CONF-IRM 2022 Proceedings

International Conference on Information Resources Management (CONF-IRM)

10-2022

Providing an efficient framework for power theft detection based on combination of Raven roosting optimization algorithm and clustering and classification techniques

Hassan Ghaedi

Sahar Soleimani

21. Providing an efficient framework for power theft detection based on combination of Raven roosting optimization algorithm and clustering and classification techniques

Hassan Ghaedi
Department of Computer, Khormuj Branch,
Islamic Azad University
hassan.ghaedi@iau.ac.ir

Sahar Soleimani
Department of Computer, Khormuj Branch,
Islamic Azad University
s.soleimani@samabushehr.ac.ir

Abstract

One of the main concerns of power generation systems around the world is electricity theft. One of the goals of the Advanced Measurement Infrastructure (AMI) is to reduce the risk of electricity theft in the electric smart grids. However, the use of smart meters and the addition of a security layer to the measurement system paved the way for electricity theft. Nowadays, machine learning and data mining technologies are used to find abnormal patterns of consumption. The lack of a comprehensive dataset about abnormal consumption patterns, the issue of choosing effective features, the balance between consumer's normal and abnormal consumption patterns, and the choice of type and number of classifiers and how to combine them are the challenges of these technologies. Therefore, a detection system for electricity theft that is capable of effectively detecting theft attacks is needed. To this end, a framework including data preparation phases, feature selection, clustering, and combined modeling have been proposed to address the aforementioned challenges. In order to balance normal and abnormal data, 6 artificial attacks have been created. Moreover, with respect to the Chief element in the Raven optimization algorithm and its two-step search feature, this algorithm has been used in feature selection and clustering phases. Stacking as a two-step combined modeler has been used to strengthen the prediction of accuracy. In the second step of this modeler, the meta-Gaussian Processes algorithm is used due to the high accuracy of detection. The Irish Social Science Data Archive (ISSDA) dataset has been used to evaluate performance. The results show that the proposed method identifies dishonest customers with higher accuracy.

Keywords: Power theft Detection, Classification, Feature selection, smart power grid, Clustering.

1. Introduction

In smart power grids, it is important to know the amount of power consumed by customers in a moment is necessary so as to accurately predict and plan the electricity demand in the future. In the smart power systems, if a user by a normal consumption is detected as a dishonest user, the system is not damaged, however, if a dishonest user is detected as a right user, the system will be damaged so that if the total number of these users is high, the system will be damaged more. Many studies have been done to explore electricity theft using machine learning and data mining technologies through customer consumption patterns, but each has some kind of challenge. One of these challenges is the lack or deficiency of existing samples of abnormal consumption of electricity customers. For example, there are no or a few abnormal samples for a customer. The existence of abnormal and manipulated samples of consumption can make it easier to identify dishonest customers for classification. Thus, the lack of a comprehensive Dataset, which includes both normal and abnormal samples, limits the discovery rate of theft. That is why a comprehensive and balanced Dataset is required. Another challenge is the lack of attention to the issue of feature selection. Removing some features may not have any effect on the final result of detection, and the failure to select some of the important features can have a significant effect on the correct detection of the classification result. By converting the property selection issue into an optimization issue, it can be solved by a meta-heuristic algorithm and extracted important features. The selection of type and number of classifiers and how to combine them are other challenges of using data mining techniques in the issue of electricity theft detection, in which often one classifier or sometimes two classifiers have been used. Therefore, an electricity theft detection system that is capable of effectively detecting theft attacks is essential. Several studies have been carried out to detect electric theft using data mining techniques through customer consumption patterns, however, each has some challenges, as mentioned before. One of these studies has done by (Jokar et al., 2016) the strength of their research is the use of single-class and multi-class support vector machines to detect theft. Moreover, the abnormal samples of customer consumption were also slightly produced in their research. One of the weaknesses of this research is the imbalance between normal and abnormal samples of customer consumption. In fact, the low number of abnormal samples has challenged classifiers to detect electricity theft. The use of the support vector machine (SVM) classifier alone and the lack of use of other methods, especially the combination methods, have challenged their research study. Therefore, the main goals of this research are to create a balance between normal and abnormal examples of customer consumption and increase the accuracy of electricity theft detection using combined classification methods and selecting important and effective features. This research study is organized as follows: Section 2 reviews the previous works. The Black Raven Optimization algorithm is reviewed in Section 3. Section 4 describes the proposed framework. Examinations and evaluations are discussed in Section 5 and Section 6 is dedicated to the research conclusions.

2. Related Works

Considering the importance of detecting and discovering electricity theft from smart grids, many research studies have been carried out over the past few years. In this section, some of them which have used data mining techniques are discussed. (Jeyakumar & Devaraj, 2018) proposed a new approach to identifying suspect customers using a consumption pattern. So that, if there is a difference between the produced electricity and the consumed electricity in a district, all customers belonging to that area are considered a suspect. In their research, customers were classified into the K cluster using the k-nearest neighbors (KNN) algorithm and by using the normal customer profiles, three types of abnormal customers were generated. A readout was made for each customer every half hour, and the customer profiles were categorized using the artificial neural network (ANN) algorithm. To evaluate the efficiency of the proposed method, the Accuracy and Error Rate parameters were used. The drawback of the proposed method is that electricity theft is based on the assumption that suspected customers are fraudulent customers. As reported by (Sowndarya & Latha, 2017), AMI provides two-way communication between the power industry and customers, which eliminates manual intervention in reading the meter. In spite of the advantages of smart meters, there are some disadvantages related to these meters. Manipulating smart meters may not be possible, however, electricity theft is still possible by bypassing smart meters. In their research, a new framework detects fraudulent customers based on customer consumption patterns. Because the KNN algorithm has the ability to retrain, it reinforces the proposed framework against unwanted changes in the consumer pattern. The disadvantage of this method is the lack of existing examples (samples) related to customer’s abnormal consumption patterns. (Yeckle & Tang, 2018) stated that AMI is a major component of grid networks that is responsible for collecting, measuring, and analyzing customers' energy consumption. Despite the advances, new problems have arisen in AMI, particularly electricity theft. To cope with these challenges, Dataset related to consumption data was analyzed. Their research provides the possible use of outlier data detection algorithms to increase AMI security. The performance of the proposed algorithm is examined on a real Dataset, and a data preprocessing method is also performed by the K-means clustering algorithm with the aim of reducing the number of sample measurements per day. (Jokar et al., 2016) used single-class and multi-class support vector machines to discover fraudulent customers. They create a balance between true and dishonest data by creating a new dataset using manipulated data by customers. (Feng et al., 2020) presented an algorithm for detecting abnormal power consumption patterns based on local matrix reconstruction (LMR). Five daily load characteristics were used instead of high daily amplitude load curves and PCA technique to calculate weighted regeneration errors. (Zhang et al., 2020) proposed a feature engineering-based method for detecting abnormal power consumption behavior. First, the main features were created by brainstorming. Then, the optimal features were obtained based on the variance and similarity between the selected features. (Nazmul Hasan et al., 2019) proposed a hybrid power theft detection system that combines a convolutional neural network (CNN) with long short-term memory (LSTM). Since power consumption is time-series data, the CNN-LSTM model was used for classification, and a technique based on local values was employed to calculate missing items. In their study, Li et al. (Li et al., 2019) proposed a hybrid model based on CNN and random forest (RF) to detect power theft. In this model, CNN was first used to learn features between different hours of the day and different days extracted from smart meter data. A dropout layer and a back-propagation algorithm were used in this model to reduce over-fitting and update network parameters in the training phase, respectively. The RF was then trained based on the characteristics obtained to determine whether the customer was robbing or not. Although much research has been done in identifying and detecting theft by using data mining, the challenges still are there such as not paying attention to abnormal` samples of customers in datasets, the issue of effective feature selection as well as the type and number of classifiers and so on. In the light of the above-mentioned facts, this research tries to address the aforementioned challenges with the help of Raven roosting algorithm in combination with other methods.

3. Raven Roosting Optimization (RRO) Algorithm

The Raven Roosting optimization (RRO) algorithm is inspired by the process of feeding and gathering a bird called the common Raven, developed by (Jokar et al., 2016). According to RRO algorithm, after selecting an accumulation place, the algorithm parameters are set. The location of the Ravens is then randomly selected (a potential food location) and the fitness value of each N Raven is evaluated, and the Raven located in the best location is selected as Chief. After that, a percentage of the Ravens (PERCfollow) is used to leave the roosting site randomly within a radius of Rchief and others will look for food at their best location.

All Ravens memorize one of their searches (SEARCH) for the source of food during the flight to the destination. Ravens perform this process by dividing their flight into Nstep so that the length of each step is randomly selected. Each Raven at each step senses the quality of its step point and searches for a range within the radius of Rpcpt. If amongst these locations, it finds a better location than its best personal one, there is one percent chance of Probstop that the Raven stopped flying at that point and doing a search at that new location; otherwise, it performs the next flight stage and keeps moving on to its destination. In Ravens' foraging algorithm, a feeling mechanism has been embedded whereby followers perceive other locations on the path, and if they find a better location than their own, they will accidentally stop in these locations. For ravens that reach their destination (in the vicinity of the Chief or the best own), if the best personal locations need to be updated, they will be updated and fitness function will be evaluated for each Raven. Finally, if the location of the best solution also needs to be updated, it will be updated.

4. The Proposed Method

Given the importance of detecting electricity theft and the existence of challenges mentioned in the previous sections, it is necessary to provide a method for addressing and solving these challenges. Here, the RRO algorithm is used in the phases of feature selection and clustering in order to better identify and increase the accuracy of electricity theft detection. According to Figure 1, the proposed method involves several executive phases. In the first phase, which is known as the data preparation phase, pre-processing operations are performed. The electricity may not be consumed during the hours of night or days of the year and the value of 0 is sent to the control center. Using preprocessing techniques, appropriate data is placed instead of empty values. Moreover, to improve the classification accuracy, the linear normalization of features is done in this phase. In order to improve classification performance, effective features are selected in phase 2. Nowadays, in many aspects of feature selection, meta-heuristic methods, for example, genetic algorithm (GA), ant colony algorithm (ACO) and particle swarm optimization (PSO) algorithms are used to select effective features. Each of these methods suffers from weaknesses.

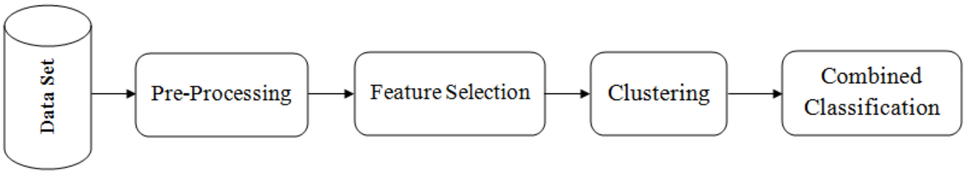

Figure 1: Steps of the proposed algorithm for theft detection

To overcome these problems, a new and more efficient way is needed. Correlation information of features is an important factor in helping the search process. In this phase, an effective 6-step method is proposed based on the RRO algorithm that uses the correlation information of features (Monirul Kabir et al., 2010) to guide the search process in the RRO algorithm. In order to determine the size of the subset of features, a random probability function is used, according to the following Eq.(1):

$$p_s = \frac{f - s}{\sum_{i=1}^{p}(f - i)} \quad (s \leq f) \qquad (1)$$

Where f and s are the number of initial Dataset features and selected features, respectively; p is the difference between f and s (p = f-s), and ps is the probability value of determining s as the initial number of features. According to Eq.(1), the probability value of the ps increases, if s value is lower. The value of s is randomly selected in the range of [3, k] (k=∈×f). Indeed, the grouping of features is the division of features into groups of similar objects with the aim of finding the relationship between features, so that the algorithm is capable to select the privileged features (Monirul Kabir et al., 2010). In this research, the Pearson correlation coefficient is used to measure the correlation between different features of a Dataset. The greater the correlation coefficient between the two features, the more similar the two features. The correlation coefficient between the two features i and j are shown by rij and is obtained according to Eq.(2):

$$\mathrm{rij} = \frac{\sum_{k=1}^{m}(\mathrm{xi(k)} - \overline{\mathrm{xi}})\big(\mathrm{xj(k)} - \overline{\mathrm{xj}}\big)}{\sqrt{\sum_{k=1}^{m}(\mathrm{xi(k)} - \overline{\mathrm{xi}})^{2})}\sqrt{\sum_{k=1}^{m}(\mathrm{xj(k)} - \overline{\mathrm{xj}})^{2})}} \qquad (2)$$

where m is the number of samples, and xi(k) and xj(k) are the values of the features i and j for the sample kth, respectively. The variables $\overline{xi}$ and $\overline{xj}$ are, the mean of xi and xj for the m sample, respectively. After calculating the correlation coefficient for all possible components of features, the correlation of each feature i is calculated according to Eq.(3).

$$cori = \frac{\sum_{j=1}^{f} |rij|}{f-1} \quad if \;\; i \neq j \tag{3}$$

where f is the total number of features. A higher correlation value indicates the higher similarity of the feature with other features. To build two groups of features, RRO-LS algorithm arranges features ascending based on their correlation values. The first half with less correlation is placed in the unlike group(U=unlike) and the second half in the Alike group(A=Alike). In the proposed method, each Raven is identified by a binary vector. The length of each vector is equal to the number of initial features, and if the values of one cell of the vector are equal to 1 or 0, it means that the corresponding feature has been selected or not been selected, respectively. Firstly, the initial location of each Raven is randomly initialized. In the RRO algorithm, on each flight, the location of each Raven changes in a Rpcpt radius of a circle according to Eq.(4).

$$\mathrm{NewLoc} = \mathrm{Rv}(\mathrm{j},:) + \mathrm{rand}(1,\mathrm{D}) + \mathrm{Rpcpt} \tag{4}$$

where Rpcpt is the radius of the flying circle and Rv(j, :) is the initial location of the Ravens.

Then, according to the group Chief, the location of the follower Ravens, in a circle with the Rchief radius, is expressed according to Eq.(5).

$$\mathrm{NewLoc} = (\mathrm{Chief} \;-\; (1 - 2 \;*\; \mathrm{rand}\,(1,\mathrm{D}) \;*\; \mathrm{Rchief}) \quad (5)$$

With a local search strategy, a classifier can learn all important information about a Dataset and plays a key role in the proposed algorithm (Kabir et al., 2011). To carry out local search operations, two stages are considered including the segmentation of features and Ravens displacement. For each Raven, the Add and Dell operators are used so as to improve their local search. In this way, a Raven uses the Add operator to add a number of desirable features and uses the Dell operator to remove a number of features from its location. In this strategy, firstly, the number of bits 1 generated from a newly Raven, for example, 01101011, is identified and is placed into a subgroup called F.

$$F = \{F_2, F_3, F_5, F_7, F_8\} \tag{6}$$

Then each element of F is compared with members of A and U sets and divided into two subgroups FA and FU. FA means that all alike features are located in both A and F sets. FU means that all unlike features are located in both U and F sets. Then all FA and FU features are arranged in ascending order based on their correlation values. The most important step is the Raven displacement. At this stage, the number of bits 1 generated in Raven must be managed. For this purpose, according to Eq.(7), the values of nA and nu, which represent the number of alike and unlike features, respectively, are calculated as follows:

$$n_u = \alpha.S, \qquad n_A = (1-\alpha)S \tag{7}$$

where α the control parameter controlling the nu; s is the size of the subset of the selected features. With two Add and Dell operators, the features are added to or removed from the Raven. Whenever $|FU| < n_u$, then, the number of (nu-FU) features in (U-FU) is added to Raven by Add operator, otherwise, with the Dell operator, the (FU-nu)features in FU should be removed from the Raven. Besides, if $|\mathrm{FA}| > \mathrm{n_A}$, the number of (FA-nA)features in FA is removed from the Raven by Dell operator, otherwise, with the Add operator, the (nA-FA) features in (A-FA) arebeing added to the Raven. Then, the new Dataset is divided into two teaching and testing groups, and the 10-flod cross-validation method is used to evaluate each Raven using the KNN classifier. At the end, the value of the fitness function is compared to the best overall value and Chief, and updates are done. In the third phase, data clustering is performed in such a way that the data is clustered using the proposed RRO-CL algorithm. The goal of this phase is to label the normal and abnormal data in order to build appropriate models for the classification of new customers. In this algorithm, data is split into the desired clusters using the RRO optimization algorithm. Figure 2 illustrates the proposed algorithm for data clustering. In the following, the steps of the proposed algorithm for clustering RRO-CL are explained. Step 1: Adjustment of the parameters including the number of Ravens; the domain of the problem, the perception radius of the Ravens, the group Chief radius, the number of execution stages, the number of followers, and so on. Then each Raven is randomly assigned an initial location of Xind. Since each Raven is represented as vector with k cluster centers, Xind is shown as:

$$Xind = (Ci1, Cij, \dots, Cik) \qquad (8)$$

Cij represents jth cluster center for the ith Raven.

Step 2: Each Raven will randomly be initialized with the initial location x. Step 3: At the start, the values of pbest and Chief are initialized with an infinity value of inf, then for each Raven the fitting value is calculated according to Eq.(9)

$$Fitness = \sum_{j=1}^{k} \left[\sum_{x \in mij} |x - Cij| \right] \qquad (9)$$

where x is the input data vector; mij is the number of data for jth cluster from the ith Raven, and Cij represents jth cluster center for the ith Raven. Step 4: Regarding the obtained fitting value, the values of pbest and Chief are updated and the location of each Raven is updated according to Eqs. (10) and (11):

$$\mathrm{NewLoc} = \mathrm{Rv}(\mathrm{j}, :) + \mathrm{rand}(1, \mathrm{D}) + \mathrm{Rpcpt} \qquad (10)$$

$$\mathrm{NewLoc} = (\mathrm{Chief} \; - \; (1 - 2 \, * \, \mathrm{rand}\,(1, \mathrm{D}) \, * \, \mathrm{Rchief}) \qquad (11)$$

Step 5: The Kmeans algorithm calculates the Euclidean distance of the input data to all cluster centers for each Raven and the cluster of each data is specified. Step 6: The cluster centers are re-calculated and the location of each Raven is calculated based on the new centers. If the result is desirable, the algorithm is finished otherwise the third step is returned.

In the fourth phase, data modeling is done. For data modeling, there are several modelers including logistic regression(LR), support vector machine (SVM), decision tree(DT), neural network(NN), and k-nearest neighbors. The combined structure of collective learning is also an effective way of machine learning, in which the results of several simple modelers are combined together to improve the learning accuracy.

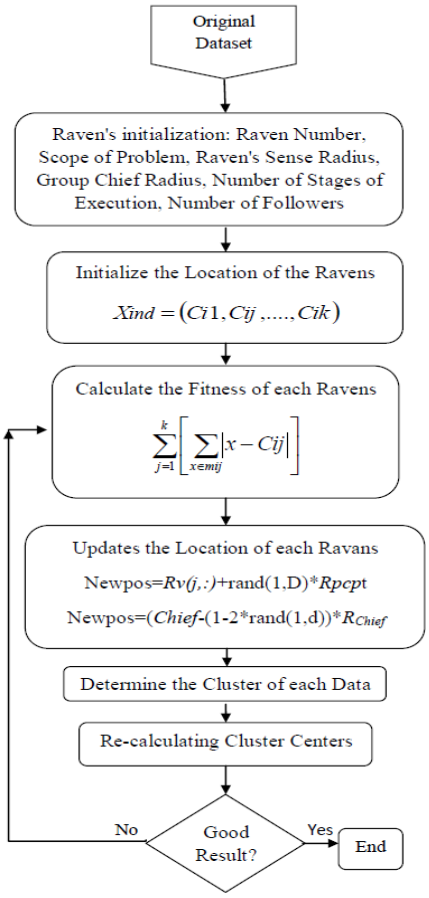
Original Dataset
Raven's initialization: Raven Number, Scope of Problem, Raven's Sense Radius, Group Chief Radius, Number of Stages of Execution, Number of Followers
Initialize the Location of the Ravens
$Xind = (Ci1, Cij ,...., Cik)$
Calculate the Fitness of each Ravens
$\sum_{j=1}^{k}\left[\sum_{x\in mij}|x - Cij|\right]$
Updates the Location of each Ravans
Newpos=$Rv(j,:)$+rand(1,D)*$Rpcpt$
Newpos=($Chief$-(1-2*rand(1,d))*R_{Chief}
Determine the Cluster of each Data
Re-calculating Cluster Centers
Good Result?
No
Yes
End

Figure 2: The proposed clustering algorithm

The classifiers'composition consists of 2 sections. The first part involves the creation of classifiers with an appropriate basis, the election of type, and the number of classifiers. The second part involves combining the output of classifiers in order to obtain the best result for the classification of patterns. In this research, the hybrid stacking algorithm (Wolpert, 1992) is used to increase the recognition rate of the hybrid classification system. Stacking consists of two phases. In the first phase, a set of basis classifiers is produced and in the second phase, a meta classifier is trained to combine the output of the basis classifiers. The most important issue in the stacking method is selecting features and an appropriate algorithm for learning at meta-level. Various algorithms such as SVN, NN, KNN, etc. are used at the meta-level. In this research, meta-level Gaussian accumulations (Xu et al., 2009) are used due to the nature of flexible non- parametric, providing predictive distribution and a simple and effective learning process. Due to study limitations and shortcomings associated with electricity theft discovery, it is highly required a method to cover these problems. Therefore, by studying the RRO optimization algorithm, its use in feature selection and clustering customers, as well as the use of feature correlation, local search and hybrid structure of collective learning, a framework, is proposed that aims to increase the recognition rate of customers who theft electricity.

5. Experiments and Analysis

In this section, according to the proposed framework, various tests are conducted in each phase, and the results are analyzed. Furthermore, the results of the proposed algorithm are compared with literature data.

5.1. Dataset

The ISSDA dataset was used to conduct tests and analyze the results by considering the importance of electricity theft and the need for a reference dataset with customers' power consumption data. The ISSDA (*Irish Social Science Data Archive*, 2012) is the consumption data of 5,000 smart meters for domestic and commercial Irish consumers, released by SEAI in 2012. The most important reasons for using this dataset in the present study are the large number of customers and their variety, as well as the length of the measurement time. In this dataset, for every customer, consumption data are saved every half an hour during the day, reducing the sampling rate per hour. The consumption data of 535 days are saved for each customer. One of the challenges of the discovery of electrical theft based on the pattern of customer consumption is the lack of a reference dataset from customers' abnormal consumptions. The ISSDA dataset only includes normal data from customers' consumption. In the first phase, after the data preprocessing, artificial attacks are created and an abnormal data sample of each customer's consumption is added to this dataset. These artificial attacks are in some way an indication of the possible manipulations of submitted data by customers in the declaration of their actual consumption. These artificial attacks can be made using the equations presented in Table 1. For example, for the ISSDA dataset in which the sampling rate is computed per hour, each sample is represented as y = {y1, ..., y24}, and 6 attack types are created.

Attack Formula	Description
$f_1(y_t)=a*y_t$, a=random(0.1,0.9)	Multiply all attributes in a constant random amount a
$f_2(y_t)=a_t*y_t$, a_t=random(0.1,0.9)	Multiply each attribute in a random amount
$f_3(y_t)=a_t*mean(y)$,a_t=random(0.1,0.9)	Multiply the mean of readings in a random amount
$f_4(y_t)=mean(y)$	Average readings during a day
$f_5(y_t)=y_{24-t}$	Changing the readings of a day.
$f_6(y_t)=\begin{cases} 0 & \forall t \in [ts,tf] \\ y_t & otherwise \end{cases}$	It is a by-pass attack that sends a value of zero at a specified time interval [ts, tf], otherwise, it sends the actual amount of consumption.

Table 1: Artificial attacks to generate the unusual samples in the ISSDA dataset

The artificial attack example created for each customer is added to the original dataset as manipulated and abnormal data.

5.2. Experiments

In this section, the experiments related to the phases of the proposed framework are explained. After normalizing and pre-processing data in the first phase, the effective features in the final detection are extracted in the second phase. The RRO-FS (raven roosting feature selection) technique is applied to the ISSDA dataset. Table 2 shows the accuracy of each algorithm on this dataset. It can be seen that the accuracy of the RRO-FS algorithm is better than the other algorithms. Table 3 presents the features extracted from the ISSDA dataset by GA, ACO, PSO, and RRO-FS algorithms.

Dataset	**GA**	**ACO**	**PSO**	**RRO-FS**
ISSDA	78.30%	75.82%	81.32%	88.11%

Table 2: The Accuracy of RRO-FS algorithm compared to GA, ACO, and PSO algorithms

Algorithm	**Selected attributes**
GA	f7,f10,f12,f13,f15,f18,f19,f20,f21,f23
ACO	f4,f6,f7,f8,f14,f15,f16,f22,f23,f24
PSO	f7,f8,f9,f12,f13,f14,f19,f20,f23,f24
RRO-FS	f11,f12,f14,f15,f19,f20,f21,f22,f23,f24

Table 3: Selected attributes from ISSDA using GA, ACO, PSO, and RRO-FS algorithms

In the third phase of the proposed framework, after selecting the desired features, the normal and abnormal data were clustered into three different clusters using the RRO-CL (raven roosting-clustering) technique, and the assigned cluster number represents the class label of each data. In this phase, in order to evaluate and analyze, the results of the proposed clustering algorithm RRO-CL were compared with the K-means, SOM, and PSO clustering algorithms. In our study, the Squared Euclidean distance (SED) index which is stronger and more robust than other indicators, was used to evaluate the cluster efficiency. The lower the SED value, the better result. Each test was carried out 30 times and the average was considered the final SED value, as shown in Table 4.

Dataset	**Clustering algorithm**			
	KMEMNS	SOM	PSO	RRO-CL
ISSDA	163.68	174.32	156.35	121.97

Table 4: The SED for clustering algorithms K-means, SOM, PSO, and RRO-CL

From the data presented in Table 4, the SED value of the proposed RRO-CL algorithm is lower than other algorithms. For example, the SED value is 121.97. This indicates a better efficiency of the RRO-CL technique in comparison to SOM, PSO, and K-means algorithms. After clustering the data and identifying the cluster number of each data, it is necessary to build a model for the classification of new data. A stacking hybrid algorithm that includes two phases was used in the proposed framework. Before using the hybrid method, the accuracy of the simple modelers such as NN, SVM, LR, and KNN for the ISSDA dataset is calculated. Table 5 elucidates the results of these predictions.

Algorithm **criterion**	NN	SVM	LR	KNN
Accuracy (%)	89.67	91.45	86.17	78.7

Table 5: The Accuracy of NN, SVM, LR and KNN Classifiers

Stacking is now used to predict the Accuracy of the ISSDA dataset with different meta-algorithms. Table 6 shows the prediction accuracy of different meta-algorithms in stacking.

In the proposed framework, the basic algorithms were used such as LR, SVM, KNN, and NN, and meta-Gaussian Processes. According to Table 6, its accuracy is higher than hybrid methods with KNN and NN meta algorithms. Furthermore, by comparing the results of Tables 5 and 6, it was found that the stacking hybrid method with the meta-Gaussian Processes algorithm is better than the basic and individual algorithms presented in Table 5. According to Table 5, the accuracy prediction value using the KNN algorithm is 78.7%, while that of KNN meta-algorithm shown in Table 6 is equal to 88.11%.

Basic Algorithms	Meta Algorithm	Accuracy(%)
LR, NN, SVM	KNN	88.11
LR, SVM,KNN	NN	91.56
LR, SVM, KNN, NN	Gaussian Processes	97.75

Table 6: Comparison of the accuracy with different meta algorithms in Stacking

Additionally, from Table 5, the accuracy prediction of the NN algorithm is 89.67%, while that of the combined method with meta-Gaussian Processes prediction is 97.75%. Therefore, these results disclose that the proposed framework has higher accuracy in comparison to other methods and can be regarded as a reliable framework. It should be noted that in the literature mostly the accuracy criterion is used as the benchmark for evaluation of classifiers performance, however, using this criterion solely is not suitable especially for the unbalance data due to the significance numerically difference between their number of positive and negative labels in the real world. Therefore, the f-score criterion, which is derived from the combination of recall and precision criteria, was used to evaluate the efficiency of classifiers. Ideally, its value is 1 and it is zero in the worst case. Another important criterion for assessing the efficiency of classifiers is the Area under Curve (AUC) criterion. This criterion represents the area below the Receiver Operating Characteristic (ROC) chart. The ROC curve is a two-dimensional curve whose X-axis is FPR and then the Y-axis is TPR. The more accurate the model, the AUC is closer to 1, and this number is closer to zero if the classifier performance is weaker. Table 7 shows the evaluation results of the F-score and AUC criteria for the proposed framework on the ISSDA dataset.

Dataset	Accuracy%	Recall%	Precision%	F-score%	AUC%
ISSDA	97.75	97.82	98.22	98.02	98.14

Table 7: The Accuracy, Recall, Precision, F-score, and AUC, using the proposed framework

The data in Table 7 shows that both the F-score and AUC criteria yielded a high percentage of desirability in the proposed framework. In another experiment, the results of the proposed algorithm were compared with the experimental results reported by (Jokar et al., 2016) on the ISSDA dataset, as shown in Table 9. In one of the experiments, only normal samples and in another experiment only the f2(yt) attack (see Table 1) were used for data training. Table 9 shows the results of these experiments.

Experiment	DR(%)	FPR(%)
Using Normal Samples for Training(Jokar et al., 2016)	76	29
Only h3(xt) Attack was Used for Training(Jokar et al., 2016)[a]	86	16
Using Normal Samples for Training in Proposed Algorithm	87	21
Only f2(yt) Attack was Used for Training in Proposed Algorithm	92	11
a h3(xt)=γtxt, γt=random (0.1, 0.8), x={x1,...,x24}, t=1,...,24		

Table 9: Comparison of the results of the proposed algorithm with (Jokar et al., 2016)

According to the results of Table 9, in the case of normal samples for training, the DR and FPR values in the proposed algorithm are 87% and 21%, respectively, however, those reported by Jokar were 76% and 29%, respectively. Moreover, it can be evident that the DR and FPR values are 86% and 16%, respectively, for the case that only h3(xt) was used for training by Jokar, while those in the proposed algorithm were 92% and 11%, respectively, indicating the improvement of these values in the proposed algorithm.

In another experiment, the AUC of the proposed framework was compared with research (Peng et al., 2021). According to the results of Table 10, the AUC of the proposed framework is 98.14%.

Method	**AUC(%)**
Research (Peng et al., 2021)	81.50
Proposed Framework	98.14

Table 10: Comparison of the AUC of the proposed algorithm with research (Peng et al., 2021)

In the last experiment, the accuracy of the proposed framework was compared with the results of literature articles Research(Khan et al., 2020), (Nabil et al., 2019) and (Ibrahem et al., 2019) on the ISSDA dataset. In this experiment, (Jokar et al., 2016) research attacks (6 attacks) were used. According to Table 11, it can be seen the Accuracy of the proposed method is higher than two.

Method	$\boldsymbol{Accuracy}$**(%)**
ETDFE(Ibrahem et al., 2019)	93.36
Model1(Nabil et al., 2019)	91.8
Model2(Nabil et al., 2019)	90.2
Research(Khan et al., 2020)	95
Proposed Framework	97.75

Table 11: Comparison of the results of the proposed algorithm with other algorithms

The results of Table 11 indicate that the value of the Accuracy of the proposed algorithm is higher than others works. Therefore, according to the proposed framework, the percentage of accurate prediction of customers who manipulate their metrics and do not send their actual consumption data to the center has improved.

6. Conclusion

In this research, due to concerns related to electricity theft by illegal and dishonest customers, a four-stage framework was proposed for the discovery of abnormal. In the first phase, normalization and pre-processing of data were performed. Then, in the second phase, because of the significance of key feature selection, effective features were selected using the RRO algorithm. Due to the fact that there is no reference dataset covering both normal and abnormal samples of customer consumption, for the ISSDA dataset, abnormal samples were created using artificial attacks before the second phase. In the third phase, inspired by the Raven algorithm, each of the normal and abnormal samples was divided into three clusters and the cluster number was labeled as the data class. In the last phase, data modeling was performed using the Stacking hybrid method, which is a two-step algorithm. In the first step, simple algorithms such as LR, SVM, KNN, and NN algorithms were used and in the second step, the meat-Gaussian Processes algorithm was used for data modeling. The results of each phase revealed that the proposed algorithms had high performance. To evaluate the efficiency of the proposed algorithm, the Accuracy, F-measure, and AUC criteria were used. The results of the proposed algorithm were compared with the literature research and it was found that the results of the proposed algorithm are better. Since robust theft detection is important, it is recommended that the peak periods of consumption in different seasons of the year are involved in the detection process as future work.

References

Feng, Z., Huang, J., Tang, W. H., & Shahidehpour, M. (2020). Data mining for abnormal power consumption pattern detection based on local matrix reconstruction. *International Journal of Electrical Power and Energy Systems*, *123*(February), 106315. https://doi.org/10.1016/j.ijepes.2020.106315

Ibrahem, M. I., Nabil, M., Fouda, M. M., Mahmoud, M., Alasmary, W., & Alsolam, F. (2019). Efficient Privacy-Preserving Electricity Theft Detection with Dynamic Billing and Load Monitoring for AMI Networks. *IEEE Access, 7*, 96334–96348. https://doi.org/10.1109/ACCESS.2019.2925322

Irish Social Science Data Archive. (2012). https://www.ucd.ie/issda/data/commissionforenergyregulationcer/

Jeyakumar, J., & Devaraj, D. (2018). Machine learning algorithm for efficient power theft detection using smart meter data. *International Journal of Engineering and Technology(UAE), 7*, 900–904.

Jokar, P., Arianpoo, N., & Leung, V. C. M. (2016). Electricity theft detection in AMI using customers' consumption patterns. *IEEE Transactions on Smart Grid*, *7*(1), 216–226. https://doi.org/10.1109/TSG.2015.2425222

Kabir, M., Shahjahan, M., & Murase, K. (2011). A new local search based hybrid genetic algorithm for feature selection. *Neurocomputing*, *74*, 2914–2928. https://doi.org/10.1016/j.neucom.2011.03.034

Khan, Z. A., Adil, M., Javaid, N., Saqib, M. N., Shafiq, M., & Choi, J. G. (2020). Electricity theft detection using supervised learning techniques on smart meter data. *Sustainability (Switzerland), 12*(19), 1–25. https://doi.org/10.3390/su12198023

Li, S., Han, Y., Yao, X., Yingchen, S., Wang, J., & Zhao, Q. (2019). Electricity Theft Detection in Power Grids with Deep Learning and Random Forests. *Journal of Electrical and Computer Engineering*, *2019*. https://doi.org/10.1155/2019/4136874

Monirul Kabir, M., Monirul Islam, M., & Murase, K. (2010). A new wrapper feature selection approach using neural network. *Neurocomputing*, *73*(16), 3273–3283. https://doi.org/https://doi.org/10.1016/j.neucom.2010.04.003

Nabil, M., Ismail, M., Mahmoud, M. M. E. A., Alasmary, W., & Serpedin, E. (2019). PPETD: Privacy-Preserving Electricity Theft Detection Scheme with Load Monitoring and Billing for AMI Networks. *IEEE Access*, *7*, 96334–96348. https://doi.org/10.1109/ACCESS.2019.2925322

Nazmul Hasan, M., Toma, R. N., Nahid, A. Al, Manjurul Islam, M. M., & Kim, J. M. (2019). Electricity theft detection in smart grid systems: A CNN-LSTM based approach. *Energies*, *12*(17), 1–18. https://doi.org/10.3390/en12173310

Peng, Y., Yang, Y., Xu, Y., Xue, Y., Song, R., Kang, J., & Zhao, H. (2021). Electricity Theft Detection in AMI Based on Clustering and Local Outlier Factor. *IEEE Access*, *9*, 107250–107259.

https://doi.org/10.1109/ACCESS.2021.3100980

Sowndarya, R., & Latha, D. P. (2017). *An Artificial Intelligent Algorithm for Electricity Theft Detection in AMI*.

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259. https://doi.org/https://doi.org/10.1016/S0893-6080(05)80023-1

Xu, Z., Kersting, K., & Tresp, V. (2009). *Multi-Relational Learning with Gaussian Processes.*

Yeckle, J., & Tang, B. (2018). Detection of Electricity Theft in Customer Consumption Using Outlier Detection Algorithms. *2018 1st International Conference on Data Intelligence and Security (ICDIS)*, 135–140. https://doi.org/10.1109/ICDIS.2018.00029

Zhang, W., Dong, X., Li, H., Xu, J., & Wang, D. (2020). Unsupervised Detection of Abnormal Electricity Consumption Behavior Based on Feature Engineering. *IEEE Access*, *8*, 55483–55500. https://doi.org/10.1109/ACCESS.2020.2980079