institute of Technological Sciences, Karimnagar, Telangana.

Abstract
The upgrading of smart grids plays a crucial role in electricity theft detection as they generate massive data which includes customer consumption data. Electricity theft is a global problem that negatively impacting on utility companies and also electricity users. it impairs utility companies' ability to grow economically, creates electric risks, and has an effect on consumers' high energy costs. Power loss in the transmission and distribution of electricity is an important problem faced by power utility companies from all over the world. Power losses are mainly categorized into technical losses and non-technical losses. Technical loss is related to electricity transmission, which is caused by internal scenes in power system components such as linear transmission; NTL is calculated as the difference between total losses and TLs, which is generally done by electricity theft. Basically, electricity theft mostly happens through physical attacks like eavesdropping, breaking the meter, or tampering with the meter reading.

Keywords: SVM, Convolutional Neural Networks, Random Forest, theft, smart grid.

Introduction

Utility providers may lose money as a result of electricity theft scams. For example, losses from power theft are projected to be more than $4.5 billion per year in the United States, while losses from electricity consuming firms globally are estimated to be above $20 billion per year. For example, the severe demand on electrical systems caused by power theft may result in fires that endanger public safety. As a result, reliable detection of energy theft is crucial to the safety and stability of the power infrastructure. Power providers have experienced huge quantities of electricity theft after implementing Advanced Metering Infrastructure (AMI) in smart grids. Yet, each coin has two distinct sides. The AMI Network opens the door to various new types of electrical thefts, which are triggered by the use of digital tools and cyberattacks. Human investigation of unlawful line diversions, comparison of malicious metre data with benign records, and study of problematic equipment or devices are all main goals of power theft detection. Unfortunately, when thoroughly examining all counters in the system, these actions are quite time consuming and costly. Therefore, these manual procedures are ineffective in preventing cyberattacks. Several approaches have been proposed in recent years to address the aforementioned issues. These approaches are mostly classified as state-based detection methods that rely on specialised hardware.

Related Work:

Authors in [1] provided the information about the types of losses by mentioning that non-technical loss during transmission of electrical energy is a major problem in developing countries and it has been very difficult for the utility companies to detect and fight the people countries to take the charge on the people who are responsible

for theft. Electricity theft forms a major cause of NTL. These losses show a greater impact on quality of supply, increase load on the generating station, and affect tariff imposed on genuine customers. This paper gives us an idea of the factors that influence the consumers to steal electricity. Holding these ill effects, various methods for detection and estimation of the theft are introduced. This paper generates an overview design of smart meter, harmonic generator, and filter circuit. The ultimate goal of this work is to detect illegal consumers and conserve and effectively utilize energy and also smart meters are designed to provide data of various parameters related to instantaneous power consumption. NTL in the distribution feeder is operated by external control station from the sending end information holding of the distribution feeder. If a considerable amount of Non-technical loss is detected, then harmonic generator is gets operated at baselines which is extracted from consumer's earlier validated usage. It uses sums of chi-square random variables, sums of squares among-group variation (SSA), among-block variation (SSBL), and random variation (SSE). Authors in [3] explained that they have included five phases. Phase 1 includes Physical Tampering Detection Solutions. It notifies some meter tamper alerts provide some of following observations like anti- that feeder for introducing additional harmonic component for destroying appliances of the illegal consumers. For analysis, cost-benefit analysis for implementation of the introduced system in India is involved. Authors in [2] proposed that state estimation-based approach for distribution transformer load estimation is expressed for detecting meter tampering and provided evidences of Non-Technical loss. In this paper, they have used Typical Taiwan Power Company (TPC) distribution feeder data. The results included the NTL detection and energy theft scenarios. It also estimates the power usage during meter reading irregularity periods and the tested results showed that ANOVA is useful in detecting individual meter data anomoly. Detecting losses is done with the process by including two phases. Phase 1 includes Localization of Anomalous Usage, which includes customer meter data from AMI/AMR, database in every hour and conjunction with network topology data from automatic mapping system also it uses DSE algorithm for the accurate measurement of electricity consumption. Phase 2 includes Detection of Consumer Meter Defect or Tampering which includes ANOVA. It is used to compare meter messade curves with the tampering alert to detect Ap1, explained that they have included five phases. Phase1 includes Physical Tampering Detection Solutions. It notifies some meter tamper alerts provide some of following observations like anti-tampering alert to detect Ap1, reverse rotation alert to detect Ap2, disconnect alert to detect Ap3, anti-tampering alert to detect Ap4. It consists AMIDS will most likely marks the triggered alert as a false positive and will not report any type of intrusion. Phase2 includes Cyber Intrusion Detection Systems which includes a remote cumulative attestation kernel (CAK). phase 3 includes Power Measurement-based Anomaly Detection uses Naïve Bayesian algorithm. In this paper, they presented AMIDS, an integrated intrusion solution to identify energy theft attempts in advanced meter infrastructure and also gives an evolution of probabilities during energy theft and energy consumption. Additive increase/multiplicative decrease is a feedback control algorithm better known for its TCP use congestion control. AMID combines linear growth of the congestion window when there is no congestion with an exponential reduction when congestion is detected. Authors in [4] introduced a method called clustering. Electricity theft detection includes a methodology consists of three phases. Phase A includes Local Outlier Factor which finds outliers which aims to find out the abnormal objects. It uses LOF method on local density and also uses reachability distance. Phase B includes clustering and local outlier factor, it is an extension of LOF called clustering and local outlier factor (CLOF). The main

aim of this phase is to cluster the dataset with k-means and to select the outliers which are deviate from pre-defined clusters.

Phace C includes Detection Framework and this phase internally holds three more modules: Pre-processsing module, the detection module and the judgement module. In these phases, after evaluating the dataset and applying the frameworks they had arranged the datasets according to the ranks obtained by the values in the k-mean algorithm. Finally, a user is noted committing electricity theft if their average rank is high. But the drawback is that it only analyses electricity consumption alone. It does not specialize in detecting linear FDI, which is adopted by most physical attacks called technical attacks

Preliminaries

Convolutional Neural Networks (CNN), Support Vector Machine (SVM), Synthetic Minority Oversampling Technique (SMOTE), Random Forest (RF), Gradient – boosted decision trees (GBDT), Logistic Regression (LR).

A. Convolutional Neural Network

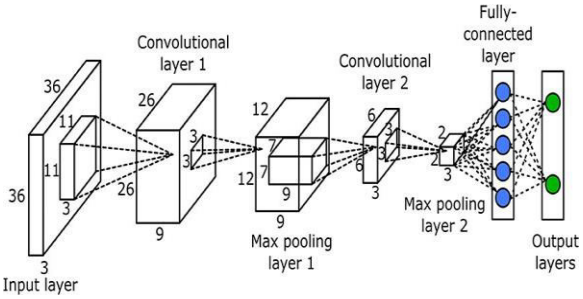


Figure 1: Convolutional Neural Network

A neural network will include an input layer, hidden layer and an output layer. CNN are motivated from the architecture of the human brain similar to a neuron in the brain processes and transfers the data continuously throughout the body. Neurons in CNN takes input, undergo processing and transfers the result as output. The input layer holds the image units as input in the form of arrays. In CNNs, there could be hidden layers, which perform feature extraction from the image by doing calculators.

CNN are feedforward networks in that information flows takes place in one direction only, from their inputs to their outputs. Just as ANN are biologically inspired. CNN architecture come in several variations; however.

In general, they consist of convolutional and pooling for subsampling layers. There are some which are being associated with the input nodes or neurons. These are assigned to satisfy for giving the actual and target output. On comparing, if the actual is not equal to target output, then we can change the old weights by assigning a new values to them.

1.Linear Activation Function

Although a linear equation is straightforward to solve, it has a restricted ability to handle more difficult issues and is less powerful when it comes to inferring complex functional mappings from data. Without an activation function, a neural network is nothing more than a simple linear regression model.

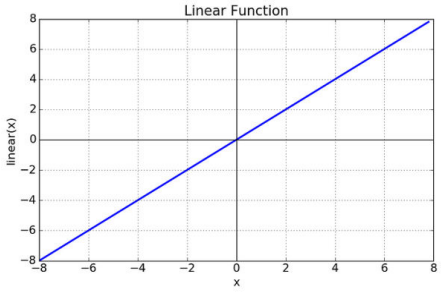


Figure 2: Identity Function

Equation : f(x) = x

Range = - infinity (∞) to + infinity (∞)

The linear activation function, also known as “no activation” in which the activation is proportional to the input. The function doesn’t do anything to the weighted sum of the input, it simply spits out the value it was given

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

2. Non-Linear Activation Function

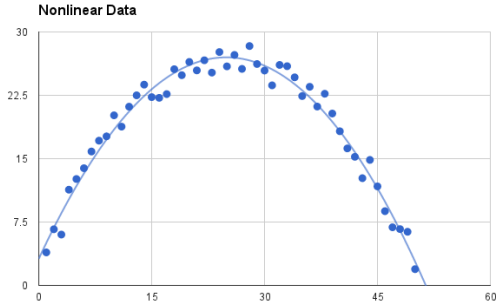


Figure 3: Non- Linear Curve

These are also known as to be the highly using activation functions and makes easy for a neural network model to adopt with a variety of data and to differentiate between the outcomes.

2.1 Sigmoid Activation Function

The sigmoid activation function is holds by the equation

$F(z) = 1/1+e^{-z}$

where input z belongs to –infinity to +infinity. Sigmoid is continuous and differentiable everywhere. Due to less computation in finding its derivative, this activation function is widely used in shallow neural networks.

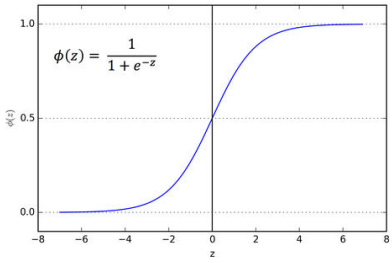


Figure 4: Sigmoid Activation Function

It is rarely used in DNN's hidden layers because of its soft saturation property which makes DNNs delay converging during training.

2.2 Hyperbolic Activation Function

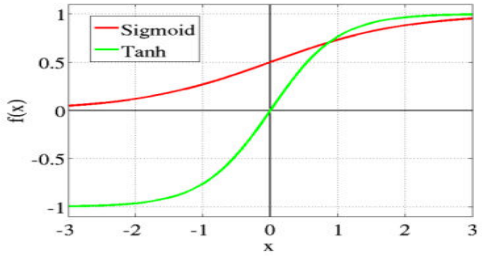


Figure 5: Tanh activation function

Like the sigmoidal function, hyperbolic tangent is continuous and differentiable everywhere. It is given by the given equation as follows

$F(x) = \tanh(x)$

The input is belongs to –infinity to +infinity and an activation lies in the range of -1 to +1.

2.3 ReLU Activation Function:

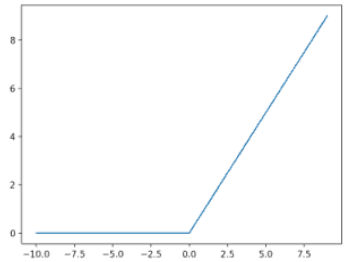


Figure 6: ReU Activation Function

$F(x) = \max(0,x)$

If x is greater than or equal to 0 then f(x) is z. Otherwise it should be equal to 0.
The ReLU is actively using activation function right now. It is widely used in all of the algorithms used in CNN and deep learning too. The function and its derivative both are monotonic.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

B. Support Vector Machine:

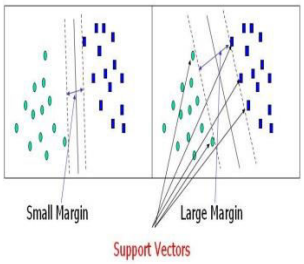


Figure7: Support Vector

Support vector machines are a part of supervised learning methods used for classification, regression and outlier detection. Influence the position and orientation of the spaces, in which the number of dimensions is greater than the number of samples.

C. Synthetic Minority Oversampling Technique

Synthetic Minority Oversampling is a statistical technique for increasing the number of cases in your dataset in a balanced way. The component works by generating new instances from existing minority cases that you supply as input. This helps us to eliminate the overfitting problem posed by random oversampling.

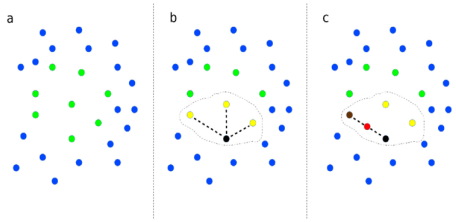


Figure 8: Oversampling

D. Random Forest

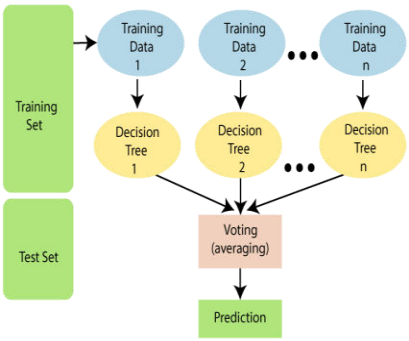


Figure 9: Random Forest Hierarchy

Random forest is a commonly used machine learning algorithm trademarked by Leo Breiman and Adle Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have adoption, as it handles both classification and regression problems. It holds the property like bagging which is the limitation for softmax classifier. The 80% train data set and 20% test data set. data set is usually divided into two sets:

E. Gradient-Boosted Decision Trees

Gradient-boosted decision trees are a machine learning technique for optimizing the predictive value of a model through successive steps in the learning process. It is mainly used in regression and classification tasks among others. It gives prediction model in the form of an ensemble of weak prediction model.

In the below figure, w_1,w_2.....w_n are the weights associated with the inputs nodes which we have inserted into the network. Let us consider, a gradient boosting algorithm with M stages m(1<=m<=M) of gradient boosting, suppose some imperfect model F_m, our algorithm should use a new estimator h(m)

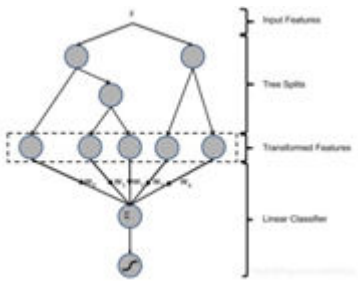


Figure 10: Gradient boosting

$_{+1}(x_i) = F_m(x_i) + h_m(x_i) = y_i$

y_i = the number of samples in y

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

F. Logistic Regression

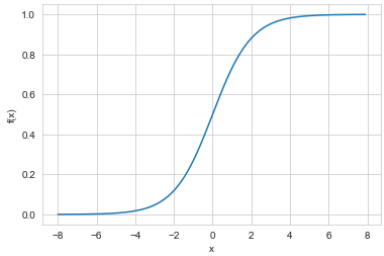


Figure 11: Logistic Regression

Logistic regression is a data ananlysis technique that uses mathematics to find the relationships between two data factors. It then uses this relationship to predict the value of one of those factors based on the other.

Table 1 : Obtained Results

Feature	F-Measure	Precision	Recall	Accuracy
CNN	100	100	100	100
CNN with SVM	99.94	99.94	99.94	99.94
RF	94.16	94.26	94.07	94.16
SVM	96.15	96.77	95.79	96.15

CNN-RF Based Electricity Theft Detection

Among all the classifiers in neural networks, the random forest classifier takes advantage of some of the machine learning techniques includes bagging and random feature selection which would overcome the limitation of the SoftMax classifier. By considering some of the methods, a novel convolutional neural network-random forest (CNN-RF) model is considered for detection of electricity theft. The CNN is proposed to capture various features of consumers consumption behaviours from meter data automatically, which is one of the key factor in the success of electricity theft detection model. Our methodology consists of nine modules and four phases. The modules are as follows

a) Upload Electricity Theft Dataset
b) Pre-process Dataset
c) Generate CNN Model
d) CNN with Random Forest
e) CNN with SVM
f) Run Random Forest
g) Run SVM algorithm
h) Predict Electricity
i) Comparison Graph

Phase 1: Uploading and pre-processing the SGCC dataset

In this phase, we have uploaded the dataset to the application. We have cleaned the dataset by removing missing values and machine learning algorithms will take characters values. Therefore, we need to pre-process the dataset by considering all non-numeric characters to convert into numeric characters. After pre-processing we have splitted dataset into train and test where 80% dataset will be used for training and 20% for testing.

Phase 2: Generating CNN model

This phase includes the model which is used to train CNN with dataset. A convolutional Neural Network is a type of artificial neural network, which is broadly uses for image or object recognition and classification.

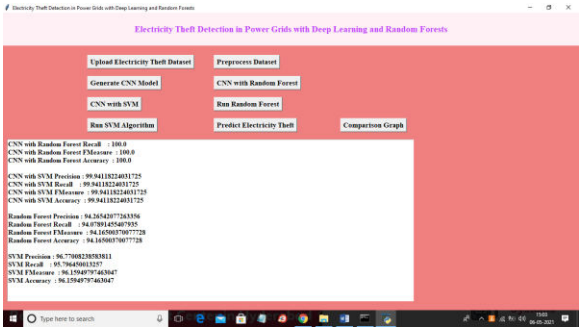


Figure 12: Obtained accuracy

Phase 3: Running CNN with Random Forest and CNN with SVM

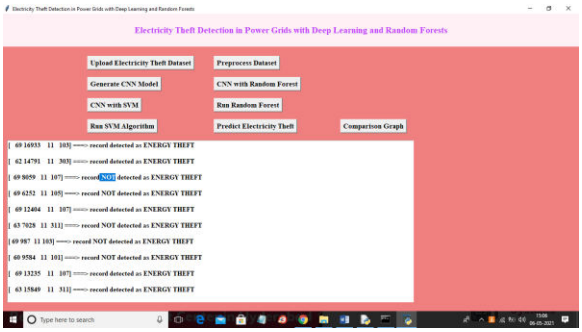


Figure 13: Obtained theft records

CNN with Random Forest is used to get accuracy. Random Forest is a mostly using machine learning algorithm which belongs to the supervised learning technique, in which it can be used for both classification and regression problems in machine learning. It is generally based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. Support vector machine is one of the most popular supervised learning algorithms, which is used for classification and also regression problems, which it can used for classification problems in machine learning.

Phase 4: Predicting Electricity theft and showing the comparison graph

This phase is identified as the final phase which involves the key factor i.e., electricity theft detection and identifies the records which are observed as theft among more than 17,000 records in SGCC 2016 dataset. The next work of this phase is to display a comparisons graph between the accuracy among the algorithms which we have used. It holds the attributes like accuracy, precision, recall and FSCORE

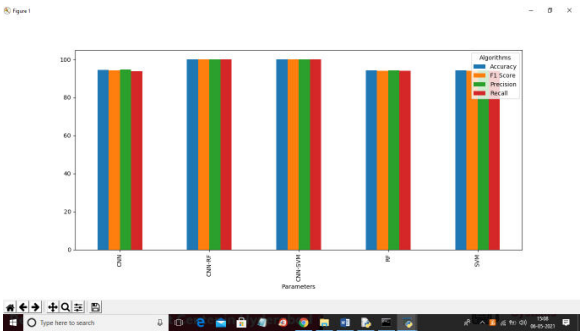


Figure 14: Comparison Graph

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.

Conclusion

The detection of electricity theft is performed using a unique CNN-RF model. In this model, the CNN is taken into account, acting similarly to an automatic feature extractor while looking into data from smart metres. Random Forest is used as the output classifier, as usual. A fully connected layer with a dropout rate of 0.4 is created during the training phase since a large number of parameters must be tuned, increasing the work that involves the risk of over-fitting. In addition, the SMOT method amused to resolve the issue of data imbalance. As a benchmark, some machine learning and deep learning techniques like SVM, RF, GBDT, and LR are applied to the same problem. All of these techniques have been tested using SEAI and LCL datasets. Customers' privacy may be impacted by the identification of electricity theft, hence future study will be focused on determining how the granularity and duration of smart metre data may impact this privacy.

REFERENCES

[1] Shuan Li, Yinghua Han, Xu Yao, Song Yingchen, Jinkuan Wang and Qiang Zhao, October 31, 2019 by Journal of Electrical and Computer Engineering, “Electricty theft:Overview,issues,prevention and smart meter based approach to control theft”.

[2] Shih-Che Huang, Yuan-Liang Lo, Chan-Nan Lu, IEEE, 2013 “Non-Technical Loss Detection Using State Estimation and Analysis of Variance”.

[3] Stephen NcLaughlin, Brett Holbert, Ahmed Fawaz, Robin Berthier and Saman Zonouz, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATION, VOL. 31,NO. 7, JULY 2013 “A Multi-Sensor Energy Theft Detection Framework for Advanced Metering Infrastructures”

[4] Yanlin Peng, Yining Yang, Yuejie Xu, Yang Xue, Runan Song, Jinping Kang, Haisen Zhao, IEEE, Digital object identifier: 10.1109/ACCESS. 2021.3100980 “Electrcity Theft Detection in AMI Based on Clustering and Local Outlier Factor”

[5] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, “Electricity theft: overview, issues, prevention and a smart meter based approach to control theft,” Energy Policy, vol. 39, no. 2, pp. 1007–1015, 2011.

[6] J. P. Navani, N. K. Sharma, and S. Sapra, "Technical and nontechnical losses in power system and its economic consequence in Indian economy," IJECSE, vol. 1, no. 2, pp. 757–761, 2012.
[7] 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.
[8] 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.
[9] T. B. Smith, "Electricity theft: a comparative analysis," Energy Policy, vol. 32, no. 1, pp. 2067–2076, 2004.
[10] J. I. Guerrero, C. Le´on, I. Monedero, F. Biscarri, and J. Biscarri, "Improving knowledge-based systems with statistical techniques, text mining, and neural networks for nontechnical loss detection," Knowledge-Based Systems, vol. 71, no. 4, pp. 376–388, 2014.
[11] 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.
[12] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, "the challenge of non-technical loss detection using artificial intelligence: a surveyficial intelligence: a survey," International Journal of Computational Intelligence Systems, vol. 10, no. 1, pp. 760–775, 2017.
[13] 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.
[14] O. Rahmati, H. R. Pourghasemi, and A. M. Melesse, "Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran region, Iran," CATENA, vol. 137, pp. 360–372, 2016.
[15] P. Kadurek, J. Blom, J. F. G. Cobben, and W. L. Kling, "Theft detection and smart metering practices and expectations in the Netherlands,"in Proc. IEEE PES Innovative Smart Grid Technologies Conf. Europe(ISGT Europe), 2010.
[16] S. McLaughlin, D. Podkuiko, and P. McDaniel, "Energy theft in the advanced metering infrastructure," in Proc. CRITIS 2009, 2010, vol. LNCS 6027, pp. 176–187.
[17] I. H. Cavdar, "A solution to remote detection of illegal electricity usage via power line communication," IEEE Trans. Power Del., vol. 19, no. 4, pp. 1663–1667, Oct. 2004.
[18] A. Pasdar and S. Mirzakouchaki, "A solution to remote detecting of illegal electricity usage based on smart metering," in Proc. 2nd IEEE Int. Workshop Soft Computing Applications (SOFA 2007), Aug. 21–23, 2007.
[19] J. Nagi, A. M. Mohammad, K. S. Yap, S. K. Tiong, and S. K. Ahmed, "Non-technical loss analysis for detection of electricity theft using support vector machines," in Proc. 2nd IEEE Int. Conf. Power and Energy, Dec. 1–3, 2008.
[20] D. Matheson, C. Jmg, and F. Monforte, "Meter data management for the electricity market," in Proc. 8th Int. Conf. Probabilistic Methods Applied to Power Systems, 2004.
[21] zhongzong yan, He wen,"performance analysis of electricity theft detection for the smart-gird: an overview", IEEE transaction on instrumentation and measurement 2021.