

## Extraction and selection of statistical harmonics features for electrical appliances identification using k-NN classifier combined with voting rules method

Fateh GHAZALI<sup>1</sup>, Abdenour HACINE-GHARBI<sup>1</sup>, Philippe RAVIER<sup>2\*</sup> ,  
Tayeb MOHAMADI<sup>3</sup>

<sup>1</sup>LMSE Laboratory, University of Bordj Bou Arreridj, Bordj Bou Arreridj, Algeria

<sup>2</sup>PRISME Laboratory, University of Orleans, Orléans, France

<sup>3</sup>Department of Electronics, Faculty of Technology, Ferhat Abbas University, Setif, Algeria

Received: 12.12.2018

Accepted/Published Online: 26.03.2019

Final Version: 26.07.2019

**Abstract:** In this paper, we propose a novel framework for electrical appliances identification using statistical harmonic features of current signals and the use of the k-NN classifier combined with a voting rule strategy. Harmonic coefficients are computed over time using short-time Fourier series of the current signals. From these sequences of coefficients, the mean, standard deviation, skewness, and kurtosis are computed, which provide the statistical harmonic features. This framework has three novelties: (i) selecting the best combination of statistical measures in the sense of classification rate (CR); (ii) combining the k-NN classifier with the voting rule method in order to search for the best number of voting vectors; and (iii) selecting relevant features for the task of appliances identification by using one of the relevant feature selection algorithms based on mutual information. Results evaluated on the Plaid dataset clearly show that the mean and standard deviation statistics combination gives the best CR of 92% with 500 features and gives the minimal computing time compared to the system based on HMM models. Moreover, combining the k-NN classifier with the voting rule using the above features increases the CR up to 95%. Using this combination, the results also show that an increase of the training dataset size further improves identification performance results in terms of precision, sensitivity, and F-score. A feature selection procedure based on joint mutual information strategy shows that using a selected subset of five features is sufficient, giving similar CR results to those obtained using the total number of features, whatever the training dataset size.

**Key words:** Electrical appliance identification, harmonics analysis, short-time Fourier series, statistical features extraction, k-nearest neighbors, voting rules method, filters features selection, mutual information

### 1. Introduction

Balancing production and electricity consumption is a daily principal concern of electricity suppliers. Today, new challenges are the reduction of global electricity production without compromising the electrical needs of the consumers and guaranteeing a high quality of electrical supplying service. This is a necessity for economic and environmental reasons and this is now possible thanks to smart grid deployment. Indeed, balance can be optimized whenever the demand and consumption data over the smart grid network are known. For electricity suppliers, electricity production optimization may involve electrical shedding, adapted billing for customer uses, or electricity exchanges with other suppliers. For customers, having access to consumption information gives the opportunity to optimize electricity consumption and energy bills. In this context, detailed consumption

\*Correspondence: [philippe.ravier@univ-orleans.fr](mailto:philippe.ravier@univ-orleans.fr)

information can be obtained with nonintrusive load monitoring (NILM) systems. The purpose of a NILM system is to disaggregate energy consumption at the main, which necessitates electrical appliance identification, achieved using current features.

The implementation of feature extraction methods from the current measurements for electrical appliance identification was widely discussed in the literature [1–4]. In [1], the authors applied a hidden Markov model (HMM)-based identification system where the current signals are represented by sequences of vectors composed of short-time Fourier series (STFS) coefficients. Also in [2], the authors considered the wavelet cepstral coefficients (WCCs) as descriptors of electrical appliances and compared their performances with those of harmonic descriptor STFS using the HMM classifier.

In [3], the authors proposed to apply a parametric model in order to represent each appliance's current signal. A single vector composed of 14 parameters was used for appliance clustering and identification using the K-means algorithm. This compact representation of signals reduces the memory storage and minimizes the classification execution time. In the literature, several papers also represent signals [5–7] in a compact way using a statistical feature extraction method. The statistical feature extraction method allows transforming a harmonic vectors chain of a signal into a single statistical feature vector. Practically, identification results have to regularly be delivered since meters continuously measure the current signals. Accordingly, statistical features such as mean, standard deviation, skewness, and kurtosis should be measured on temporal segments of fixed duration, which leads to converting each signal into a sequence of statistical features vectors. One possible classification method for current signals consists of classification of each statistical feature vector using the k-nearest neighbors (k-NN) classifier, then applying the voting rule to the sequence of the class's indices to make a decision about the signal class [8].

The method of voting rules originates from the social sciences [9]. This method has become a widely used technique in various disciplines, namely in engineering disciplines, and especially in pattern recognition [10–14]. The voting rules procedure is an alternative approach applied to a set of possible statistical vectors. The motivation is to let them act as individual classifiers and combine these classifiers as competitors for the final classification [13] using a voting rules procedure. The method is interesting because of its simplicity and efficiency.

The present work consists of identifying electrical appliances from their signatures (electric currents). It is clear that such an identification system requires two main phases, one for learning and the other for testing. Many models and mathematical methods can be developed for the identification system, such as HMM [15], support vector machines (SVMs) [16], artificial neural networks (ANNs) [17], and k-NN [18]. These models require information extracted from the current signals that are considered as relevant descriptors of the appliance. In this work, we propose to extract a vector of statistical features of harmonics representing the complete signal and to investigate the best combination of statistical measures. We also study the best number of statistical vectors in the voting rule procedure.

However, classification problems often have a large number of features, but not all of them are useful for classification. Irrelevant and redundant features may even reduce the classification accuracy. Therefore, feature selection, also known as variable selection or attribute selection, is proposed as a data preprocessing step to reduce or eliminate irrelevant and redundant features [19]. In this context, the application of a feature selection technique is essential.

Feature selection procedures can globally be grouped into two approaches that are either dependent on

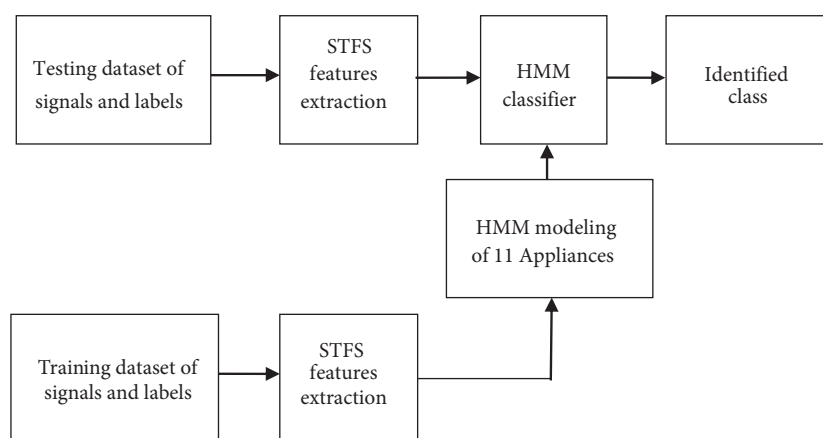
the classifiers (wrapper methods) or independent of the classifiers (filter methods) [20]. Wrapper methods look for the space of the features subset using classification accuracy as a measure of utility for a candidate subset. It has the disadvantage of considerable computing expense since a classification system has to be built for each subset trial. In contrast, filter methods [21] define a heuristic scoring criterion that serves as an indirect measure of classification accuracy and evaluate the statistics of the data independently of any classifier [21, 22]. Thus, a filter feature selection algorithm searches for the optimal feature subset in the search space based on a certain evaluation criterion, which is independent of any learning/classification algorithm.

Several filter methods are present in the literature [19, 22–24]; among them, the mutual information (MI)-based methods are the most used. In this work, we have applied a filter selection algorithm with the joint mutual information (JMI) criterion, which uses MI estimation as a relevance measure of the features. We selected the JMI strategy because of its good trade-off in terms of accuracy, stability, and flexibility with small data samples [23].

## 2. State-of-the-art appliance identification and feature extraction methods

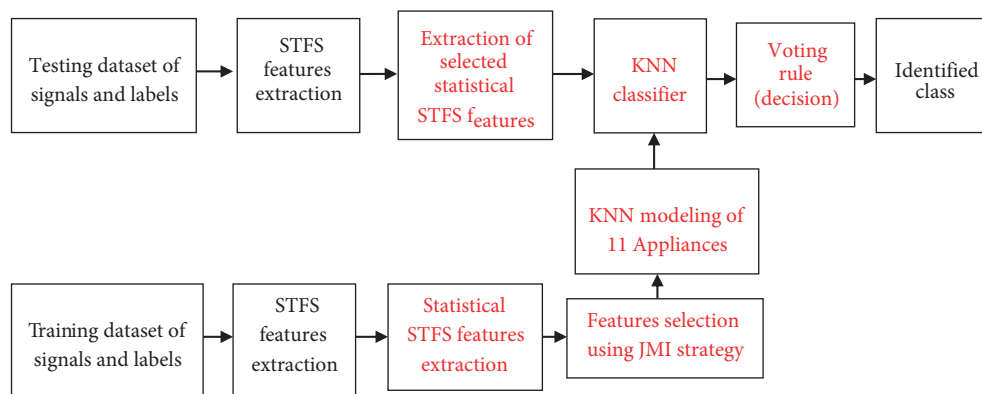
The purpose of NILM systems is to monitor and control domestic and industrial energy consumption, using only one meter at the input of the network supply. The general framework of NILM starts from measurements made on the input of total electricity consumption to finally break it down into individual contributions of each appliance.

Hart introduced the first work on NILM methods in 1989 [25, 26]. He was the first to analyze the variations of the total power to identify electrical appliances, but the method does not allow identifying some appliances (multistate appliances as washing machines). The basic works of Hart launched several research works: Sultanem [27] used current variations and active and reactive power as relevant parameters for identification. HMM was first used in [28]. In order to identify the structural nature of electrical appliances, the authors of [29] used current harmonics. Leeb proposed the first works that gave interest to the transient part of the current inrush in 1993 [30]. He investigated the form of active and reactive power transients. Many other works followed [31–39] with the state of the art presented in [40].



**Figure 1.** Flowchart of the identification system proposed in [2].

Generally, several classification methods have been proposed in appliance identification systems with the aim of performance improvement, either in terms of complexity or in terms of accuracy. The electrical



**Figure 2.** Flowchart of the identification system based on k-NN classifier combined with voting rules method. Blocks with red characters highlight novelties added with respect to Figure 1.

appliance classification methods can be grouped into two categories. In the first category, a features extraction stage converts each signal in a sequence of features vectors. Next, a classifier identifies the class of this sequence. In [1] and [2], the authors used STFS features extraction methods for electrical appliance identification based on the HMM classifier. A flowchart of the system proposed in [2] is illustrated in Figure 1. The second category uses and classifies a single feature vector (instead of a sequence of feature vectors) that represents the complete signal. The single feature vector is used in the input of a suitable classifier like SVM [16], ANN [41], or k-NN [42]. This second category can reduce complexity in terms of computation cost and memory space without decreasing the accuracy. In this work, we chose the second category using statistical features applied on STFS features (harmonics) in order to extract one feature vector and to classify it using the k-NN classifier for appliance identification from the current signal. Practically, though, the electric current is continuously delivered by meters and the feature vectors are delivered online. This means that the statistical features are estimated only after a fixed duration period. Then, considering a sequence of such consecutive periods, each signal is converted into a sequence of statistical features vectors that are classified into a sequence of the class's indices. Hence, the voting rule is proposed to be applied on this sequence using a k-NN classifier to make a global decision about the signal class.

### 3. Proposed method

In this work, we propose to use statistical harmonics vectors (mean, standard deviation, skewness, and kurtosis) of electric current and to evaluate these descriptors on k-NN classifiers combined with the voting rules method.

The first contribution of this paper is the use of statistical features as a new descriptor for appliance identification. Statistical features have the ability of catching dynamic behaviors in a very compact way. This use requires a new schema of the identification system. The second contribution is thus the proposal of a new scheme based on the combination of the k-NN classifier with the voting rule procedure (Figure 2). In order to search for the most relevant features among a large set of possible ones, we finally apply a filter feature selection algorithm based on the JMI strategy. This feature selection procedure is a third contribution.

#### 3.1. Harmonic analysis

The universal method of harmonic analysis is that of the Fourier transform (FT). In the discrete case, the vector of the current signal  $s[n]$  of length  $L$  can be analyzed using the discrete Fourier transform (DFT):

$$S[k] = \sum_{n=0}^{L-1} s[n] \exp\left(\frac{-j2\pi kn}{L}\right), k = 0, \dots, L-1. \quad (1)$$

Our approach for vectors chain extraction is primarily based on the division of each electrical current signal in overlapped windows by a half period and then computation of the discrete Fourier series (DFS) in each window. The DFS decomposition is [1, 43]:

$$s[n] = \sum_{k=0}^{N-1} C_k \exp\left(\frac{j2\pi kn}{N}\right), \quad (2)$$

where  $N$  is the period of  $s[n]$  in samples and  $C_k$  are the Fourier series coefficients expressed as:

$$C_k = \frac{1}{N} \sum_{n=0}^{N-1} s[n] \exp\left(\frac{-j2\pi kn}{N}\right). \quad (3)$$

The obtained coefficients are called STFS coefficients. They can be stacked in an STFS coefficient vector (calculated on a window), which is called an observation. Thus, each electric current signal is represented by a vector sequence, which can be put into a harmonic matrix  $MH(P, R)$  representing the complete signal, where  $P$  is the harmonic number (its optimal value will be discussed later) and  $R$  is the number of windows related to the duration of the input current signal.

### 3.2. Statistical features

The originality of our approach lies in the computation of the statistical features, which are vectors, computed from the harmonic matrix obtained by the application of the STFS on the set of windows of the electric current signal. Computation of the statistical features is achieved on the lines of the harmonic matrix resulting in statistical vectors with length equal to  $P$  (number of harmonics). We selected statistical measures up to the fourth order computed on the absolute values of the matrix:

- the mean

$$\mu = E[MH] \quad (4)$$

is estimated as

$$\mu(p) = \frac{1}{R} \sum_{r=1}^R MH(p, R), p = 1, \dots, P; \quad (5)$$

- the standard deviation

$$\sigma = \sqrt{E[(MH - \mu)^2]} \quad (6)$$

is estimated as

$$\sigma(p) = \sqrt{\frac{1}{R} \sum_{r=1}^R (MH(p, R) - \mu(p))^2}; \quad (7)$$

- the skewness

$$Sk = E\left[\left(\frac{MH - \mu}{\sigma}\right)^3\right] \quad (8)$$

is estimated as

$$Sk = \frac{\frac{1}{R} \left( \sum_{r=1}^R (MH(p, R) - \mu(p))^3 \right)}{(\sigma(p))^3}; \text{ and} \quad (9)$$

- the kurtosis

$$Ku = E \left[ \left( \frac{MH - \mu}{\sigma} \right)^4 \right] \quad (10)$$

is estimated as

$$Ku = \frac{\frac{1}{R} \left( \sum_{r=1}^R (MH(p, R) - \mu(p))^4 \right)}{(\sigma(p))^4}. \quad (11)$$

Finally, each appliance electric current is represented by a single vector  $V = [V(1) \dots V(P)]$ , which is the concatenation of  $P$  individual vectors  $V(p)$ , each one being composed of four temporal statistical values characterizing the harmonic of order  $p$ :

$$V(p) = [\mu(p) \ \sigma(p) \ Sk(p) \ Ku(p)]. \quad (12)$$

For example, a pure resistive load powered by a pure sine wave should theoretically produce a collection of zero  $V(p)$  vectors except  $V(1)$  where the first component is the current magnitude, the other ones being zero.

### 3.3. k-NN combined with voting rule method

The voting method is one of the most effective techniques in pattern recognition, where the final classification decision is the majority decision reached among a set of several concurrent systems [44, 45]. In [44], the authors gave an overview and comparison of voting methods for pattern recognition. The voting method is also applied in electrical appliance identification using a fusion of several classifiers [45]. Unlike applying voting rules between several classifiers, as done in [45], we propose in our framework to apply voting rules for a single k-NN classifier, between different numbers of statistical vectors, for the same observation duration. The purpose is to improve the classification rate of the identification system by the choice of number of statistical vectors that contribute to the vote. Theoretically increasing the number of vectors gives better results in the voting sense, but at the same time, the constraint of statistical data will decrease the results. A compromise therefore has to be found between the number of vectors and the statistical data. Section 4.3 explains the optimal number of statistical vectors in the sense of classification rate (CR). Figure 2 shows our identification system scheme and the combination synopsis of the k-NN classifier with the voting rule method.

### 3.4. Feature selection

Feature selection is an important step in pattern recognition, especially in the case of high dimensions. It has the aim of identifying the relevant features subset that keeps maximum information and minimum redundancy to explain the class variable, from the global space of features. Feature selection has the advantages of reducing the computing time and space memory and probably improves the accuracy by avoiding the curse of dimensionality. In this work, we use the mutual information as a measure of features relevance. For understanding this concept, we present in the following a brief review of two basic elements of information theory, which are entropy and mutual information [46].

### 3.4.1. Entropy and mutual information

The entropy  $H(X)$  is a measure of the average information amount contained in the random variable  $X$  [47]. It can be defined as a measure of the uncertainty of variable  $X$ . Let  $X$  be a discrete random variable and  $p(x)$  its probability function. The entropy of  $X$  is defined by [47]:

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log_2(p(x)). \quad (13)$$

In the continuous case, the entropy  $h$  is defined by:

$$h(X) = - \int_{-\infty}^{+\infty} f_x(x) \log_2(f_x(x)) dx, \quad (14)$$

where  $f_x(x)$  is the probability density function (pdf) of the random variable  $X$ .

The entropy definition can be extended to joint entropy that takes more than one variable [47]. In the case of two discrete random variables  $X, Y$  with joint probability function  $p(x, y)$ , the joint entropy  $H(X, Y)$  is defined as:

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log_2(p(x, y)). \quad (15)$$

In the continuous case, the joint entropy is given as:

$$h(X, Y) = - \int \int_{-\infty}^{+\infty} f_{xy}(x, y) \log_2(f_{xy}(x, y)) dx dy, \quad (16)$$

where  $f_{xy}(x, y)$  is the joint pdf of the random variables  $X, Y$ .

Another important element in information theory is the mutual information  $I(X; Y)$ , which is the reduction in uncertainty about one random variable due to the knowledge of the other [47]. It can be interpreted as a measure of the shared information between two random variables  $X, Y$ . The mutual information between two random variables  $X$  and  $Y$  is defined by:

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log_2 \left( \frac{p(x, y)}{p(x)p(y)} \right). \quad (17)$$

In the continuous case, the mutual information  $\mathcal{I}$  is defined by:

$$\mathcal{I}(X, Y) = \int \int_{-\infty}^{+\infty} f_{xy}(x, y) \log_2 \left( \frac{f_{xy}(x, y)}{f_x(x)f_y(y)} \right) dx dy, \quad (18)$$

where  $f_{xy}(x, y)$  is the joint distribution of  $X, Y$  and  $f_x(x)$  and  $f_y(y)$  are the marginal distributions.

Practically, the histogram approaches have been mostly used for pdf estimation from data for their simplicity and low complexity [22]. However, the mutual information suffers practically from the bias estimation caused by the insufficient representation of the pdf by the histogram approach, and by the small number of samples. Furthermore, the finite number of samples can induce a large bias of MI estimation in high dimensions. In [22], the authors proposed new formulas of bin number of histograms in order to reduce the bias and the mean square error (MSE) of MI and entropy estimation.

### 3.4.2. Feature selection based on joint mutual information strategy

Theoretically, feature selection aims to select from an initial set of  $n$  features,  $F = \{Y_1, Y_2, \dots, Y_n\}$ , a subset of  $k$  relevant features  $S = Y_{P_1}, Y_{P_2}, \dots, Y_{P_k}$  that produces the maximal mutual information with the following class variables:

$$S = \arg \max_{S \in F} I(C; S). \quad (19)$$

This can be performed using the forward ‘greedy’ algorithm, which selects at each iteration  $j$  one feature  $Y_{P_j}$  that verifies the following equation:

$$Y_{P_j} = \arg \max_{Y_i \in F - S_{j-1}} I(C; S_{j-1}, Y_i), \quad (20)$$

where  $S_{j-1} = Y_{P_1}, Y_{P_2}, \dots, Y_{P_{j-1}}$ . Because of the chaining rule  $I(C; S_{j-1}, Y_i) = I(C; Y_i) + I(C; S_{j-1} \setminus Y_i)$  that can be developed in  $I(C; S_{j-1}, Y_i) = I(C; Y_i) + I(C; S_{j-1}) - I_3(C; S_{j-1}; Y_i)$ , Eq. (20) can be reduced to:

$$Y_{P_j} = \arg \max_{Y_i \in F - S_{j-1}} [I(C; Y_i) - I_3(C; S_{j-1}; Y_i)], \quad (21)$$

where  $I_3(C; Y_i; S_{j-1})$  represents the redundancy between the feature  $Y_i$  and  $S_{j-1}$ .

Practically,  $I(C; S_{j-1}, Y_i)$  or  $I_3(C; S_{j-1}; Y_i)$  cannot be accurately estimated with increasing numbers of features. Hence, several heuristic strategies have been proposed like MIM, CMIM, MIFS, MRMR, CMI, JMI, DISR, CIFE, TMI, and ICAP [7]. In this work, we used the JMI strategy [48], which considers MI between three variables instead of multivariate MI of Eq. (21):

$$Y_{P_j} = \arg \max_{Y_i \in F - S_{j-1}} \left[ I(C; Y_i) - \frac{1}{j-1} \sum_{k=1}^{j-1} I_3(C; Y_i; Y_{P_k}) \right], \quad (22)$$

where the  $I_3$  terms are estimated using  $I_3(C; Y_i; Y_{P_k}) = I(Y_i; Y_{P_k}) - I(Y_i; Y_{P_k} \setminus C)$ .

## 4. Experiences and results

In this section, we present the different experiments that we carried out to evaluate the performance of our algorithm for statistical harmonics features extracted from the electric currents of the Plaid dataset [49]. A k-NN classifier was also used. We tested the following:

1. The application of all possible combinations of the statistical parameters in order to extract the optimal combination, using the k-NN classifier with 15 possible combinations as given in Table 1. Performance results are established based on the identification results evaluated using the classification rate ( $CR$ ) defined as:

$$CR(\%) = \frac{T - M}{T} \times 100, \quad (23)$$

where  $T$  is the total number of tested waveforms (each one representing an appliance) given to the input of the classifier and  $M$  is the number of misclassified tested waveforms.

2. The voting rule method, with our k-NN classifier applied on the optimal combination of the statistical parameters obtained in Test 1. In the Plaid dataset, the sampling frequency is  $F_s = 30$  kHz. In our experimental phases, the [0:15 kHz] frequency band practically gives 250 harmonics, knowing the 60-Hz power line frequency.



3. The dataset distribution configuration organization effect, using other performance evaluation metrics (sensitivity, precision, F-score, and confusion matrix).
4. The dimension reduction, because the number of parameters (harmonics) is very large, probably causing the effect of the curse of dimensionality. In order to overcome this problem, we apply the forward greedy algorithm combined with the JMI strategy to select the relevant features and reduce the dimensions.

**Table 1.** The best combination for statistical features using k-NN classifier is MEAN  $\mu$  and STD  $\sigma$ .

	Feature number	CR at 50 harmonics in	CR at 250 harmonics in
$\mu$	250	86.96	84.73
$\sigma$	250	84.35	83.79
Sk	250	75.79	72.62
Ku	250	67.59	67.78
$\mu, \sigma$	500	93.10	92.36
$\mu, \text{Sk}$	500	75.79	72.62
$\mu, \text{Ku}$	500	67.59	67.78
$\sigma, \text{Sk}$	500	75.79	72.62
$\sigma, \text{Ku}$	500	67.59	67.78
Sk, Ku	500	68.52	68.34
$\mu, \text{Sk, Ku}$	750	68.52	68.34
$\mu, \sigma, \text{Ku}$	750	67.59	67.78
$\sigma, \text{Sk, Ku}$	750	68.52	69.34
$\mu, \sigma, \text{Sk}$	750	75.79	72.62
$\mu, \sigma, \text{Sk, Ku}$	1000	68.52	68.34

#### 4.1. Presentation of the Plaid dataset

Our analysis is based on the Plaid dataset [49] composed of 1074 recordings of currents and voltages of 11 types of electrical appliances from a variety of 56 manufacturers. The recordings are made at a frequency of 30 kHz. Table 2 presents a summary of the Plaid dataset and the different types and numbers of appliances.

#### 4.2. Best combination of statistical features

The purpose is to select the most relevant combination among the four types of statistical features (mean, standard deviation, skewness, kurtosis) using the k-NN classifier (taking k by default equal to 1). The statistical parameters are computed on the complete signal. This leads to converting each signal into a single vector. Therefore, k-NN classification has been performed without application of the voting rule. Table 1 shows 15 possible combinations of statistical features and their CR (%) with 50 and 250 harmonics numbers, respectively.

It is clear from Table 1 that the best combination is that of the mean and the standard deviation, which gives the highest classification rates of up to 93.10% for 50 harmonics and 92.36% for 250 harmonics. The combination of the mean with standard deviation is therefore the optimal experimental one in the sense of classification rate. These statistic features are thus chosen in the following study.

Furthermore, the application of the HMM models of 7 states, each one being associated to a GMM model of 3 Gaussians, on the harmonic vector chain [2] gives a classification rate of 93.30% at 250 harmonics, but with

**Table 2.** Summary of the appliances found in the Plaid dataset.

N	Appliance type	Number of instances
1	Compact fluorescent lamp	175
2	Vacuum cleaner	38
3	Hair-dryer	156
4	Microwave	139
5	Air conditioner	66
6	Laptop	172
7	Fridge	38
8	Incandescent light bulb	114
9	Fan	115
10	Washing machine	26
11	Heater	35
	Total	1074

a long computing time of about 59 s (without considering the feature extraction time). On the other hand, the k-NN classifier (with the statistical features  $\mu, \sigma$ ) gives a classification rate of 92.36% with a calculation time of 0.41 s, using a computer with an Intel Core i3 processor and 6 GB of RAM memory using MATLAB language. This result shows that the proposed approach presents a comparable CR to that of the HMM approach but gives the minimal calculation time.

Further, Table 1 also shows that in most combinations, the use of only 50 harmonics gives better CR results than the use of 250 harmonics. This is probably caused by the curse of dimensionality. In fact, the optimal combination of 500 statistical features needs a dimensionality reduction. This can be performed using a feature selection algorithm, which will be the subject of Section 4.6.

### 4.3. The optimal number of statistical vectors for voting rule method

In this experiment, we applied our identification algorithm to the combination of the statistical features (mean, standard deviation) using 250 harmonics, for a period of 1 s, so 60 harmonic vectors in windows are recovered by a half, which gives 120 harmonic vectors. By dividing the 1-s duration into consecutive segments, the number of statistical vectors can vary from 1 to 60, as shown in Table 3.

For example, three statistical vectors consider three k-NN classifiers, one for each segment, and the voting rule makes its decision based on the most frequent decision result given by the classifiers. Table 3 shows the CR obtained with 250 harmonics, with different numbers of statistical vectors. The table also shows the elapsed time [seconds] for each statistical vector for 250 harmonics.

From Table 3, it can be seen that 8, 12, and 15 statistical vectors (which are also the numbers of classifiers) participating in the vote give the best CR of 94.97% for all numbers using 250 harmonics. This also respectively corresponds to statistical vectors being computed on  $R = 15, 10,$  and  $8$  windows. Moreover, the results show that the computational time when using 8 statistical vectors gives the shortest running time (2.03 s) compared with 2.83 s for 12 statistical vectors and 3.39 s for 15 statistical vectors (same hardware configuration as in Section 4.2).

From these results, we conclude that the optimal number of statistical vectors that participate in the vote is eight. This means that the statistical vectors evaluated in 15 time windows practically give the optimal classification rate at 250 harmonics (94.97%) and the short time running of 2.03 s. Tables 1 and 3 clearly show that the large number of harmonics (250) probably causes the effect of the dimensionality curse. We propose in the next section to study the effect of the training and testing datasets on identification performance.

**Table 3.** The best number of statistical vectors given by the voting rule method, using 250 harmonics.

Number of statistical vectors	CR (%)	Elapsed time [seconds]
1	92.36	0.47
2	94.41	0.80
3	94.78	0.98
4	94.22	1.20
6	94.22	1.68
8	94.97	2.03
10	94.78	2.46
12	94.97	2.83
15	94.97	3.39
20	94.78	4.46
30	94.59	6.50
40	94.78	9.04
60	94.78	11.79

#### 4.4. Dataset size effect on identification performance

In order to study the effect of dataset distribution, Table 4 shows the effect of the different dataset distribution configurations on the CR, taking 8 statistical vectors (optimal choice in Table 3).

**Table 4.** Effect of the dataset distribution on the CR results.

Training/Testing	50%/50%	60%/40%	80%/20%
CR (%)	94.97	97.20	97.20

Table 4 shows that increasing the training dataset size improves the identification performance results. This can be justified by a higher amount of information that is added for describing the different classes when using large data samples.

#### 4.5. Performance evaluation

The classification rate (CR), also called the recognition rate or mean accuracy, gives a global view of the performance for our system. For deeper evaluations, we use other evaluation metrics, like sensitivity, precision, f-score, and confusion matrix, frequently used in the literature [41, 50–52], taking the case of 8 vectors (best

voting rule case). These metrics give the performance for each electrical appliance, so we can consider them as local evaluation criteria [38, 41, 50]:

- sensitivity  $S$  is defined as

$$S = \frac{TP}{TP + FN} * 100,$$

- precision  $P$  is defined as

$$P = \frac{TP}{TP + FP} * 100, \text{ and}$$

- F-score  $F_s$  is defined as

$$F_s = \frac{2 * S * P}{S + P},$$

where  $TP$  is the number of true positives, i.e. the positive samples correctly classified;  $FP$  is the number of false positives, i.e. the negative samples incorrectly classified;  $FN$  is the number of false negatives, i.e. the positive samples incorrectly classified; and  $TN$  is the number of true negatives, i.e. the negative samples correctly classified. The metric F-score is a weighted average of precision and sensitivity [41]. These metrics can be obtained from the confusion matrix, which is an useful tool for deeply evaluating the efficiency of supervised classification systems. The columns of this matrix represent the sensitivity of each appliance and the rows represent the precision of each electrical appliance [41]. The main diagonal gives for each class the correct classification number. The nonzero off-diagonal values are incorrect classification numbers [50].

Table 5 shows the results obtained for the different metrics (sensitivity  $S$ , precision  $P$ , and F-score  $F_s$ ) of each electrical appliance type and for different dataset distributions. Results obtained show the detailed performance of our framework, for each electrical appliance. For example, microwave and fridge were classified with 100% values for the three metrics and for all dataset distribution configurations, demonstrating an excellent identification performance. The other appliances also show very acceptable performance, almost all above 90%, except vacuum cleaner with remarkably weak results falling to 56% for F-score. Furthermore, Table 5 shows that an increase of the training dataset size improves identification performance results in terms of precision, sensitivity, and F-score.

Table 6 shows the confusion matrix taking a 50%-50% training and testing dataset distribution configuration. It can be observed that the vacuum cleaner's electrical signature may be confused with many other appliances, suggesting that the descriptors are not relevant enough for a correct discrimination.

We propose in the next section to apply one of the dimensionality reduction methods to reduce the number of harmonics by selecting the most relevant ones for the identification system.

#### 4.6. Feature selection results

This part is devoted to the selection of the most relevant features explaining the classes or types of electrical appliances among the 500 statistical harmonics features, with 8 statistical vectors per second using voting rules method. The features are the 250  $\mu(p)$  mean values followed by the 250  $\sigma(p)$  standard deviation values. The critical point in the histogram-based MI estimation procedure is the discretization phase, or more precisely the choice of bins number. In the literature, several formulas exist for this choice, like the formula proposed by Sturges, Scott, Freedman-Diaconis, or Shimazaki. In our simulation, we used the Sturges formula for calculating the number of bins in the discretization phase.

**Table 5.** Sensitivity  $S$ , precision  $P$ , and F-score  $F_s$  of each electrical appliance type and for different dataset distributions.

Training % / Testing % (NBRTrai / NBRTTest)		50/50 (537/537)			60/40 (645/429)			80/20 (859/215)		
N	Appliance type	$S$	$P$	$F_s$	$S$	$P$	$F_s$	$S$	$P$	$F_s$
1	Compact fluorescent lamp	95.45	97.67	96.55	100	97.22	98.59	100	100	100
2	Vacuum cleaner	50.00	63.63	56.00	60	81.81	69.23	62.50	83.33	71.42
3	Hair-dryer	97.56	96.38	96.96	100	93.93	96.87	100.00	93.93	96.87
4	Microwave	100	100	100	98.21	100	99.09	100	100	100
5	Air conditioner	87.87	100	93.54	96.15	100	98.03	92.30	100	96.00
6	Laptop	97.67	95.45	96.55	97.10	100	98.52	100	100	100
7	Fridge	100	100	100	100	100	100	100	100	100
8	Incandescent light bulb	94.73	90.00	92.30	100	97.87	98.92	100	95.83	97.87
9	Fan	94.82	94.82	94.82	100	97.87	98.92	100	95.83	97.87
10	Washing Machine	92.30	100	96.00	100	100	100	100	100	100
11	Heater	94.11	76.19	84.21	85.71	85.71	85.71	71.42	83.33	76.92
	Overall	91.32	92.19	91.54	94.28	95.85	94.90	93.29	95.66	94.27

**Table 6.** Confusion matrix obtained with 50% training dataset and 50% testing dataset distribution. For each active appliance type (number  $N$  of the top line) the matrix gives the number of correctly identified appliances and the incorrectly identified ones that are assigned to another appliance type number in the training dataset. The value in bold is the global CR. Dots stand for zero values.

N	1	2	3	4	5	6	7	8	9	10	11	$P$ (%)
1	84	.	.	.	.	2	.	.	.	.	.	97.67
2	.	7	.	.	.	.	.	3	.	.	1	63.64
3	.	2	80	.	1	.	.	.	.	.	.	96.39
4	.	.	.	70	.	.	.	.	.	.	.	100
5	.	.	.	.	29	.	.	.	.	.	.	100
6	4	.	.	.	.	84	.	.	.	.	.	95.45
7	.	.	.	.	.	.	19	.	.	.	.	100
8	.	2	.	.	1	.	.	54	3	.	.	90.00
9	.	1	.	.	1	.	.	.	55	1	.	94.83
10	.	.	.	.	.	.	.	.	.	12	.	100
11	.	2	2	.	1	.	.	.	.	.	16	76.19
$S$ (%)	95.45	50.00	97.56	100	87.88	97.67	100	94.74	94.83	92.31	94.12	<b>94.97</b>

Figure 3 clearly illustrates the effectiveness of feature selection and its contribution to the reduction of dimensionality. It can be seen that 10 relevant features are sufficient and even better by giving a higher CR rate than the CR obtained with the total number of 500 features.

The CR results for the 10 first selected features are reported in Table 7. These simulation results indicate that only 5 features are sufficient in order to exactly reach the same CR compared to 250 harmonics, whatever the dataset distribution (with a CR of 94.97% reached in the 50%/50% case and a higher CR of 97.20% in the 60%/40% and 80%/20% cases, as previously discussed). This shows that a high dimensionality reduction may be achieved without any performance loss. The curse of dimensionality is also observed in Table 7 and Figure 3 since a higher CR superior to the final value of CR can be reached, whatever the training dataset size. From Figure 3, the curve corresponding to training dataset size of 50% presents the peaking phenomenon, which explains the curse of dimensionality caused probably by the limited number of dataset samples. This phenomenon does not exist in other curves, justified probably by increasing the sample number. Furthermore, the results demonstrate that the mean features of the first odd orders (harmonic) are predominantly the relevant features for this task. Moreover, results show that the dataset size has no effect on the first eight selected features set and has a slight effect on their order, as well as a weak effect on the corresponding accuracies. For training datasets with size larger than 60%, the selected features roughly follow the same order. For the 50% case, high harmonic orders are selected from the ninth feature, probably due to the weak training dataset size, which first gives less information about the classes and second causes a poor estimation of MI [22].

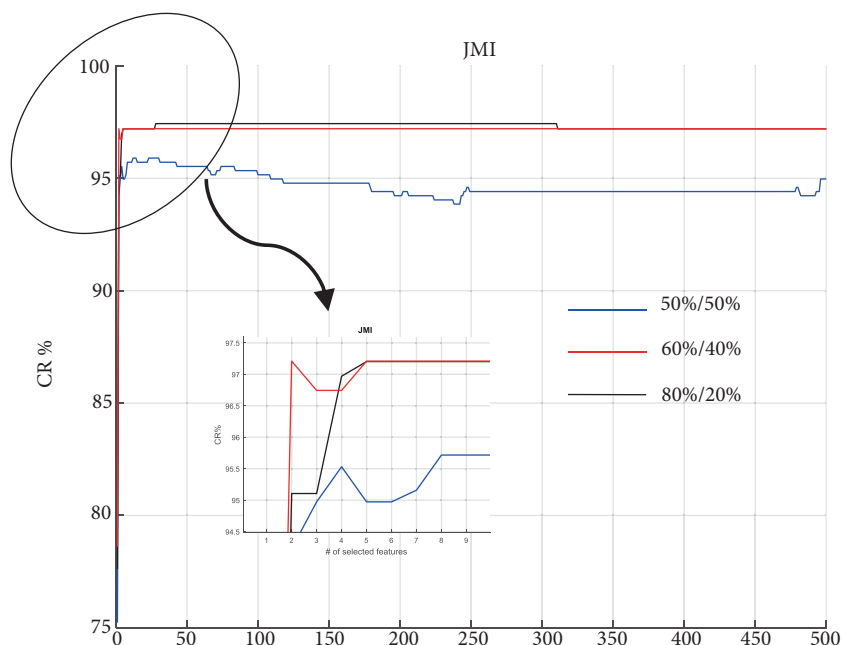


Figure 3. CR (%) for the JMI strategy of feature selection with the three dataset distribution configurations.

From these results, we can conclude that using the feature selection method has effectively and advantageously contributed to reducing the number of features.

### 5. Conclusions

In this paper, a novel framework for electrical appliance identification is proposed. Our first key ideas are the use of statistical features of harmonics and the application of the k-NN classifier combined with the voting rule method. Feature extraction from current signals has been performed on STFS coefficient statistics (mean, standard deviation, skewness, and kurtosis). The first results evaluated on the Plaid dataset demonstrate that

**Table 7.** The classification rate CR (%) of the 10 first selected features and for 500 features (NBR is the number of selected features) as a function of the training/testing dataset distribution.

NBR	CR (%) at 50%/50%	CR (%) at 60%/40%	CR (%) at 80%/20%
1	75.23	77.62	78.60
2	94.22	95.10	97.20
3	94.97	95.10	96.74
4	95.53	96.96	96.74
5	94.97	97.20	97.20
6	94.97	97.20	97.20
7	95.15	97.20	97.20
8	95.71	97.20	97.20
9	95.71	97.20	97.20
10	95.71	97.20	97.20
500	94.71	97.20	97.20

the combination of the mean and standard deviation features provided the optimal performance results in terms of classification rate. These results have been improved with 2.6% gain by the application of the voting rule with the choice of an optimal number of voting vectors.

In order to reduce the high dimensionality, we have also applied feature selection algorithms based on the mutual information JMI strategy. The application of the JMI strategy showed that the feature selection procedure was effective for reducing the number of features while outperforming the CR values.

Our results also showed the robustness of the proposed system with respect to different dataset training and testing distributions in terms of several performance evaluation metrics.

One possible issue is the relevance of this novel framework to the case of larger datasets. Moreover, other statistics such as entropy measures may be used for statistical feature estimation. The statistics may also be applied on other features like wavelet coefficients.

## References

- [1] Nait-Meziane M, Hacine-Gharbi A, Ravier P, Lamarque G, Le Bunetel JC et al. HMM-based transient and steady-state current signals modeling for electrical appliances identification. In: 5th International Conference on Pattern Recognition Applications and Methods; Rome, Italy; 2016. pp. 670-677.
- [2] Hacine-Gharbi A, Ravier P. Wavelet cepstral coefficients for electrical appliances identification using hidden Markov models. In: 7th International Conference on Pattern Recognition Applications and Methods; Funchal, Portugal; 2018. pp. 541-549.
- [3] Nait-Meziane M, Hacine-Gharbi A, Ravier P, Lamarque G, Le Bunetel JC et al. Electrical appliances identification and clustering using novel turn-on transient features. In: 6th International Conference on Pattern Recognition Applications and Methods; Porto, Portugal; 2017. pp. 647-654.
- [4] Thiruvaran T, Phung T, Ambikairajah E. Automatic identification of electric loads using switching transient current signals. In: IEEE TENCON Spring Conference; Sydney, Australia; 2013. pp. 252-256.
- [5] Hsueh-Hsien C, Nguyen VL. Statistical feature extraction for fault locations in nonintrusive fault detection of low voltage distribution systems. *Energies* 2017; 10: 611.

- [6] Esmael B, Arnaout A, Fruhwirth RK, Thonhauser G. A statistical feature-based approach for operations recognition in drilling time series. *International Journal of Computer Information Systems and Industrial Management Applications* 2013; 5: 454-461.
- [7] Hacine-Gharbi A, Ravier P, Nemo F. Local and global feature selection for prosodic classification of the word's uses. In: *6th International Conference on Pattern Recognition Applications and Methods*; Porto, Portugal; 2017. pp. 711-717.
- [8] Nilanon T, Yao J, Hao J, Purushotham S, Liu Y. Normal/abnormal heart sound recordings classification using convolutional neural network. In: *Computing in Cardiology Conference*; Vancouver, Canada; 2016. pp. 585-588.
- [9] Grofman B, Owen G, Feld SL. Thirteen theorems in search of truth. *Theory and Decision* 1983; 15: 261-278.
- [10] Xu L, Krzyzak A, Suen CY. Method of combining multiple classifiers and their application to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics* 1992; 22 (3): 418-435.
- [11] Ho TK, Hull JJ, Srihari SN. Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1994; 16 (1): 66-75.
- [12] Lam L, Suen CY. Application of majority voting to pattern recognition: an analysis of the behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics* 1997; 27 (5): 553-567.
- [13] Zou J, Nagy G. A comparative study of local matching approach for face recognition. *IEEE Transactions on Image Processing* 2007; 16 (10): 2617-2628.
- [14] Ogawara K, Fukutomi M, Uchida S, Feng Y. A voting-based sequential pattern recognition method. *PLoS ONE* 2013; 8 (10): e76980. doi: 10.1371/ journal.pone.0076980
- [15] Ridi A, Hennebert J. Hidden Markov models for ilm appliance identification. *Procedia Computer Science* 2014; 32: 1010-1015.
- [16] Mittelsdorf M, Hüwel A, Klingenberg T, Sonnenschein M. Submeter based training of multi-class support vector machines for appliance recognition in home electricity consumption data. In: *Proceedings of SmartGreens*; Aachen, Germany; 2013. pp. 151-158.
- [17] De Paiva Penha D, Castro ARG. Convolutional neural network applied to the identification of residential equipment in nonintrusive load monitoring systems. In: *3rd International Conference on Artificial Intelligence and Applications*; Chennai, India; 2017. pp. 11-21.
- [18] Dat Nguyen T, Dong Do T, Ha Le M, Le NT, Benjapolakul W. Appliance classification method based on k-nearest neighbors for home energy management system. In: *First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics*; Bangkok, Thailand; 2019. pp. 53-56.
- [19] Xue B. Particle swarm optimisation for feature selection in classification. PhD, Victoria University, Wellington, New Zealand, 2014.
- [20] Fleuret F. Fast binary feature selection with conditional mutual information. *Journal of Machine Learning Research* 2004; 5: 1531-1555.
- [21] Duch W. Feature extraction: foundations and applications. In: Guyon I, Gunn S, Nikravesh M, Zadeh L (editors). *Studies in Fuzziness & Soft Computing*. Berlin, Germany: Springer, 2006, pp. 89-117.
- [22] Hacine-Gharbi A, Deriche M, Ravier P, Harba R, Mohamadi T. A new histogram-based estimation technique of entropy and mutual information using mean squared error minimization. *Computers and Electrical Engineering* 2013; 39 (3): 918-933.
- [23] Brown G, Pocock A, Lujan MJ, Zhao M. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *Journal of Machine Learning Research* 2012; 13 (1): 27-66.
- [24] Liu H, Motoda H. *Feature Selection for Knowledge Discovery and Data Mining*. Dordrecht, the Netherlands: Kluwer Academic Publishers, 1998.



- [25] Hart GW. Residential energy monitoring and computerized surveillance via utility power flows. *IEEE Technology and Society Magazine* 1989; 8 (2): 12-16.
- [26] Hart GW. Nonintrusive appliance load monitoring. *Proceedings of the IEEE* 1992; 80 (12): 1870–1891.
- [27] Sultanem F. Procédé et appareil d'analyse de signaux de courant et de tension en vue du repérage de charges à usage domestique. Brevet Français FR 2645968, 1990 (in French).
- [28] Bons M. Modèles à sources markovienne cachée appliqués à l'analyse non-intrusive des principaux usages domestiques de l'électricité. PhD, Université de Rennes I, Rennes, France, 1996 (in French).
- [29] Onada T, Nakano Y, Yoshimoto K. System and method for estimating power consumption of electric apparatus, and abnormality alarm system utilizing the same. US patent 6,816,078 B2, 2004.
- [30] Leeb S. A conjoint pattern recognition approach to nonintrusive load monitoring. PhD, Massachusetts Institute of Technology, Cambridge, MA, USA, 1993.
- [31] Cole AI, Albicki A. Data extraction for effective nonintrusive identification of residential power loads. In: *IEEE Instrumentation and Measurement Technology Conference*; St. Paul, MN, USA; 1998. pp. 812-815.
- [32] Drenker S, Kader A. Nonintrusive monitoring of electric loads. *IEEE Computer Applications in Power* 1999; 12 (4): 47-51.
- [33] Chan WL, So ATP, Lai LL. Harmonics load signature recognition by wavelets transforms. In: *International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*; London, UK; 2000. pp. 666-671.
- [34] Baranski M, Voss J. Nonintrusive appliance load monitoring based on an optical sensor. In: *IEEE Bologna Power Tech Conference*; Bologna, Italy; 2003. p. 8.
- [35] Patel SN, Robertson T, Kientz JA, Reynolds MS, Abowd GD. At the flick of a switch: detecting and classifying unique electrical events on the residential power line. In: *International Conference on Ubiquitous Computing*; Innsbruck, Austria; 2007. pp. 271–288.
- [36] Chang HH, Lin CL, Yang HT. Load recognition for different loads with the same real power and reactive power in a nonintrusive load monitoring system. In: *12th IEEE International Conference on Computer Supported Cooperative Work in Design*; Xi'an, China; 2008. pp. 1122–1127.
- [37] Du Y, Du L, Lu B, Harley R, Habetler T. A review of identification and monitoring methods for electric loads in commercial and residential buildings. In: *IEEE Energy Conversion Congress and Exposition*; Atlanta, GA, USA; 2010. pp. 4527–4533.
- [38] Zeifman M, Roth K. Nonintrusive appliance load monitoring: review and outlook. *IEEE Transactions on Consumer Electronics* 2011; 57 (1): 76-84.
- [39] Carrie Armel K, Gupta A, Shrimali G, Albert A. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy* 2013; 52: 213-234.
- [40] Najmeddine H, El Khamlichi Drissi K, Pasquier C, Faure C, Kerroum K et al. States of art on load monitoring methods. In: *2nd IEEE International Conference on Power and Energy*; Johor Baharu, Malaysia; 2008. pp. 1256-1258.
- [41] De Paiva Penha D, Castro ARG. Home appliance identification for NILM systems based on deep neural networks. *International Journal of Artificial Intelligence and Applications* 2018; 9 (2): 69-80.
- [42] Liu Q, Wu H, Liu X, Linge N. Single appliance recognition using statistical features based k-nn classification. In: *International Conference on Cloud Computing and Security*; Nanjing, China; 2017. pp. 631-640.
- [43] Mallat S. *A Wavelet Tour of Signal Processing*. 2nd ed. San Diego, CA, USA: Academic Press, 1999.
- [44] van Erp M, Vuurpijl L, Schomaker L. An overview and comparison of voting methods for pattern recognition. In: *Proceedings of Eighth International Workshop on Frontiers in Handwriting Recognition*; Niagara on the Lake, Canada; 2002. pp. 195-200.

- [45] Bilski P, Winiecki W. Non-intrusive appliance load identification with the ensemble of classifiers. In: Proceedings of the International Workshop on Non-Intrusive Load Monitoring; Vancouver, Canada; 2016. pp. 1-5.
- [46] Pascoal C, Oliveira MR, Pacheco A, Valadas R. Theoretical evaluation of feature selection methods based on mutual information. *Neurocomputing* 2016; 226: 168-181.
- [47] Cover T, Thomas J. *Elements of Information Theory*. 2nd ed. New York, NY, USA: Wiley and Sons, 2006.
- [48] Yang H, Moody J. Data visualization and feature selection: new algorithms for non-Gaussian data. *Advances in Neural Information Processing Systems* 1999; 12: 687-693.
- [49] Gao J, Giri S, Kara EC, Berges M. Plaid: A public dataset of high-resolution electrical appliance measurements for load identification research: demo abstract. In: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings; New York, NY, USA; 2014. pp. 198–199.
- [50] Kelly J, Knottenbelt W. Neural NILM: Deep neural networks applied to energy disaggregation. In: 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments; Seoul, South Korea; 2015. pp. 55-64.
- [51] Batra N, Parson O, Berges M, Singh A, Rogers A. A comparison of nonintrusive load monitoring methods for commercial and residential buildings 2014. arXiv preprint, arXiv: 1408.6595.
- [52] Djordjević S, Simić M. Nonintrusive identification of residential appliances using harmonic analysis. *Turkish Journal of Electrical Engineering & Computer Sciences* 2018; 26 (2): 780-791.