



Comparative Evaluation of Non-Intrusive Load Monitoring Methods Using Relevant Features and Transfer Learning

Sarra Houdi, Dominique Fourer, François Auger, Houda Ben Attia Sethom,
Laurence Miègeville

► To cite this version:

Sarra Houdi, Dominique Fourer, François Auger, Houda Ben Attia Sethom, Laurence Miègeville. Comparative Evaluation of Non-Intrusive Load Monitoring Methods Using Relevant Features and Transfer Learning. *Energies*, 2021, 14 (9), pp.2726. 10.3390/en14092726 . hal-03223322

HAL Id: hal-03223322

<https://hal.science/hal-03223322>

Submitted on 10 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article

Comparative Evaluation of Non-Intrusive Load Monitoring Methods Using Relevant Features and Transfer Learning

Sarrah Houdi ¹, Dominique Fourer ^{1,*}, François Auger ², Houda Ben Attia Sethom ³ and Laurence Miègeville ²

¹ Laboratoire IBISC (Informatique, BioInformatique, Systèmes Complexes), EA 4526, University Evry/Paris-Saclay, 91020 Evry CEEEX, France; sarrah.houdi@univ-evry.fr

² Institut de Recherche en Energie Electrique de Nantes Atlantique (IREENA), EA 4642, University of Nantes, 44602 Saint-Nazaire, France; francois.auger@univ-nantes.fr (F.A.); laurence.miegeville@univ-nantes.fr (L.M.)

³ Laboratoire des Systèmes Electriques, Université de Tunis El Manar, Tunis 1002, Tunisia; houda.benattia@enicarthe.rnu.tn

* Correspondence: dominique@fourer.fr

Abstract: Non-Intrusive Load Monitoring (NILM) refers to the analysis of the aggregated current and voltage measurements of Home Electrical Appliances (HEAs) recorded by the house electrical panel. Such methods aim to identify each HEA for a better control of the energy consumption and for future smart grid applications. Here, we are interested in an event-based NILM pipeline, and particularly in the HEAs' recognition step. This paper focuses on the selection of relevant and understandable features for efficiently discriminating distinct HEAs. Our contributions are manifold. First, we introduce a new publicly available annotated dataset of individual HEAs described by a large set of electrical features computed from current and voltage measurements in steady-state conditions. Second, we investigate through a comparative evaluation a large number of new methods resulting from the combination of different feature selection techniques with several classification algorithms. To this end, we also investigate an original feature selection method based on a deep neural network architecture. Then, through a machine learning framework, we study the benefits of these methods for improving Home Electrical Appliance (HEA) identification in a supervised classification scenario. Finally, we introduce new transfer learning results, which confirm the relevance and the robustness of the selected features learned from our proposed dataset when they are transferred to a larger dataset. As a result, the best investigated methods outperform the previous state-of-the-art results and reach a maximum recognition accuracy above 99% on the PLAID evaluation dataset.

Keywords: Non-Intrusive Load Monitoring (NILM); Home Electrical Appliances (HEAs); identification; feature selection; transfer learning; deep learning



Citation: Houdi, S.; Fourer, D.; Auger, F.; Ben Attia Sethom, H.; Miègeville, L. Comparative Evaluation of Non-Intrusive Load Monitoring Methods Using Relevant Features and Transfer Learning.

Energies **2021**, *14*, 2726.

<https://doi.org/10.3390/en14092726>

Academic Editors: Junemo Koo and Jinkyun Cho

Received: 16 March 2021

Accepted: 23 April 2021

Published: 10 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

During the last decades, the electricity consumption in the residential sector increased steadily with the worldwide population growth and became a major ecological issue. Prior studies show that a real-time feedback down to the HEA level can help to effectively reduce consumption, with almost 15% of energy savings [1]. For consumers, the main advantages are the control and the understanding of their electricity consumption through a transparent access and a promptly forwarded information. For utilities, it can improve the load-forecasting accuracy and provide a basic scheme to set up energy management strategies [2]. In this context, Non Intrusive Load Monitoring (NILM) methods offer an efficient answer, since they provide a breakdown of the residential energy consumption without instrumenting each HEA.

Here, we are interested in event-based NILM systems, where the current and voltage measurements are recorded using a single sensor connected to the house electrical panel [3–7]. An event detection method is used to predict the changes in the aggre-

gated power signals that occur at each HEA's operating state [6]. Then, relevant features that meet the additive criterion [8] (which is required for the subsequent steps) are computed to recognize the HEA electrical signature that triggered an event using a pattern matching method.

NILM encounters several challenges that concern the correct identification of HEAs. Indeed, in a house, a broad range of HEAs can be present, and many of them can rely on the same electrical behavior [9–11]. Hence, discerning the most relevant and informative features is of paramount importance for any NILM application whose performances depend on the HEA signature uniqueness [4,9,12]. This is the reason why an HEA must be described by a reduced number of relevant features. To date, prior studies have focused on HEA recognition performance and often involve non-physics-related and difficult to interpret features. This is the case for the features provided by deep convolutional neural networks (CNN), which can suffer from the robustness issue with adversarial examples [13,14], which require the use of attention mechanisms [15]. However, only a few works investigate in detail the role and the meaning of Feature Selection (FS) methods in NILM problems when addressed through a pattern recognition approach [16–20].

Thus, this study aims at filling the gap by investigating different FS techniques to tackle the HEA identification problem in a supervised machine learning scenario when a large number of electrical features are available and when HEAs from distinct manufacturers belonging to the same device category are considered. The goal is to show the benefits provided by FS in terms of HEAs' classification performance and interpretability, and for enhancing the generalization capability by reducing overfitting of the trained models. Thus, transfer learning [21], which deals with the generalization capability of the selected features for different datasets, is also an important NILM challenge that is investigated in this study.

This paper is organized as follows. In Section 2, the addressed problem is formulated. In Section 3, we introduce the two HEAs' datasets considered in this study. In particular, we present a novel dataset of HEAs' current and voltage measurements in steady-state conditions, and we detail the electrical features computed from these measurements. In Section 4, several FS methods are presented. Two of them that are novel are detailed: a heuristic forward method and a method based on a trained Dense neural network. Finally, in Section 5, the selected features are used in combination with several classification algorithms on both considered datasets to demonstrate the importance of selecting the suitable combination of features and classification algorithm. The paper is concluded with future work directions in Section 6.

2. Problem Statement

2.1. Supervised HEA Identification

From the values of a set of features describing a unique HEA signature, we aim at identifying the closest pre-registered HEA, referred to (according to the machine learning literature) as a class. To this end, we use a supervised machine learning framework trained on an annotated dataset. The overall HEAs' recognition problem can be illustrated in the flowchart of Figure 1, which also includes references to the paper sections where each method is detailed. During the training step, several features (or descriptors) deduced from voltage and current HEAs' measurements are first computed and used to derive a discriminative model for each HEA. Then, the most discriminant features are automatically selected to obtain the optimal classification accuracy. During the test and operating procedures, the selected features are computed from the observed measurements to predict the class of the corresponding HEA.

Since this work is motivated by HEA recognition for energy consumption estimation and prediction, the same appliance type can thus be considered from a different class when the HEA is recorded at different power levels, or when its energy consumption sufficiently differs due to a different manufacturing brand. Hence, we use a larger and more complicated classification taxonomy than those commonly used for HEA identification

in the literature. From another hand, we only consider loads in steady-state condition because switching on (or off) introduces transient signals or fluctuations which are often not sufficient to accurately characterize an HEA. In fact, transients are also affected by HEA-independent factors such as the time on the voltage waveform where the HEA is switched on/off, the network impedance, supply voltage distortion, the sampling frequency or the switching on/off mechanism [22]. To deal with transient signals and state change detection for multiple HEA recognition, the reader can for example refer to a multivariate statistical approach recently proposed in [6,23].

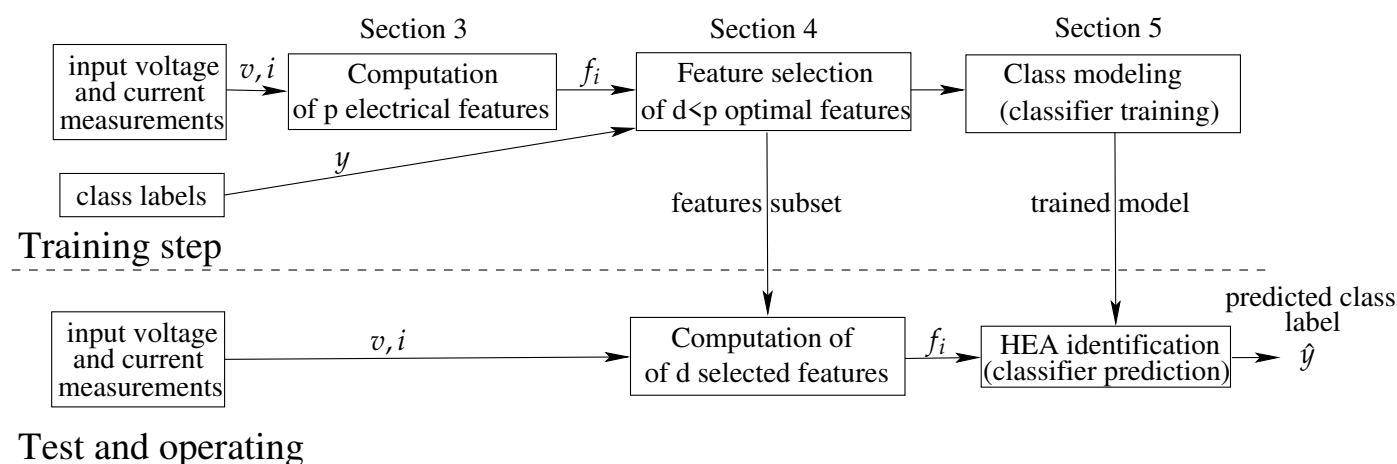


Figure 1. Flowchart of the general process for supervised HEA identification.

2.2. Features Selection for HEA Identification

Let $F = \{f_1, f_2, \dots, f_p\}$, with $f_i \in \mathbb{R}, \forall i \in [1, p]$ be the overall set of p features. The FS process aims at finding the optimal subset of features $F' \subseteq F$ that maximizes the identification accuracy such that its cardinality verifies $\text{card}(F') = d \leq p$ [24]. This process induces several benefits such as avoiding the curse of dimensionality and overfitting phenomenon [25], improving the classification performance with the removal of non-discriminating features and decreasing the computational cost by only collecting the needed features. In contrast to the other dimensionality reduction techniques [24], FS does not change the original meaning of features and therefore allows further interpretation by a domain expert. Our work consists in evaluating several feature selection techniques when applied to an HEA recognition scenario, in terms of recognition accuracy and in terms of robustness by investigating two distinct datasets and several classification methods.

2.3. Transfer Learning

The last challenge addressed in this study concerns transfer learning [21], which consists in using the knowledge extracted from a given dataset to tackle a new problem based on a different dataset with a different setup and classification taxonomy. The investigated datasets can be of different natures since they use different recording protocols, different grid properties and different annotation taxonomies. This main motivation is to show the validity and the generalization capability of the selected features from one dataset to another. To this end, we introduce a novel dataset recorded on a French grid (utility frequency at 50 Hz) which is different from the other existing publicly available datasets such as PLAID [26], which was recorded in the USA grid with utility frequency at 60 Hz. Hence, our new proposed dataset contains a large variety of HEA types including new ones such as LCD and Plasma TV, Coffee maker, and oven.

3. Materials

3.1. HEAs Datasets

To test the performance of an NILM technique, it is important to consider real data. Due to the challenges in the creation of such datasets, mainly related to the required time and the high costs involved, we investigated in this study two data sets that are freely available: the PLAID dataset [26], which is shared online and is used as common reference, and our novel publicly available dataset freely accessible at <http://dx.doi.org/10.21227/ww76-d733>). Both datasets contain individual HEAs measurements which are convenient for extracting features, training models, conducting performance evaluation and performing benchmarking on a common basis. Indeed, some existing datasets include scenarios of multiple simultaneous loads [27]. However, before conducting disaggregation (i.e., decomposing the whole energy consumption of a dwelling into the energy usage of individual HEAs), it is important to build an initial signature database that is key to many NILM techniques.

For both datasets considered in this study, a class of HEA corresponds to a brand of a category of HEA, e.g., the class “Incandescent light bulb-Electrix-soft white”, which is distinct from the HEA class “Incandescent light bulb-Philips Durama”. Furthermore, we only consider steady-state conditions, and we extracted for each class of HEAs in both datasets several periods of the current and voltage steady-state waveforms. Indeed, during HEAs’ switching On/Off, momentary fluctuations of the current and voltage signals occur before settling in to a steady-state value. These fluctuations are called transients and can characterize a given HEA. However, one major drawback of the switching transients is their reproducibility, since they can be affected by HEA-independent factors.

3.1.1. PLAID Dataset

This dataset [26] contains current and voltage measurements sampled at 30 kHz from 11 different HEA types present in more than 60 households in Pittsburgh, Pennsylvania, USA. The goal of this dataset is to provide a public library for high-frequency measurements that can be used to assess existing or novel HEA classification algorithms. There are 11 categories of HEAs, which correspond to: air conditioner, compact fluorescent lamp, fridge, hairdryer, laptop, microwave, washing machine, bulb, vacuum, fan, and heater. Each category of HEA is represented by more than ten different instances. For each HEA, three to six measurements are collected for each state transition (i.e., on/off changing state). As mentioned, we only consider the steady-state operations, and we extracted for each class of HEAs in the PLAID dataset several periods of the current and voltage steady-state waveforms such that we have a total of 71 HEAs (with different categories and brands) and $n = 36,720$ distinct recordings (also called individuals in the statistical terminology).

3.1.2. New Proposed Dataset

We introduce a novel publicly available dataset containing 24 categories of HEAs (e.g., fans, fridges, washers, etc.) considered with distinct brands (35 types) and that were recorded at several power levels to lead us to a total of 61 considered HEAs. For HEAs with a wide variety of operational programs or adjustable settings such as temperature or intensity, we recorded all power consumption patterns and considered that an HEA power level corresponds to a specific class of HEA, although power consumption patterns refer to the same device. Figure 2 lists the considered categories of HEAs.

Number of HEAs	Category of HEA	Number of power levels	Number of HEAs	Category of HEA	Number of power levels	Number of HEAs	Category of HEA	Number of power levels
2	Flat iron	1	1	Plasma TV	1	1	Microwave	1
1	LCD TV	1	2	Hand Mixer	1	1	Washing machine	9
1	Fryer	1	1	Coffee Maker	1	1	Rice cooker	2
1	Waffle iron	1	2	Laptop	1	1	Vacuum 1	5
1	Iron	1	2	Fridge	1	1	Vacuum 2	2
1	Fan	3	1	Oven	2	1	Radiator	3
2	Hair dryer	4	2	LED lamp	1	3	Kettle	1
3	Incandescent lamp	1	1	Iron	1	2	Low energy lamp	1

Figure 2. Measured appliances included in our new proposed dataset.

The HEAs have been recorded in steady-state conditions in a French 50 Hz electrical grid. The measurement setup consists of an AC current probe (E3N Chauvin Arnoux) with a 10 mV/A sensitivity and a differential voltage probe with a 1/100 attenuation. Voltage and current waveforms were captured by an 8-bit resolution digital oscilloscope (RIGOL DS1104Z). The sampling rate was set at $f_{s1} = 250$ kHz for some of the recordings and $f_{s2} = 50$ kHz for the other part. Electrical assembly and instrumentation are depicted in Figure 3. Hence, each HEA of this dataset is described by 8 periods of the current and voltage steady-state waveforms, resulting consequently in a set of $n = 8 \times 61 = 488$ distinct individuals.

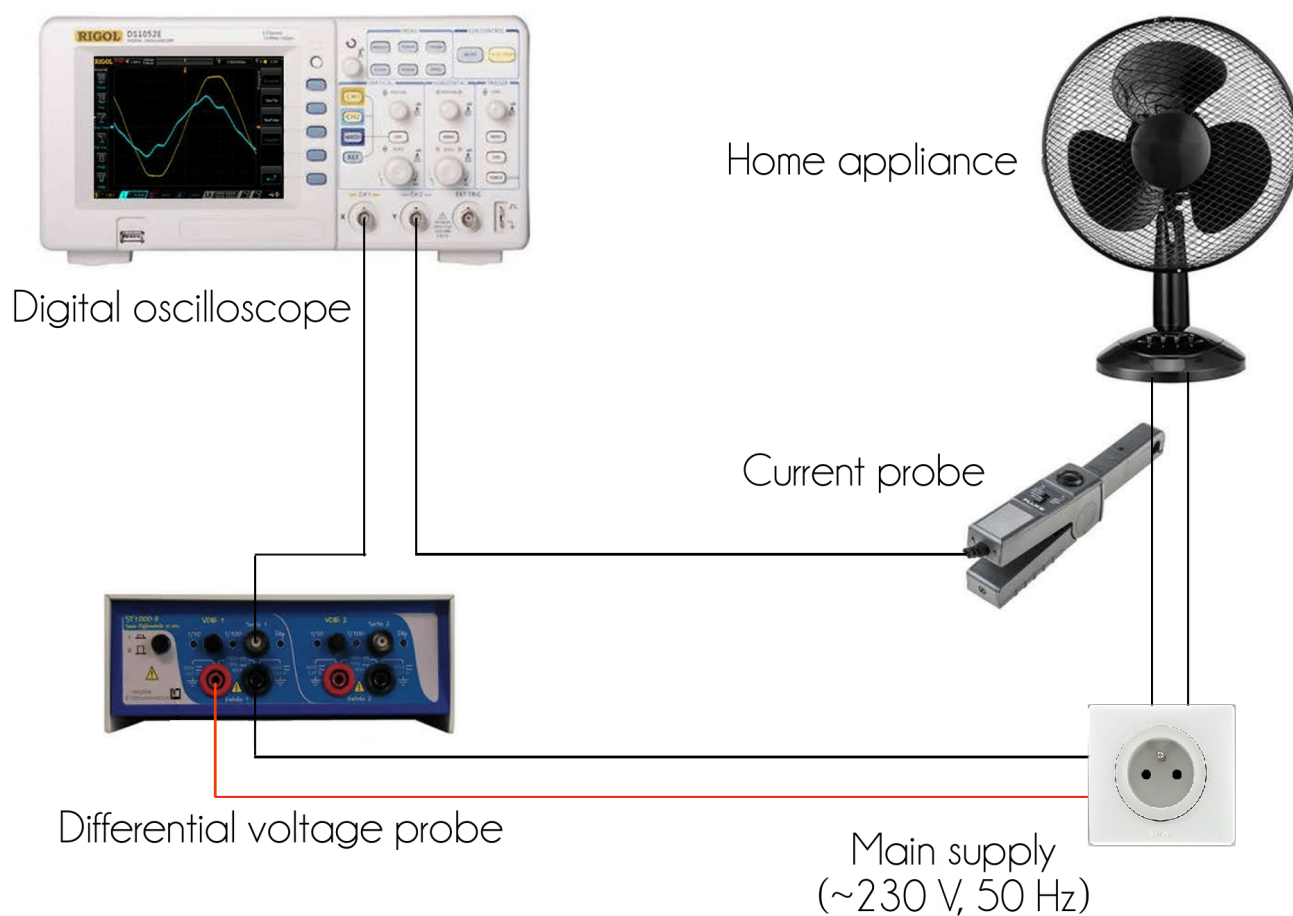


Figure 3. Electrical assembly and instrumentation.

3.2. Electrical Features Computed from Current and Voltage Measurements

We can describe the different HEAs using 90 features extracted at each voltage period summarized in Table 1. The detail of their computation was introduced in [19], based on the latest IEEE 1459-2010 standard for the definition of single phase physical components under non-sinusoidal conditions [28,29]. From the voltage $v(t)$ and current $i(t)$ signals, we compute the Fourier coefficients v_{ak} , v_{bk} , i_{ak} and i_{bk} , using the following formulas:

$$x_{ak} = \frac{2}{M} \sum_{m=0}^{M-1} x[m] \cos(2\pi mkf_0/f_s), \quad x_{bk} = \frac{2}{M} \sum_{m=0}^{M-1} x[m] \sin(2\pi mkf_0/f_s) \quad (1)$$

where $x[m] = x(m/f_s)$ is the sampled discrete-time signal, f_s is the sampling frequency, f_0 the utility frequency (e.g., 50 Hz in France) and $M = f_s/f_0$. From these Fourier coefficients, the following features can be computed for any harmonic rank $k \in \mathbb{N}$.

- The Root Mean Squared (RMS) value of the k^{th} harmonic component of the voltage V_k and currents I_k , and their sum V_H and I_H :

$$V_k = \sqrt{\frac{v_{ak}^2 + v_{bk}^2}{2}}, \quad I_k = \sqrt{\frac{i_{ak}^2 + i_{bk}^2}{2}}, \quad V_H = \sqrt{\sum_{k=2}^{15} V_k^2}, \quad I_H = \sqrt{\sum_{k=2}^{15} I_k^2} \quad (2)$$

- The RMS voltage V and current I :

$$V = \sqrt{V_1^2 + V_H^2}, \quad I = \sqrt{I_1^2 + I_H^2} \quad (3)$$

- The k^{th} harmonic component of the active, reactive, and apparent powers P_k , Q_k , S_k , and their sums P_H , Q_H , S_H :

$$P_k = \frac{1}{2}(v_{ak}i_{ak} + v_{bk}i_{bk}), \quad P_H = \sum_{k=2}^{15} P_k \quad (4)$$

$$Q_k = \frac{1}{2}(v_{ak}i_{bk} - v_{bk}i_{ak}), \quad Q_H = \sum_{k=2}^{15} Q_k \quad (5)$$

$$S_k = V_k \times I_k, \quad S_H = V_H \times I_H \quad (6)$$

- The active, reactive apparent, and distortion powers P , Q , S , and D :

$$P = P_1 + P_H, \quad Q = Q_1 + Q_H, \quad S = V \times I, \quad D = \sqrt{S^2 - P^2 - Q^2} \quad (7)$$

- The voltage and current total harmonic distortion THD_V and THD_I :

$$\text{THD}_V = \frac{V_H}{V_1}, \quad \text{THD}_I = \frac{I_H}{I_1} \quad (8)$$

- The voltage and current distortion powers D_V and D_I :

$$D_V = V_H I_1, \quad D_I = V_1 I_H \quad (9)$$

- The non-fundamental apparent power S_N :

$$S_N = \sqrt{D_I^2 + D_V^2 + S_H^2} \quad (10)$$

- The voltage and current crest factors for $m \in [0, M-1]$:

$$F_{CV} = \frac{\max |v[m]|}{V}, \quad F_{CI} = \frac{\max |i[m]|}{I} \quad (11)$$

- Finally, the global and harmonic power factors F_p and F_{pk} :

$$F_p = \frac{P}{S}, \quad F_{pk} = \frac{P_k}{S_k} \quad (12)$$

Table 1. Summary of the proposed electrical features [19]. The features colored in red meet the additivity criterion given by Equation (13).

Electrical Features Name	Number of Computed Features
Effective current I and its harmonics I_k for $k \in \{1, \dots, 15\}$ and I_H (A)	17
Active power P and its harmonics P_k for $k \in \{1, \dots, 15\}$ and P_H (W)	17
Reactive power Q and its harmonics Q_k for $k \in \{1, \dots, 15\}$ and Q_H (VAR)	17
Apparent power S and its harmonics S_k for $k \in \{1, \dots, 15\}$ and S_H, S_N (VA)	18
Current harmonic distortion THD_I	1
Distortion power D, D_I, D_V (VAD)	3
Power factor F_p and F_{pk} for $k \in \{1, \dots, 15\}$	16
Current crest factor F_{CI}	1
Total	90

Since the NILM problem consists in the breakdown of an unknown mixture of HEAs into a set of identifiable HEAs signatures possibly belonging to a database, it is important to consider the electrical features that meet the additivity criterion [8] such that:

$$f(\mathbf{v}, \mathbf{i}) = \sum_{n=1}^{N_s} f(\mathbf{v}, \mathbf{i}_n) \quad \text{for} \quad \mathbf{i} = \sum_{n=1}^{N_s} \mathbf{i}_n \quad (13)$$

where $f(\mathbf{v}, \mathbf{i})$ is an electrical feature, $\mathbf{v}$ is a vector of voltage, $\mathbf{i}$ a vector of current samples acquired during one voltage period and N_s is the number of HEAs that are simultaneously switched on. Thus, when an HEA is connected (resp. disconnected) to the power network, an “additive” feature is increased (resp. decreased) by an amount equal to that produced by this HEA operating individually. Among the 90 features, $p = 34$ features meet the additivity criterion and are reported in red in Table 1. This property is required for extracting HEAs features from an aggregated signal and for comparing them with the dataset of individual HEAs signatures. For example, a change detection method as proposed in [6] can be used in order to separate the distinct contribution of each HEA when several ones are switched on simultaneously. Hence, this study only focuses on the additive features computed for a unique observed HEA. Both datasets are stored in an $n \times p$ normalized matrix $\tilde{X}$, where the $p = 34$ features detailed earlier are in columns and have been normalized with a zero mean and a unit standard deviation.

4. Feature Selection

Existing approaches for feature selection can be categorized into filter-based methods and wrapper-based methods [30,31]. Filter-based methods perform FS independently of the classification process. Features are individually assigned to relevance scores, which are assumed to reflect their usefulness in the classification task. Hence, the features are sorted by descending order of the obtained score of relevance [32]. Filter-based methods are often computationally faster but are known to be less accurate than other approaches. On the other hand, wrapper-based methods use the classifier of interest to score feature subsets according to the classification accuracy. This allows selecting an optimal subset of features that maximizes the classification accuracy with an improvement of the computational cost [32].

4.1. Investigated Feature Selection Methods

This paper investigates the methods listed below, which are comparatively assessed in the remainder. More details are given to the sequential forward FS method and our original contributions concerning the DNN methods, in light of the other methods' literature review.

4.1.1. Existing Methods

- Principal Component Analysis (PCA) can be used as a filter-based method based on the maximization of the dataset dispersion [33,34].
- Linear Discriminant Analysis (LDA) can be used as a supervised filter-based method maximizing the separation between classes (see Section 5.1).
- Mutual Information (MI) is a filter-based method measuring the amount of information each feature conveys from the class labels [35].

4.1.2. New Proposed Sequential Forward Method

This approach is an iterative wrapper-based FS method that aims at maximizing the classifier accuracy when adding each feature one-by-one [30,36–39]. The first iteration starts with an empty set of selected features, and each feature $f \in F$, where F is the set of p features, is tested individually using a given classifier (in this study, we used LDA and K-nearest-neighbor (KNN) presented in Section 5). The feature that maximizes the classification accuracy (Acc) [40] of the training dataset is the first one selected. Then, each remaining feature is added to the set of previously selected features, and the same process is applied to find again the highest accuracy. The iterative feature selection ends once further feature addition yields no accuracy improvement. The proposed method is presented hereafter in Algorithm 1.

Algorithm 1 Sequential forward FS algorithm

Input: dataset $X_{n,p}$, whole set of features F , ground-truth labels Y

Output: Set of sorted features F_p

```

1: Initialization:  $F_0 \leftarrow \emptyset$ 
2: for  $k \leftarrow 0$  to  $p - 1$  do
3:   for  $f \in F \setminus F_k$  do
4:      $F_+ \leftarrow F_k \cup \{f\}$ 
5:     evaluate (LDA or KNN classifier,  $X_n(F_+)$ ,  $Y$ )
6:     compute  $Acc(F_+)$ 
7:   end for
8:   Select the best remaining feature:
      $f_k \leftarrow \arg \max_{f \in F \setminus F_k} Acc(F_k \cup \{f\})$ 
9:   Update feature set:  $F_{k+1} \leftarrow F_k \cup \{f_k\}$ 
10: end for
11: return  $F_p$ 

```

4.1.3. New Proposed Deep Neural Network (DNN) Feature Selection Method

This method is inspired by [41], which is a filter-based approach that uses our proposed neural architecture. It uses the sum of the trained weights of the first layer neurons as a score of relevance of the input features.

Deep learning uses a combination of agents (the neurons) to learn high-level non-linear relationships and correlations in the analyzed data. Such methods can tackle complicated problems and become a promising approach for smart grid applications [42]. Here, we

use a dense fully connected DNN architecture [43], where each neuron of the input layer is associated with a feature f_i as presented in Figure 4. Our architecture is made of one input layer with p neurons and 2 hidden blocks containing one layer of 128 dense formal neurons including a Batch Normalization (BN) combined with a dropout layer. The output y of each neuron can be expressed as a function of an input vector $x \in \mathbb{R}^N$ such as:

$$y = g\left(w_0 + \sum_{i=1}^N w_i x_i\right) \quad (14)$$

where g is the neuron activation function chosen as REctified Linear Unit (RELU) defined by $g(x) = \max(0, x)$ except for the last output layer, which uses the softmax activation function [44]. The w_i coefficients (w_0 being the bias) are the synaptic weights, which are learned during the training process. In our implementation, we choose $w_0 = 0$ and we use a BN of the output, which was shown to improve the training performances [45] when applied on the activation of each neuron. Thus, we have $\text{BN}(y) = \frac{y - \mu}{\sigma}$, where μ and σ are the mean and the standard deviation computed from the processed training batch defined with a size equal to 64. Each dropout layer is defined to randomly discard 10 % of the connected input layer and is used in order to reduce overfitting [46]. The last layer of our proposed DNN model applies a softmax activation function on each output neuron (each output neuron is associated with a prediction class) returning output values in $[0, 1]$. This value corresponds to the probability of predicting the analyzed input as a member of class i . Hence, the final output label corresponds to the class index which maximizes the resulting probability of the last layer such as $\hat{y} = \arg \max_i y_i$, y_i being the output of neuron i of the last layer. The proposed DNN is designed for classification or regression problems; however, we propose here a new feature selection method based on the trained weights of each neuron. Following this idea, we consider the sum of the weights of all the neurons of the first layer linked to the same input feature as a score of relevance for this feature. The DNN is trained on the studied datasets using cross-entropy [43] as a loss function.

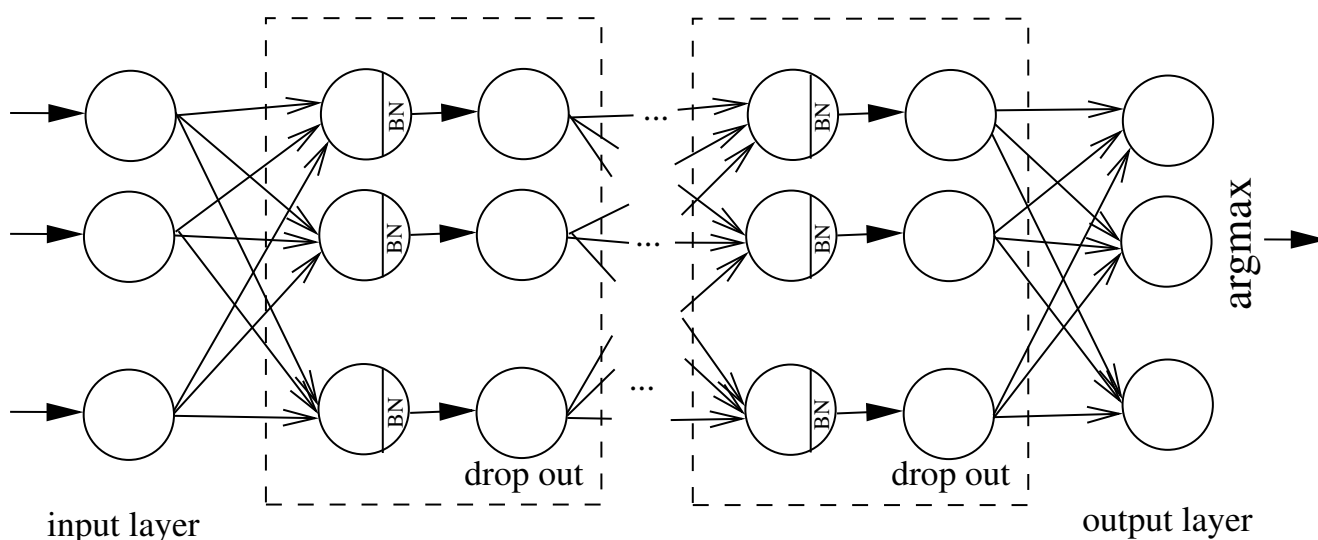


Figure 4. Diagram of the proposed deep neural network architecture with L hidden layers.

4.2. Feature Selection Results

In the context of event-based NILM [6], the selection of features meeting the additivity criterion matters. Each FS method, among the five ones previously presented in Section 4.1, is then applied on the two datasets considered in this study (see Section 3.1), where each HEA is represented by a vector of $p = 34$ features meeting the additivity criterion (see Section 3.2). Datasets are centered and reduced in order to get an $(n \times p)$ matrix $\tilde{X}$ with a zero mean and a unit standard deviation. This is obtained by applying the same operation

as used for the Batch Normalization described above (subtracting the mean and dividing by the standard deviation of each individual). For the FS methods based on a feature relevance score, we sort the features by descending order of relevance, and then we select the subset before the highest decrease in the relevancy score. The results obtained on our own dataset and the PLAID dataset are reported in Tables 2 and 3.

Table 2. Results of the features selection methods applied on the proposed dataset.

Method	Features Meeting the Additive Criterion	
	# Features	Selected Features
KNN based sequential forward FS method	12	$P_1, Q_1, P_7, Q_3, Q, P, P_3, P_H, P_5, Q_5, Q_H, Q_9$
LDA based sequential forward FS method	18	$P_1, Q, P, Q_9, Q_3, P_3, P_2, P_{10}, Q_4, P_4, P_6, P_9, Q_8, P_{13}, P_8, Q_{15}, Q_5, Q_{11}$
MI	20	$P_1, P, P_5, Q, Q_1, Q_H, P_H, P_7, P_{15}, P_9, Q_5, Q_7, P_{13}, Q_3, Q_9, P_{11}, Q_{13}, P_3, Q_{15}, Q_{11}$
PCA	17	$Q, Q_1, P_3, P_9, P_7, P_H, Q_5, P_{11}, Q_9, P_6, Q_3, Q_{11}, Q_7, P_{15}, P_{12}, Q_6, P_8$
LDA	14	$P_7, Q, Q_1, P_9, P_3, Q_5, P_H, Q_7, P_{11}, P_5, Q_9, P, P_1, Q_{11}$
DNN	27	$Q_9, P_3, P_H, P_{12}, Q_{15}, Q_5, Q_{12}, P_6, P_9, Q_8, P_7, Q_4, P_5, Q_H, P_{11}, Q_2, Q_3, Q_{11}, Q_{14}, P_{10}, P_1, P, P_2, P_4, P_8, P_{13}, P_{14}$

Table 3. Results of the features selection methods applied on PLAID dataset.

Method	Features Meeting the Additive Criterion	
	# Features	Selected Features
KNN based sequential forward FS method	25	$P_H, Q, P_1, P_3, Q_H, Q_5, P_5, Q_3, Q_1, Q_7, P_{15}, Q_2, P, Q_{13}, P_{12}, P_7, Q_{15}, Q_9, P_{13}, P_9, P_{10}, Q_{10}, Q_{11}, Q_{12}, Q_8$
LDA based sequential forward FS method	33	All features except Q_9
MI	20	$P_3, P_H, P_1, P, Q_1, Q, Q_9, Q_7, P_7, P_5, Q_5, Q_H, Q_3, P_9, Q_{11}, P_{11}, Q_{13}, Q_{15}, P_{15}, P_{13}$
PCA	31	All features except P_{14}, Q, Q_1
LDA	12	$P_3, P_H, P_{15}, Q_{13}, Q_5, Q_H, Q_{15}, P_9, P_7, Q_3, Q_9, Q_7$
DNN	17	$P_H, Q_{11}, P_3, P_9, Q_9, P_{15}, P_5, Q_7, P_7, P_{10}, Q_3, Q_H, Q_{12}, Q_{13}, Q_{15}, P_{13}, P_6$

The following observations can be made according to the obtained results:

- For both datasets, some features such as P , P_1 , P_H , Q , Q_1 , or Q_H are present regardless of the used FS method;
- For both datasets, the features selected by the DNN method for FS and by the PCA method are diversified in terms of harmonic orders;
- For both datasets, the features selected by the MI and LDA methods are related to odd-order harmonics, which describe the power supply structures included in most of the HEAs;
- For the sequential forward FS method, our experiments compare the results provided by the Euclidean-based KNN classifier (where the neighborhood parameter is set to $K = 7$) and to the LDA classifier. The number of nearest neighbors is set to 7 because it is the closest odd number to the number of instances in a class in the proposed dataset, so that for each neighborhood, there is a majority vote. The selected features are those that reach the maximum accuracy [47] reported in Tables 2 and 3. For our dataset, only 12 features allow maximizing the KNN classifier accuracy and 18 features maximize the accuracy of the LDA classifier (see Figure 5). For the PLAID dataset, 25 features allow maximizing the KNN classifier accuracy and 33 features maximize the LDA classifier accuracy (see Figure 6). The low accuracy reached by the LDA classifier in the PLAID dataset can be explained by the unbalanced training sets (where one or several classes outnumber the other classes) [48]. Indeed, LDA is known to not provide good performances in this setting since classification is generally biased towards the majority classes. It is observed that for the LDA classifier, the accuracies less rapidly reach a plateau than for the KNN. Indeed, the KNN is known to be affected by the overfitting phenomenon [49,50], which increases the distance between individuals of the same class and decreases the accuracy. The classifier cannot deal with the feature relevance and is more sensitive to FS than other classifiers such as LDA, which can handle irrelevant features [51].

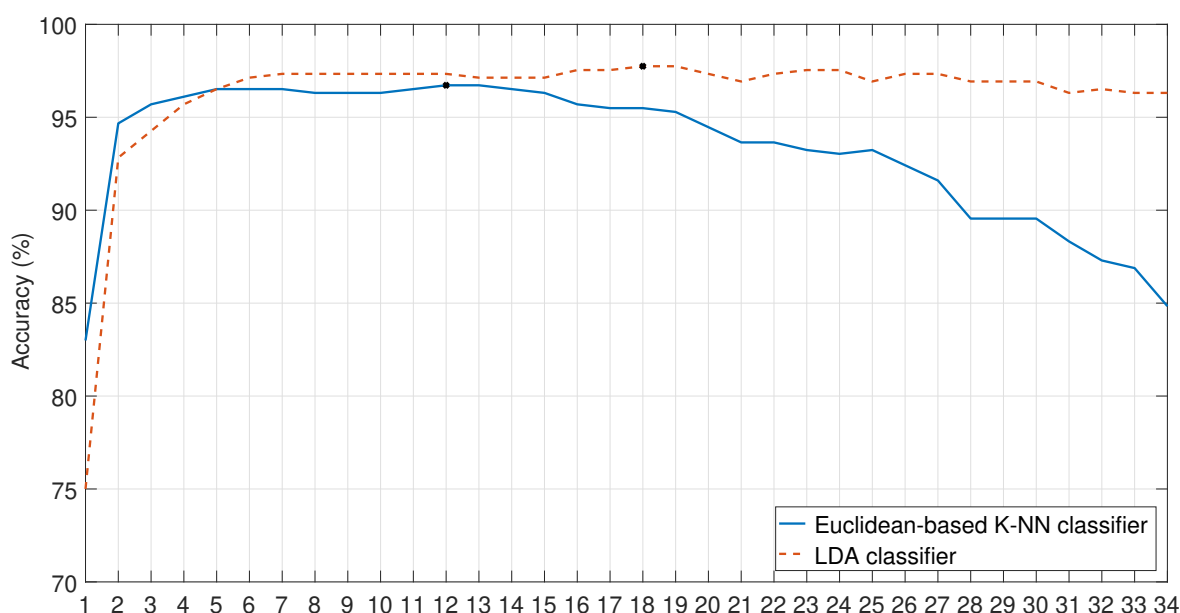


Figure 5. Classification success rate as a function of the number of the features on our dataset using KNN and LDA classifiers with the subset of $p = 34$ features meeting the additive criterion.

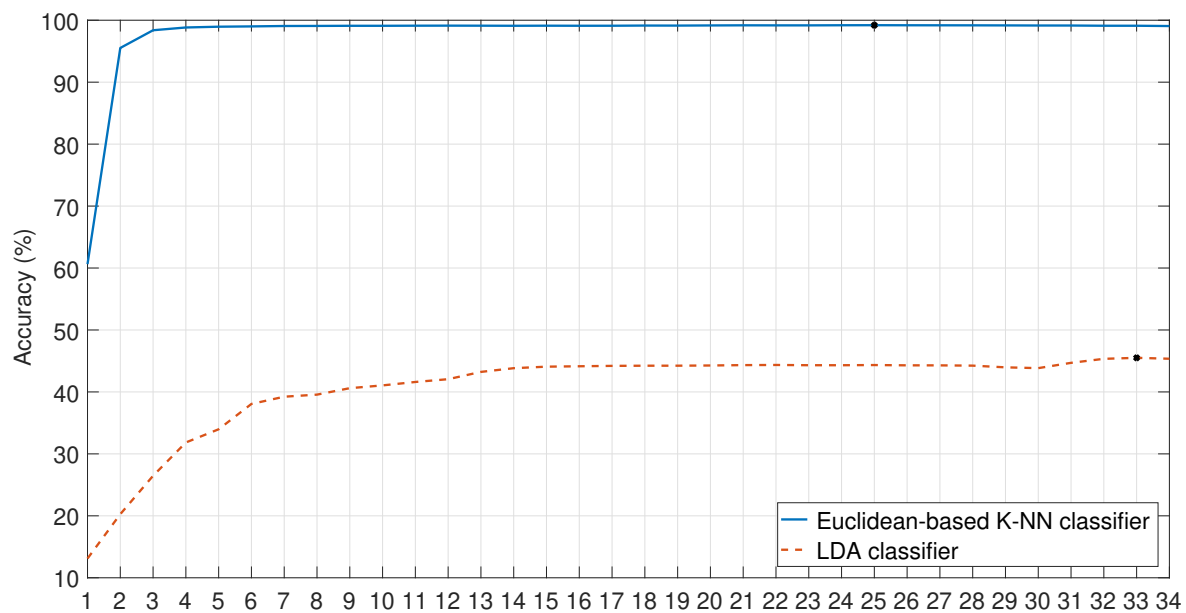


Figure 6. Classification success rate as a function of the number of the features on the PLAID dataset using KNN and LDA classifiers with the subset of $p = 34$ features meeting the additive criterion.

5. Home Electrical Appliances Classification Results

5.1. Investigated Classification Methods

We investigate four classification methods that can be combined with the different feature subsets provided by the FS methods presented in Section 4.

- The KNN method is widely used by the NILM community for HEAs' identification [12,52]. We use the Euclidean distance and $K = 7$, which corresponds to the closest odd number to the number of instances in a class in the proposed dataset. Hence, the predicted class corresponds to the most represented one in the neighborhood through majority voting.
- The LDA method estimates the optimal linear combination between features using the eigenvectors of the projection matrix $(B + S)^{-1}B$ of dimension $p \times p$, where $B = \frac{1}{n} \sum_{k=1}^K n_k (g_k - g)(g_k - g)^T$ and $S = \frac{1}{n} \sum_{k=1}^K n_k V_k$, where V_k are the covariance matrices built from the corresponding n_k individuals (number of individuals of class $k \in \{1, \dots, K\}$); $g = \frac{1}{n} \sum_{i=1}^n \bar{x}_i$ and $g_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \bar{x}_i$ correspond to the mean over all the individuals of the whole dataset $\bar{X}$ and the mean over all the individuals in the class k , respectively. Then, the tested individuals are projected into the discriminative linear space before being assigned to the class whose centroid is the closest in terms of the Euclidean distance.
- The proposed DNN classification method uses the same fully connected DNN architecture as presented in Section 4.1.3 for FS. Our implementation is based on tensorflow/keras (<https://keras.io/>). The training is completed with a batch size equal to 64 and a maximal number of 350 epochs (one epoch is reached each time the whole training dataset is processed once). The optimization is completed using the RMSprop algorithm [43] with a learning rate set to $\eta = 10^{-3}$.
- The Random Forest (RF) classification method creates a set of decision trees and aggregates the votes from the decision trees to predict the class of the tested individuals [53]. The number of trees was set to 5 after experimental tuning to get the best results.

Other classification methods such as Support Vector Machine (SVM) [8] or Adaboost [54] were not evaluated in this study because they require a very high computation cost, and our preliminary results did not reveal a significant accuracy improvement in comparison to the four investigated methods.

5.2. Classification Test Procedure

The evaluation of the classification performances uses an 8-fold cross validation methodology, which randomly splits the dataset into eight equal partitions, which are individually tested using the seven remaining partitions for training. This process is repeated eight times, until all subsamples have been used for cross validation [55]. The number of folds was arbitrarily set to 8 so that 12.5% of the studied dataset corresponds to the test set and the remaining 87.5% to the training set. The final metrics for classification performances were then computed by merging the results of each partition.

Our results are expressed in terms of the classical classification metrics used for NILM—Accuracy (Acc), $F_{Measure}$ (F_M), Recall (Rec) and Precision ($Prec$) [40,47]—which are deduced from the computed confusion matrices. Thus, if we denote C a resulting confusion matrix of dimension $I \times I$ (I being the considered number of classes), where C_{ij} corresponds to the number of individuals of the true class i (row) predicted as being in the class j (column), the Accuracy, Precision, and Recall are computed as:

$$Acc = \frac{\sum_i C_{ii}}{\sum_i \sum_j C_{ij}} \quad Prec = \frac{1}{I} \sum_i \frac{C_{ii}}{\sum_j C_{ji}} \quad Rec = \frac{1}{I} \sum_i \frac{C_{ii}}{\sum_j C_{ij}} \quad (15)$$

where $n = \sum_i \sum_j C_{ij}$ is the total number of individuals. The F-measure (also called F- or F_1 -score) is the harmonic mean of the Precision and Recall, computed as:

$$F_M = 2 \frac{Prec \cdot Rec}{Prec + Rec} \quad (16)$$

We also present the ratio $\frac{Acc}{\#features}$, which measures the efficiency and allows us to judge on the right combination of FS method and classifier. Indeed, the goal is to achieve the highest accuracy for a small number of descriptors. A large ratio means that a high accuracy is reached with a small number of selected features. Each method is evaluated on both datasets.

5.3. Self-Database Results

Sections 5.3.1 and 5.3.2 present the classification success rate as a function of the provided subset of selected features used to describe each HEA. Overall results using our new proposed dataset are summarized in Table 4 for which the details are provided in Table 5. The results obtained using the PLAID dataset are summarized in Table 6 where the details are presented in Table 7. The confusion matrices of the best methods are also presented in Appendix A. We also compare the results with the active and reactive powers (P,Q) which are the usual features proposed in the NILM literature [56]. The results show the clear improvement provided by the FS methods on each evaluated dataset. In addition, to study how noise influences our training, we have trained the studied classifiers using the selected features and Data Augmentation (DA) [57]. For this, both studied datasets are 100 % augmented by adding a white Gaussian noise to the current signals of each HEA class, to obtain an Signal to Noise Ratio (SNR) = 20 dB (defined as $10 \log_{10} \left(\frac{\|x\|^2}{\|b\|^2} \right)$, with x the original signal amplitude and b the noise signal). This leads to new augmented datasets containing noisy current signals, from which we compute the 34 features meeting the additive criterion. Then, for each original dataset, we apply cross-validation for classification evaluation using the selected features in Section 4.2 and the classification methods used in Section 5. The experiment setup also consists of an 8-fold cross-validation experiment, where the original datasets were partitioned into eight. For

each of the 8 simulations, 20% of each HEA class of the noisy generated dataset was added to the training set. No noise was added to the test part that was used for performance generalization for each of the eight simulations in our experiments. The success rate of a classification method for a specific subset of selected features was obtained by calculating the average scores of the performances metrics used in Section 5. All the results presented in the following tables are ranked in descending order of the best F_M scores.

5.3.1. Proposed Dataset

Table 5 shows that the RF classifier outperforms all the other classifiers. This is consistent with the findings of X. Wu et al. in [53], where an accuracy equal to 98.0% was found using eight steady-state features (which do not meet the additive criterion). These authors showed the advantages of RF classifier over KNN. In our experiment, RF classifier obtains the best recognition rate equal to 99.18% with the features selected by the MI method and DA, and 98.15% of accuracy when combined with LDA feature selection method without DA. This leads to a ratio (Acc/# feat.) of 7.01 and a computational time that is very low. It can be observed that the classifier performances are also better when data is artificially augmented during training. The confusion matrix depicted in Figure A1 shows that 20% of tested individuals that belong to class index 4 “Fan-Coala level 1” are classified as class index 5 “Fan -Coala Level 2” and 10% of the tested individuals belonging to class index 1 “Electric mixer Moulinex” are classified as class index 7 “Fan-Coala Level 3”. Interestingly, the performance of the DNN method is improved by a suitable choice of relevant descriptors as confirmed by the usage of the MI and the DNN features selection methods. Indeed, DNN classifier usually obtains the best results when using all the considered features. This point is of interest to develop new strategies to improve the training efficiency of DNN when used on a small training dataset. Finally, it can be observed that the results obtained with the DNN classifier are lower than those obtained for the PLAID dataset in Table 7. This can be explained by the small size of the training dataset. Indeed, DNN is known to require a large number of data to efficiently be trained. It can also be observed that for most of the classifiers, data augmentation by the addition of white Gaussian noise can significantly improve the results. Table 4 depicts the average F_M scores obtained for each FS method considering all the studied classifiers. It can be seen that the odd order harmonics features selected by the MI method are the ones for which the best average F_M score is reached.

Table 4. Average F_M scores obtained for the several different feature subsets selected for the proposed dataset (61 classes, $n = 488$) over all considered classifiers. Results are sorted in descending order of F-measure.

Feature Selection Method	F_M Average (%)
MI	92.96
KNN based Seq. forw. FS method	92.58
LDA	92.42
All features	90.92
LDA-based Seq. forw. FS method	90.14
PCA	87.64
P,Q features	86.62
DNN	84.87

Table 5. Performance (in percentage) of the classification methods applied on the proposed dataset using different feature subsets from the additive feature set (61 classes, $n = 488$). Results are sorted in descending order of F-measure.

Methods (Selected features/Classifier)				Results				
Feat.	# feat.	Classifier	D.A	Acc.	F _M	Rec.	Pre.	Acc. #feat.
MI	20	R.F	Yes	99.18	99.17	99.18	99.30	4.95
LDA	14	R.F	Yes	98.56	98.52	98.56	98.71	7.04
P,Q features	2	R.F	Yes	98.15	98.15	98.15	98.24	49.07
All features	34	R.F	Yes	98.15	98.14	98.15	98.28	2.88
KNN-based Seq. forw. FS (K = 7)	12	R.F	Yes	97.74	97.71	97.74	97.99	8.14
LDA	14	R.F	No	98.15	97.67	97.43	98.15	7.01
KNN-based Seq. forw. FS (K = 7)	12	R.F	No	98.15	97.60	97.33	98.15	8.17
MI	20	R.F	No	98.15	97.54	97.23	98.15	4.90
DNN	27	R.F	Yes	97.34	97.32	97.34	97.60	3.60
P,Q features	2	R.F	No	97.74	97.06	96.72	97.74	48.87
MI	20	LDA	Yes	96.92	96.83	96.92	97.55	4.84
LDA-based Seq. forw. FS	18	LDA	No	97.54	96.72	96.31	97.54	5.41
DNN	27	LDA	Yes	96.72	96.62	96.72	97.49	3.58
LDA-based Seq. forw. FS	18	LDA	Yes	96.72	96.57	96.72	97.43	5.37
LDA-based Seq. forw. FS	18	R.F	Yes	96.31	96.29	96.31	96.77	5.35
PCA	17	R.F	Yes	96.31	96.25	96.31	96.70	5.66
All features	34	LDA	Yes	96.31	96.18	96.31	96.84	2.83
DNN	27	LDA	No	96.93	96.14	95.77	96.91	3.59
LDA	14	LDA	Yes	95.69	95.59	95.69	96.32	6.83
KNN-based Seq. forw. FS (K = 7)	12	KNN	No	96.72	95.76	95.29	96.72	8.06
All features	34	KNN	Yes	95.69	95.53	95.69	96.82	2.81
MI	20	KNN	Yes	95.49	95.27	95.49	96.73	4.77
DNN	27	R.F	No	96.31	95.22	94.67	96.31	3.57
MI	20	LDA	No	96.31	95.21	94.67	96.31	4.81
LDA-based Seq. forw. FS	18	R.F	No	96.31	95.21	94.67	96.31	5.35
MI	20	DNN	No	96.31	95.08	94.47	96.31	4.81
KNN-based Seq. forw. FS (K = 7)	12	KNN	Yes	95.28	95.02	95.32	96.68	7.94
LDA	14	KNN	Yes	95.28	95.01	95.28	96.65	6.80
LDA-based Seq. forw. FS	18	KNN	Yes	95.27	95.01	95.28	96.49	5.29
P,Q features	2	KNN	Yes	95.28	94.97	95.28	96.66	47.64
All features	34	LDA	No	96.10	94.94	94.36	96.10	2.82
All features	34	R.F	No	96.10	94.91	94.33	96.10	2.82
PCA	17	R.F	No	95.90	94.87	94.39	95.90	5.64
KNN-based Seq. forw. FS (K = 7)	12	LDA	Yes	95.28	94.86	95.28	96.34	7.94
KNN-based Seq. forw. FS (K = 7)	12	LDA	No	95.90	94.60	93.95	95.90	7.99
PCA	17	LDA	Yes	94.46	94.37	94.46	95.48	5.55
LDA	14	LDA	No	95.69	94.33	93.65	95.69	6.83
P,Q features	2	KNN	No	95.08	93.44	92.62	95.08	47.54
PCA	17	LDA	No	94.67	93.06	92.28	94.67	5.57

Table 5. Cont.

Methods (Selected features/Classifier)				Results				
Feat.	# feat.	Classifier	D.A	Acc.	F _M	Rec.	Pre.	Acc. #feat.
LDA	14	KNN	No	94.46	92.96	92.21	94.46	6.74
PCA	17	DNN	No	93.64	92.15	91.48	93.65	5.51
P,Q features	2	LDA	Yes	93.23	91.85	93.23	93.69	46.61
All features	34	DNN	No	93.24	91.34	90.44	93.24	2.74
P,Q features	2	LDA	No	93.23	90.98	89.85	93.23	46.61
KNN-based Seq. forw. FS (K = 7)	12	DNN	No	92.62	90.40	89.34	92.62	7.72
MI	20	KNN	No	91.60	89.21	88.01	91.59	4.58
LDA-based Seq. forw. FS	18	DNN	No	90.98	88.48	87.39	90.98	5.05
LDA	14	DNN	No	92.01	89.89	88.99	92.01	6.57
DNN	27	KNN	Yes	87.09	86.18	87.09	89.46	3.23
LDA-based Seq. forw. FS	18	KNN	No	88.11	84.35	82.51	88.11	4.89
DNN	27	DNN	No	86.27	82.70	88.99	92.01	3.20
DNN	27	KNN	No	85.86	81.39	79.17	85.86	3.18
All features	34	KNN	No	84.43	80.31	78.36	84.43	2.48
PCA	17	KNN	Yes	81.14	79.87	81.14	82.58	4.77
PCA	17	KNN	No	82.37	78.18	76.25	82.37	4.84
All features	34	DNN	Yes	77.05	76.07	77.05	79.51	2.27
LDA	14	DNN	Yes	76.84	75.40	76.84	77.41	5.49
MI	20	DNN	Yes	76.84	75.33	76.84	77.18	3.84
KNN-based Seq. forw. FS (K = 7)	12	DNN	Yes	76.02	74.70	76.02	77.51	6.33
PCA	17	DNN	Yes	74.59	72.35	74.59	75.47	4.39
LDA-based Seq. forw. FS	18	DNN	Yes	70.70	68.45	70.70	70.50	3.93
P,Q features	2	DNN	No	71.52	66.43	64.28	71.52	35.76
P,Q features	2	DNN	Yes	63.32	60.05	63.32	59.19	31.66
DNN	27	DNN	Yes	45.08	43.36	45.08	45.34	1.67

5.3.2. PLAID Dataset

In Table 7, we can notice the improvement brought by FS approaches combined with the KNN classifier. Indeed, the obtained ratios (Acc/# feat.) are all greater, which denotes the fact that KNN classifier needs a small number of features to reach excellent accuracy. The best ratio (Acc/# feat.) of 47.29 is obtained when using KNN with P, Q features, which are usually considered for HEAs' identification in the NILM literature, but the reached accuracy is only of 94.58 %. The best accuracy of 99.19% is reached for only 25 features selected by the KNN-based sequential forward-FS method with and without data augmentation (according to the confusion matrix depicted in Figure A2, 10% of the tested individuals belonging to class index 7 "Incandescent light bulb-Electrix-soft white" are classified as class index 8 "Incandescent light bulb-Philips Duramax"). However, an accuracy of 99.13 % with a very good ratio $\frac{Acc}{\#feat} = 4.96$ is obtained when considering the subset of 20 features selected by the MI method with the KNN classifier. As we seek to achieve the best identification rates (close to 99 % of accuracy) for the smallest number of selected features, this combination classifier/ selected features is a good trade-off. This allows the KNN classifier to slightly exceed the performances reached by the RF classifier with the 20 features selected by the MI method (98.91 % of accuracy).

In contrast to our dataset, the high number of individuals ($n = 36720$) eases the DNN to learn the features from the dataset and to figure out that an important number of features is reliable and should be used. Our results outperform the previous ones reported in [58] for the PLAID dataset using the VI trajectories features, where an F-measure of $F_M \approx 77\%$ is given. This validates the efficiency of our approach consisting in carefully selecting features before classification. Combining RF classifier with features selected by the KNN-based sequential forward FS method also allows us to exceed the performances obtained in [18], where authors also used an RF classifier applied on PLAID dataset using an optimal subset of 20 steady-state and transient features (selected through a systematic feature elimination process) and were able to reach an accuracy of 93.2%. The performances obtained by the LDA are the worst ones. As mentioned, this can be explained by the fact that the LDA classifier struggles when the number of individuals is very large, resulting in individuals overlapping between classes. In almost all the cases, data augmentation improves or slightly modifies the classification performance for all methods (exception for the KNN classifier). Finally, Table 6 depicts the average F_M scores obtained for each FS method considering all the studied classifiers. The same observation as the one made for the proposed dataset can be made: the odd order harmonics features selected by the MI method are the ones for which the best average F_M score is reached.

Table 6. Average F_M scores obtained for the several different feature subsets selected for the PLAID dataset (71 classes, $n = 36,720$) over all considered classifiers. Results are sorted in descending order of F-measure.

Feature Selection Method	F_M Average (%)
MI	85.80
LDA based Seq. forw. FS method	85.07
KNN based Seq. forw. FS method	85.05
All features	84.56
PCA	83.53
DNN	82.52
LDA	81.61
P,Q features	58.31

Table 7. Performance (in percentage) of the classification methods applied on PLAID dataset using different feature subsets from the additive feature set (71 classes, $n = 36,720$). Results are sorted in descending order of F-measure.

Methods (Selected Features/Classifier)					Results			
Feat.	# feat.	Classifier	D.A	Acc.	F_M	Rec.	Pre.	Acc. #feat.
KNN-based Seq. forw. FS (K = 7)	25	KNN	No	99.19	98.63	98.81	98.56	3.97
KNN-based Seq. forw. FS (K = 7)	25	KNN	Yes	99.09	98.54	98.25	98.90	3.96
LDA-based Seq. forw. FS	33	KNN	No	99.07	98.50	98.71	98.39	3.00
LDA-based Seq. forw. FS	33	KNN	Yes	99.02	98.50	98.21	98.83	3.00
All features	34	KNN	No	99.07	98.49	98.69	98.40	2.91
PCA	30	KNN	No	99.00	98.46	98.60	98.41	3.30
DNN	17	KNN	No	98.99	98.45	98.62	98.41	5.82
MI	20	KNN	No	99.13	98.43	98.43	98.53	4.96
DNN	17	KNN	Yes	98.94	98.41	98.14	98.74	5.82
All features	34	KNN	Yes	99.07	98.38	98.09	98.81	2.91

Table 7. Cont.

Methods (Selected Features/Classifier)				Results				
Feat.	# feat.	Classifier	D.A	Acc.	F _M	Rec.	Pre.	Acc. #feat.
MI	20	KNN	Yes	99.03	98.33	98.23	98.48	4.95
PCA	30	KNN	Yes	98.95	98.28	98.00	98.69	3.30
MI	20	R.F	No	98.91	98.27	98.50	98.18	4.95
LDA	12	KNN	No	98.92	98.09	98.06	98.18	8.24
LDA-based Seq. forw. FS	33	R.F	No	98.75	97.97	98.20	97.91	2.99
All features	34	R.F	No	98.79	97.88	98.09	97.89	2.91
KNN-based Seq. forw. FS (K = 7)	25	R.F	No	98.80	97.85	98.00	97.85	3.95
LDA	12	KNN	Yes	98.76	97.84	97.64	98.06	8.23
DNN	17	R.F	No	98.58	97.74	98.14	97.54	5.80
PCA	30	R.F	No	98.67	97.66	97.84	97.60	3.29
LDA	12	R.F	No	98.60	97.62	97.99	97.51	8.22
LDA-based Seq. forw. FS	33	R.F	Yes	98.23	97.22	96.88	97.70	2.98
KNN-based Seq. forw. FS (K = 7)	25	R.F	Yes	98.19	97.14	96.69	97.78	3.93
All features	34	R.F	Yes	98.25	97.04	96.67	97.53	2.89
PCA	30	R.F	Yes	98.16	96.90	95.53	97.43	3.27
MI	20	R.F	Yes	98.19	96.89	96.48	97.34	4.91
LDA	12	R.F	Yes	97.88	96.86	96.43	97.40	8.16
MI	20	DNN	Yes	98.32	96.78	96.38	97.63	4.92
DNN	17	R.F	Yes	97.96	96.61	96.07	97.62	5.76
All features	34	DNN	No	98.04	96.35	96.64	96.67	2.88
KNN-based Seq. forw. FS (K = 7)	25	DNN	No	97.56	96.30	96.59	96.58	3.90
MI	20	DNN	No	96.26	96.05	96.99	96.28	4.81
LDA-based Seq. forw. FS	33	DNN	No	96.29	94.40	95.42	94.85	2.89
LDA	12	DNN	No	95.52	94.10	95.10	94.09	7.96
LDA	12	DNN	Yes	95.51	92.88	91.57	95.88	7.96
DNN	17	DNN	No	94.55	92.41	94.03	92.52	5.56
PCA	30	DNN	No	94.45	92.15	92.85	92.63	3.15
LDA-based Seq. forw. FS	33	DNN	Yes	94.81	91.66	90.72	94.43	2.87
KNN-based Seq. forw. FS (K = 7)	25	DNN	Yes	94.87	91.22	90.29	93.27	3.79
P,Q features	2	KNN	Yes	94.43	90.91	91.75	91.68	47.22
P,Q features	2	KNN	No	94.58	90.82	91.39	90.86	47.29
PCA	30	DNN	Yes	91.19	90.82	88.99	94.33	3.04
DNN	17	DNN	Yes	93.28	90.52	88.14	94.84	5.49
P,Q features	2	R.F	No	93.79	90.31	90.50	90.60	46.90
All features	34	DNN	Yes	90.33	86.33	85.00	89.60	2.66
P,Q features	2	R.F	Yes	90.63	85.84	86.14	85.70	45.32
LDA-based Seq. forw. FS	33	LDA	No	45.52	53.46	59.45	54.25	1.38
All features	34	LDA	No	45.31	53.25	59.42	54.05	1.33
MI	20	LDA	No	44.55	53.11	58.49	53.84	2.23
KNN-based Seq. forw. FS (K = 7)	25	LDA	No	44.37	51.50	57.62	52.32	1.77

Table 7. Cont.

Methods (Selected Features/Classifier)				Results				
Feat.	# feat.	Classifier	D.A	Acc.	F _M	Rec.	Pre.	Acc. #feat.
P,Q features	2	DNN	Yes	65.56	50.48	51.27	55.29	32.79
P,Q features	2	DNN	No	65.52	49.75	52.58	52.45	32.76
PCA	30	LDA	No	42.54	48.58	54.28	49.23	1.42
LDA-based Seq. forw. FS	33	LDA	Yes	42.89	48.87	49.68	57.03	1.30
All features	34	LDA	Yes	42.91	48.78	49.75	56.29	1.26
KNN-based Seq. forw. FS (K = 7)	25	LDA	Yes	42.72	48.67	49.61	55.40	1.71
MI	20	LDA	Yes	42.45	48.50	49.49	53.77	2.12
DNN	17	LDA	No	40.65	46.79	54.21	46.85	2.39
PCA	30	LDA	Yes	40.59	45.42	45.64	52.09	1.35
LDA	12	LDA	No	38.66	40.40	47.43	40.75	3.22
DNN	17	LDA	Yes	37.09	39.20	39.66	45.50	2.18
LDA	12	LDA	Yes	36.08	35.88	35.49	41.93	3.00
P,Q features	2	LDA	Yes	12.12	4.29	7.11	4.19	6.06
P,Q features	2	LDA	No	11.59	4.05	3.65	7.55	5.80

5.4. Transfer Learning Results

Now, we propose evaluating if the features selected from a dataset are able to be transferred to another dataset. For this, a cross-learning strategy is adopted for both studied datasets. The goal is to study to what extent the features selected for a particular dataset are invariant across HEAs and can be used to get good classification performances in another dataset. Indeed, as several common features are selected from each dataset separately, we assume that they convey common information that can be transferred from one dataset to another. This approach allows us to reduce the number of training samples from unknown HEAs. First, the subsets of selected features from the proposed dataset are tested using the several classifiers presented previously and applied to the PLAID dataset. The results are presented in Table 8. Very good results are obtained with the KNN, the DNN, and the RF classifiers with the different subsets of selected features. The best ones are obtained with the KNN classifier and particularly when combined to the MI feature selection method, which allows us to obtain the best performances with only 20 features. Figure A3 shows the tested individuals that were misclassified, such as the following: 10% of the individuals belonging to class index 7 “Incandescent bulb light - Electrix soft white” were classified as class index 8 “Incandescent light bulb- Philips Duramax” or 10% of the individuals in class index 39 “Laptop-HP-C24” were classified as class index 40 “Laptop-Apple macbook air”. Second, the subsets of selected features from the PLAID dataset were tested using the several classifiers presented previously and applied to the proposed dataset. The results are presented in Table 9. Very good results were obtained with the RF and the LDA classifiers with the different subsets of selected features. An accuracy of 98.77% was obtained when combining the RF classifier to the subset of features selected with the LDA sequential forward-FS method. The confusion matrix in Figure A4 shows that 50% of the individuals belonging to class index 5 “Fan-Coala Level 2” were classified as class index 4 “Fan-Coala Level 1”, and 50% of the individuals that belong to class index 53 “Washing machine-LG state (b)” were classified as class index 58 “Washing machine-LG state (g)”. Through cross transfer learning, we show that in both situations, the features selected from another dataset improve the classification rate. There is therefore transfer of knowledge on the discriminating “power” of the selected features.

Table 8. Performance (in percentage) of the classification methods applied on the PLAID dataset using the different feature subsets selected in the proposed dataset. Results are sorted in descending order of F-measure.

Methods (Selected Features/Classifier)			Results				
Feat.	# feat.	Classifier	Acc.	F_M	Rec.	Pre.	Acc. #feat.
MI	20	KNN	99.12	98.42	98.40	98.53	4.96
LDA	14	KNN	99.10	98.36	98.53	98.31	7.08
KNN-based Seq. forw. FS (K = 7)	12	R.F	98.99	98.36	98.48	98.35	8.25
KNN-based Seq. forw. FS (K = 7)	12	KNN	99.04	98.29	98.39	98.30	8.25
PCA	17	KNN	98.95	98.28	98.54	98.17	5.82
DNN	27	KNN	98.95	98.20	98.42	98.09	3.66
LDA-based Seq. forw. FS	18	KNN	98.92	98.14	98.331	98.13	5.49
LDA	14	R.F	98.96	97.95	98.09	97.94	7.07
MI	20	R.F	98.89	97.86	97.90	97.91	4.94
PCA	17	R.F	98.77	97.84	98.04	97.79	5.81
LDA-based Seq. forw. FS	18	R.F	98.76	97.64	97.79	97.62	5.49
DNN	27	R.F	98.68	97.61	97.84	97.53	3.65
KNN-based Seq. forw. FS (K = 7)	12	DNN	98.08	96.83	97.40	96.87	8.17
LDA	14	DNN	97.85	96.35	97.10	96.32	6.99
MI	20	DNN	94.85	94.82	95.86	95.10	4.74
PCA	17	DNN	96.87	94.68	95.55	94.89	5.70
DNN	27	DNN	93.76	93.66	94.86	94.17	3.47
LDA-based Seq. forw. FS	18	DNN	95.76	93.18	94.22	93.43	6.32
MI	20	LDA	44.14	52.43	58.21	53.18	2.21
DNN	27	LDA	42.36	48.29	54.22	48.48	1.57
PCA	17	LDA	41.84	43.63	50.10	43.87	2.46
LDA	14	LDA	40.75	43.54	48.95	44.3	2.46
KNN-based Seq. forw. FS (K = 7)	12	LDA	37.73	38.23	42.81	40.30	3.14
LDA-based Seq. forw. FS	18	LDA	38.65	37.70	43.20	38.74	2.15

Table 9. Performance (in percentage) of the classification methods applied on the proposed dataset using the different feature subsets selected in PLAID dataset. Results are sorted in descending order of F-measure.

Methods (Selected features/Classifier)			Results				
Feat.	# feat.	Classifier	Acc.	F_M	Rec.	Pre.	Acc. #feat.
LDA-based Seq. forw. FS	33	R.F	98.77	98.36	98.15	98.77	2.99
MI	20	R.F	98.76	98.36	98.15	98.77	4.93
KNN-based Seq. forw. FS (K = 7)	25	R.F	97.74	96.99	96.61	97.74	3.90
PCA	30	LDA	96.51	95.46	94.95	96.51	3.21
LDA	12	R.F	96.51	95.45	94.94	96.51	8.04
LDA-based Seq. forw. FS	33	LDA	96.31	95.18	94.63	96.31	2.92
KNN-based Seq. forw. FS (K = 7)	25	LDA	96.31	95.15	94.56	96.31	3.85

Table 9. Cont.

Methods (Selected features/Classifier)			Results				
Feat.	# feat.	Classifier	Acc.	F_M	Rec.	Pre.	Acc. #feat.
MI	20	LDA	96.31	95.15	94.56	96.31	4.81
DNN	17	R.F	95.70	94.33	93.65	95.70	5.63
PCA	30	R.F	95.49	93.98	93.23	95.49	3.18
MI	20	DNN	93.85	92.14	91.32	93.85	4.69
LDA	12	LDA	93.64	92.07	91.29	93.64	7.80
DNN	17	LDA	93.44	91.67	90.78	93.44	5.50
LDA	12	LDA	92.62	90.50	89.48	92.62	7.72
KNN-based Seq. forw. FS (K = 7)	25	DNN	93.24	91.02	89.92	93.24	3.73
LDA-based Seq. forw. FS	33	DNN	92.62	90.67	89.75	92.62	2.80
MI	20	KNN	91.80	89.27	88.01	91.80	4.59
DNN	17	DNN	90.98	88.38	87.16	90.98	5.35
PCA	30	DNN	90.37	87.56	86.27	90.37	3.01
PCA	30	KNN	85.04	80.66	78.58	85.04	2.83
LDA-based Seq. forw. FS	33	KNN	84.42	79.84	77.66	84.42	2.55
LDA	12	KNN	83.81	79.57	77.52	83.81	6.98
KNN-based Seq. forw. FS (K = 7)	25	KNN	84.01	79.37	77.11	84.01	3.36
DNN	17	KNN	77.87	72.96	70.65	77.87	4.58

6. Conclusions

In this paper, we addressed one of the main challenges of the NILM problem consisting in the HEA identification from electrical measurements. To this end, we covered a broad sweep of existing supervised FS and classification methods to show the efficiency of this approach for identifying distinct HEAs using the suitable set of relevant features.

As a first contribution, in addition to a novel publicly available dataset, we introduced a comparative evaluation of a large number of methods in an event-based NILM context, involving all the possible combinations of these techniques applied on two distinct annotated HEA datasets where each HEA signature made of relevant features is extracted from different categories and manufacturers.

Second, thanks to our proposed data augmentation and feature selection, we have improved the best HEA identification results on the PLAID dataset by obtaining a resulting classification rate above 99%. To our knowledge, this result outperforms the best available results obtained with the PLAID dataset using state-of-the-art methods. Furthermore, in this regard, validating our solution using two datasets has helped in (i) showing its high performance despite the data collection procedure being different and (ii) proving its capability to give very good results even if HEAs are from different categories and manufacturers.

Moreover, our results show that the number of extracted features can significantly be reduced to efficiently perform HEAs recognition. A cross transfer learning strategy was adopted by using the subsets of features selected by FS approaches applied to our dataset to classify HEAs of the PLAID dataset and conversely. Very good results were obtained, which confirms that the features selected with one dataset can be transferred to another one.

Several conclusions can be safely drawn from this study. First, the selected electrical features can be justified by the power supply topologies included in an HEA (the front-end

circuitry that connects them to the power grid), which affect their current waveforms. Each subset of selected features contains for the most part odd-order harmonics related to power components. The performance of a designed classifier can be improved by the use of an optimal subset of features. Some features, such as P , P_1 , P_H , Q , Q_1 , and Q_H , are retrieved in most of the subsets of selected features, which shows their importance for HEAs' identification. Secondly, several observations can be made according to the chosen classifier: DNN requires many training data and works poorly on the small proposed dataset (and works better on the PLAID dataset); LDA is sensitive to unbalanced datasets, and it works less well on PLAID; KNN method is efficient and sensitive to the choice of descriptors; RF is a state-of-the-art method in automatic classification before the arrival of deep learning and is robust and gives good results in most cases. In addition, the augmentation of data shows that it is possible to improve in almost all cases classification performance for all methods (exception for PLAID with KNN classifier). Finally, features selected through FS methods in a dataset could be used to correctly identify unknown HEAs from another dataset using, for example, an unsupervised statistical modeling approach [6]. This addresses one of the biggest NILM challenges: generalization.

Author Contributions: Investigation, S.H., D.F. and F.A.; Methodology, S.H. and D.F.; Supervision, D.F., F.A., H.B.A.S. and L.M. They have read and agreed to the published version of the manuscript

Funding: This research was partly funded by Tunisian Ministry of Higher Education and Scientific Research (LSE-ENIT-LR11ES15) of University of Tunis El Manar and the French ANR (investissements d'avenir) ORACLES project with Univ. Paris-Saclay.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The matlab and python codes used in this research are available for academic purposes only by contacting the corresponding author (D. Fourer). The new dataset used in this study is freely available at <https://iee-dataport.org/documents/home-electrical-appliances-recordings-nilm>

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

Abbreviations

The following abbreviations are used in this manuscript:

BN	Batch Normalization
DA	Data Augmentation
DNN	Deep Neural Network
FS	Feature Selection
HEA(s)	Home Electrical Appliance(s)
KNN	K-nearest-neighbor
LDA	Linear Discriminant Analysis
MI	Mutual Information
NILM	Non-Intrusive Load Monitoring
PCA	Principal Component analysis
PCC	Point of Common Coupling
PLAID	Public Dataset of High Resolution for Load Identification Research
RELU[RELU]	REctified Linear Unit
RF	Random Forest
SNR	Signal-to-Noise Ratio

Appendix A. Confusion matrices

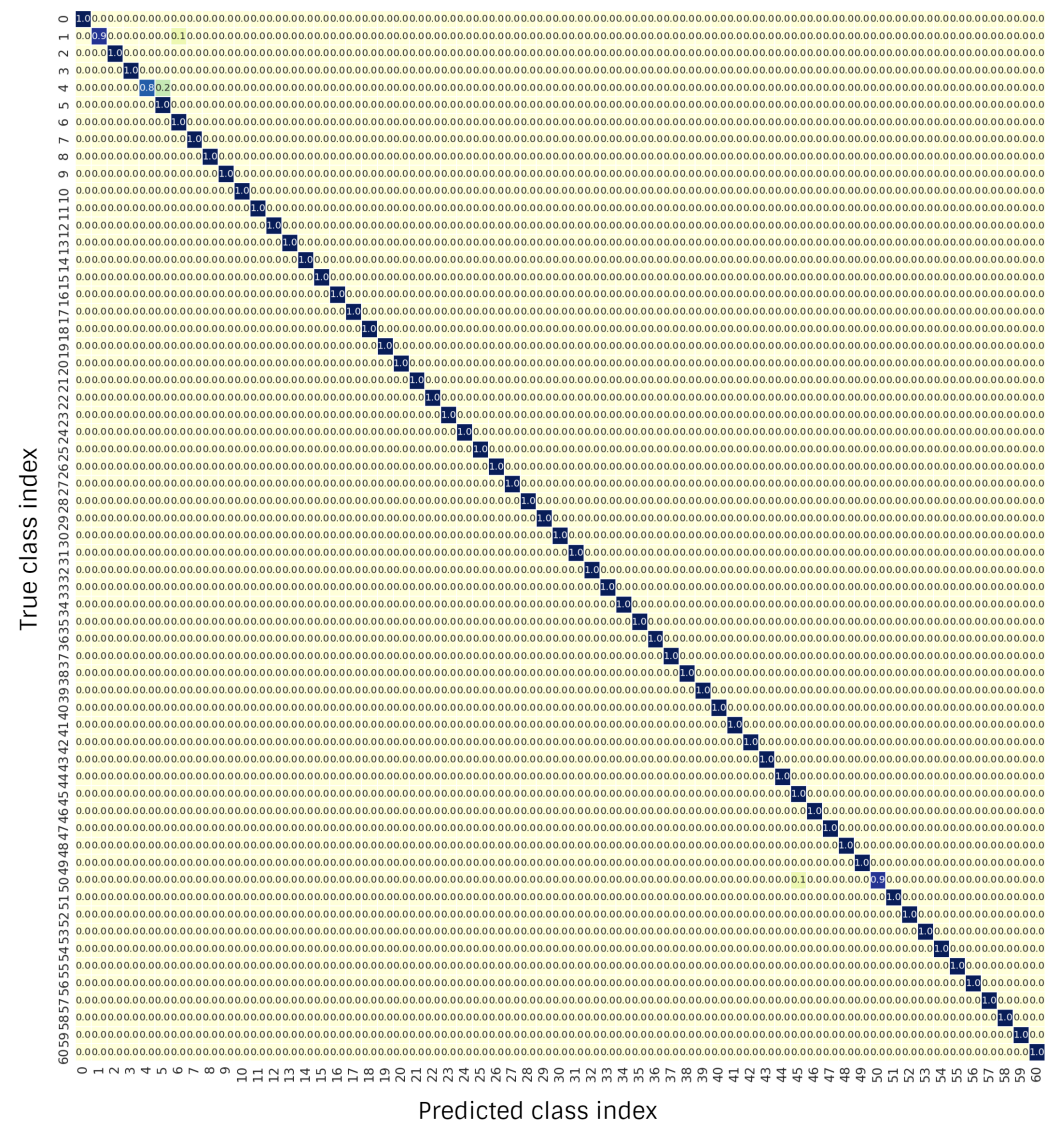


Figure A1. Confusion matrix from the proposed dataset obtained with RF classifier using the subset of 20 features selected with the MI FS method.

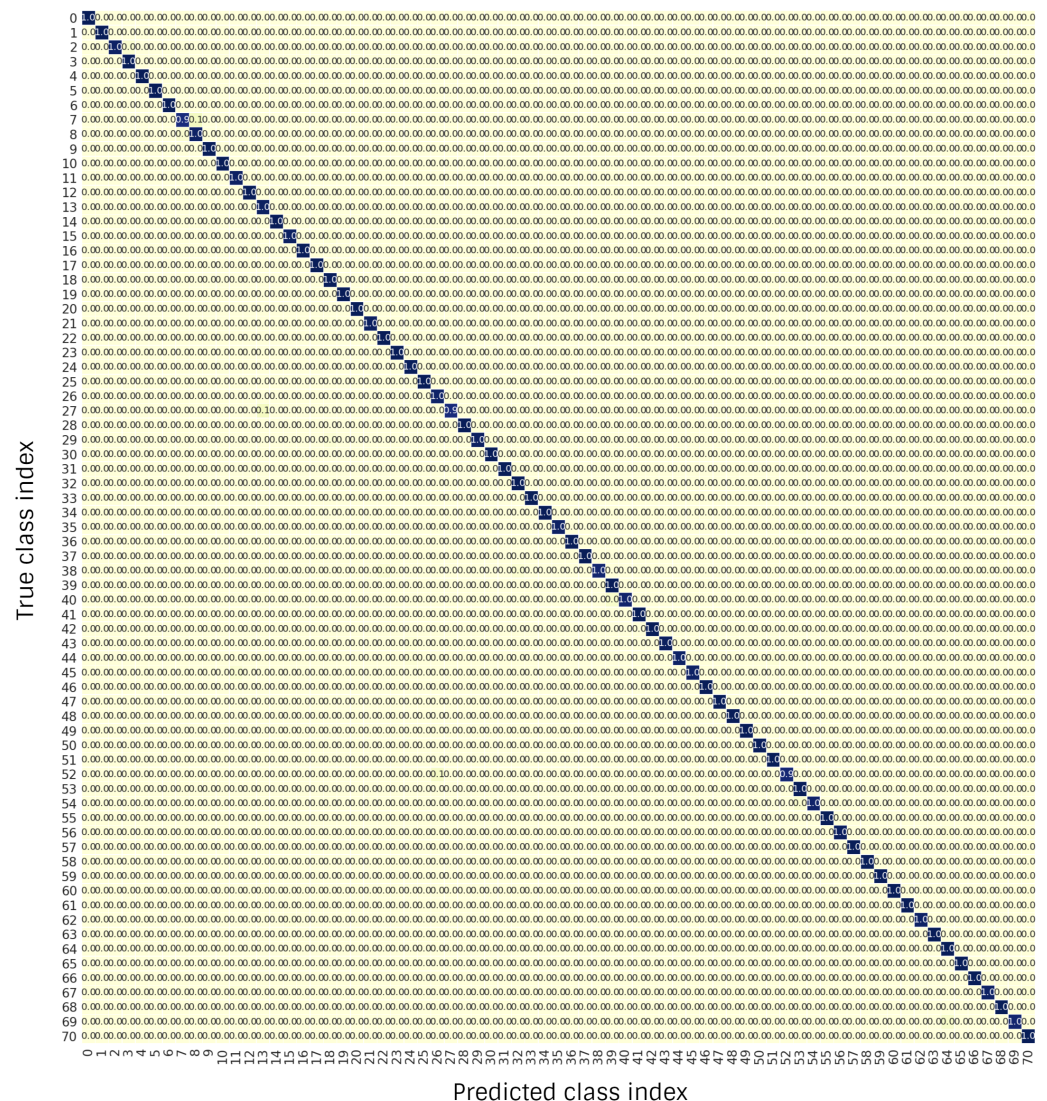


Figure A2. Confusion matrix from PLAID dataset obtained with KNN classifier using the subset of 25 features selected with the KNN based Seq. forw. FS method.

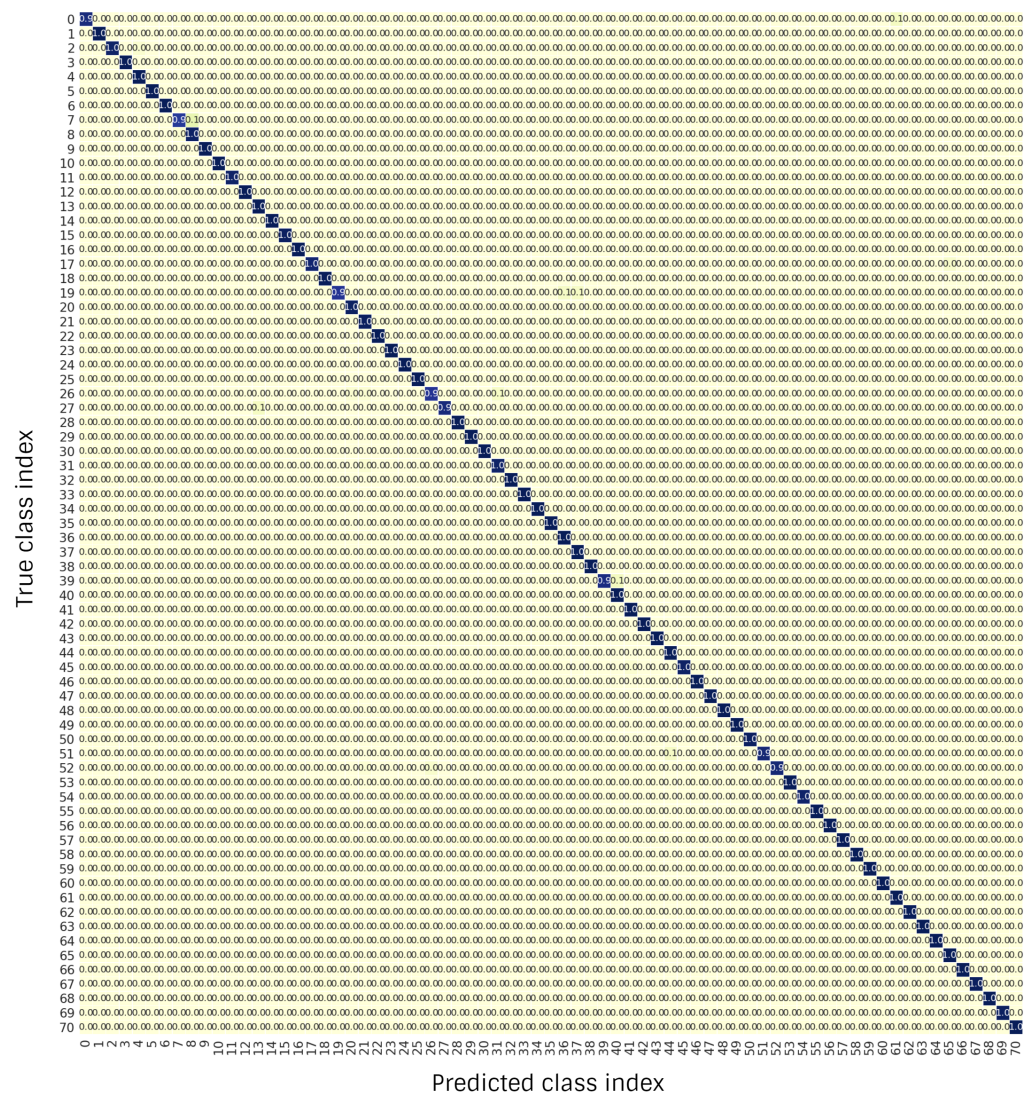


Figure A3. Confusion matrix from the PLAID dataset obtained with KNN classifier using the subset of 20 features selected with the MI FS method from the proposed dataset.

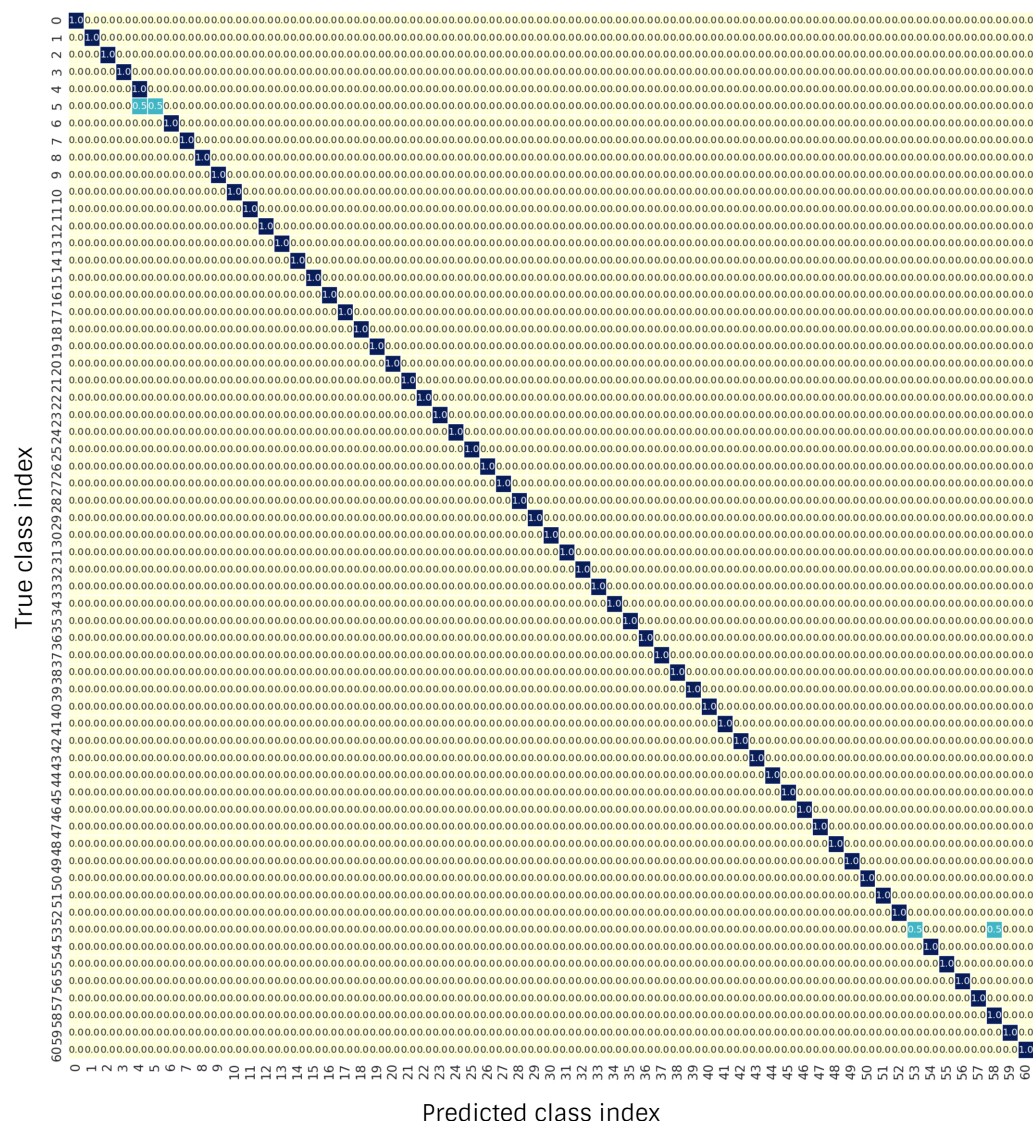


Figure A4. Confusion matrix from the proposed dataset obtained with the RF classifier using the subset of 33 features selected with the LDA based Seq. forw. FS method from the PLAID dataset.

References

1. Darby, S. The Effectiveness of Feedback on Energy Consumption. In *A Review for DEFRA of the Literature on Metering, Billing and direct Displays*; University of Oxford: Oxford, UK, 2006; Volume 486.
2. Wang, Y.; Chen, Q.; Hong, T.; Kang, C. Review of smart meter data analytics: Applications, methodologies, and challenges. *IEEE Trans. Smart Grid* **2018**, *10*, 3125–3148.
3. Hart, G.W. Non-intrusive appliance load monitoring. *Proc. IEEE* **1992**, *80*, 1870–1891.
4. Zoha, A.; Gluhak, A.; Imran, M.A.; Rajasegarar, S. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors* **2012**, *12*, 16838–16866.
5. Faustine, A.; Mvungi, N.H.; Kaijage, S.; Michael, K. A survey on non-intrusive load monitoring methodologies and techniques for energy disaggregation problem. *arXiv* **2017**, arXiv:1703.00785.
6. Houidi, S.; Auger, F.; Sethom, H.B.A.; Fourer, D.; Miègeville, L. Multivariate event detection methods for non-intrusive load monitoring in smart homes and residential buildings. *Energy Build.* **2020**, *208*, 109624.
7. Houidi, S.; Fourer, D.; Auger, F. On the Use of Concentrated Time-Frequency Representations as Input to a Deep Convolutional Neural Network: Application to Non Intrusive Load Monitoring. *Entropy* **2020**, *22*(9), 911.
8. Liang, J.; Ng, S.K.K.; Kendall, G.; Cheng, J.W.M. Load Signature Study Part I: Basic Concept, Structure, and Methodology. *IEEE Trans. Power Del.* **2010**, *25*, 551–560, doi:10.1109/TPWRD.2009.2033799.
9. Zeifman, M.; Roth, K. Non intrusive appliance load monitoring: Review and outlook. *IEEE Trans. Consum. Electron.* **2011**, *57*, 76–84.

10. Kalogridis, G.; Efthymiou, C.; Denic, S.; Cepeda, R. Privacy for smart meters: Towards undetectable appliance load signatures. In Proceedings of the First IEEE International Conference on Smart Grid Communications, Gaithersburg, MD, USA, 4–6 October 2010; pp. 232–237.
11. Ruano, A.; Hernandez, A.; Ureña, J.; Ruano, M.; Garcia, J. NILM Techniques for Intelligent Home Energy Management and Ambient Assisted Living: A Review. *Energies* **2019**, *12*, doi:10.3390/en12112203.
12. Figueiredo, M.; De Almeida, A.; Ribeiro, B. An Experimental Study on Electrical Signature Identification of Non-Intrusive Load Monitoring (NILM) Systems. In Proceedings of the 10th international Conference on Adaptive and Natural Computing Algorithms (ICANNGA), Ljubljana, Slovenia, 14–16 April 2011; pp. 31–40.
13. Do Nascimento, P.P.M. Applications of Deep Learning Techniques on NILM. Ph.D. Thesis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, 2016.
14. De Paiva Penha, D.; Castro, A.R.G. Home appliance identification for NILM systems based on deep neural networks. *Int. J. Artif. Intell. Appl.* **2018**, *9*, 69–80.
15. Piccialli, V.; Sudoso, A.M. Improving Non-Intrusive Load Disaggregation through an Attention-Based Deep Neural Network. *Energies* **2021**, *14*, 847.
16. Cannas, B.; Carcangiu, S.; Carta, D.; Fanni, A.; Muscas, C. Selection of Features Based on Electric Power Quantities for Non-Intrusive Load Monitoring. *Appl. Sci.* **2021**, *11*, doi:10.3390/app11020533.
17. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
18. Sadeghianpourhamami, N.; Ruyssinck, J.; Deschrijver, D.; Dhaene, T.; Develder, C. Comprehensive feature selection for appliance classification in NILM. *Energy Build.* **2017**, *151*, 98–106.
19. Houidi, S.; Auger, F.; Ben Attia Sethom, H.; Fourer, D.; Miègeville, L. Relevant feature selection for home appliances recognition. In Proceedings of the Electrimacs 2017 Conference, Nancy, France, 4–6 July 2017.
20. Kato, T.; Cho, H.S.; Lee, D.; Toyomura, T.; Yamazaki, T. Appliance Recognition from Electric Current Signals for Information-Energy Integrated Network in Home Environments. In Proceedings of the Smart Homes and Health Telematics (ICOST), Tours, France, 1–3 July 2009; pp. 150–157.
21. D’Incecco, M.; Squartini, S.; Zhong, M. Transfer learning for non-intrusive load monitoring. *IEEE Trans. Smart Grid* **2019**, *11*, 1419–1429.
22. Houidi, S. Classification des Charges Électriques Résidentielles en vue de Leur Gestion Intelligente et de Leur comptabilisation. Ph.D. Thesis, University of Nantes, Saint-Nazaire France, 2020.
23. Houidi, S.; Auger, F.; Sethom, H.B.A.; Miègeville, L.; Fourer, D.; Jiang, X. Statistical Assessment of Abrupt Change Detectors for Non Intrusive Load Monitoring. In Proceedings of the 2018 IEEE International Conference on Industrial Technology (ICIT), Lyon, France, 20–22 February 2018.
24. Molina, L.C.; Belanche, L.; Nebot, À. Feature selection algorithms: A survey and experimental evaluation. In Proceedings of the IEEE International Conference on Data Mining (ICDM), IEEE, Lyon, France, 20–22 February 2002; pp. 306–313.
25. Jain, A.; Zongker, D. Feature selection: Evaluation, application, and small sample performance. *IEEE Trans. Pattern Analysis.* **1997**, *19*, 153–158.
26. Gao, J.; Giri, S.; Kara, E.C.; Bergés, M. PLAID: A Public Dataset of High-resolution Electrical Appliance Measurements for Load Identification Research: Demo Abstract. In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, Seoul, Korea, 4–5 November 2014; pp. 198–199, doi:10.1145/2674061.2675032.
27. Renaux, D.P.B.; Pottker, F.; Ancelmo, H.C.; Lazzaretti, A.E.; Lima, C.R.E.; Linhares, R.R.; Oroski, E.; Nolasco, L.d.S.; Lima, L.T.; Mulinari, B.M.; et al. A Dataset for Non-Intrusive Load Monitoring: Design and Implementation. *Energies* **2020**, *13*, doi:10.3390/en13205371.
28. Langella, R.; Testa, A. IEEE standard definitions for the measurement of electric power quantities under sinusoidal, nonsinusoidal, balanced, or unbalanced conditions. *Rev. IEEE Std. 1459–2000* **2010**, *1*–40. doi:10.1109/IEEESTD.2010.5439063.
29. Eigeles, E.A. On the Assessment of Harmonic Pollution. *IEEE Trans. Power Del.* **1995**, *10*, 693–98.
30. Aha, D.W.; Bankert, R.L. A comparative evaluation of sequential feature selection algorithms. In *Learning From Data*; Springer: NY, USA, 1996; pp. 199–206.
31. Yu, L.; Liu, H. Efficient feature selection via analysis of relevance and redundancy. *J. Mach. Learn. Res.* **2004**, *5*, 1205–1224.
32. Cadenas, J.M.; Garrido, M.C.; Martínez, R. Feature subset selection filter—wrapper based on low quality data. *Expert Syst. Appl.* **2013**, *40*, 6241–6252.
33. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*; John Wiley and Sons Inc.: NY, USA, 1958; Volume 2.
34. Boutsidis, C.; Mahoney, M.W.; Drineas, P. Unsupervised feature selection for principal components analysis. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), Las Vegas, NV, USA, 24–27 August 2008; pp. 61–69.
35. Pohjalainen, J.; Rasanen, O.; Kadioglu, S. Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits. *Comput. Speech Lang.* **2015**, *29*, 145–171.
36. Blanchet, F.G.; Legendre, P.; Borcard, D. Forward selection of explanatory variables. *Ecology* **2008**, *89*, 2623–2632, doi:10.1890/07-0986.1.

37. Marcano-Cedeño, A.; Quintanilla-Domínguez, J.; Cortina-Januchs, M.; Andina, D. Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network. In Proceedings of the IECON 2010—36th Annual Conference on IEEE Industrial Electronics Society, Glendale, AZ, USA, 7–10 November 2010; pp. 2845–2850.
38. Kozbur, D. Testing-Based Forward Model Selection. *Am. Econ. Rev.* **2017**, *107*, 266–269, doi:10.1257/aer.p20171039.
39. Le, T.T.H.; Kim, Y.; Kim, H. Network intrusion detection based on novel feature selection model and various recurrent neural networks. *Appl. Sci.* **2019**, *9*, 1392.
40. Pereira, L.; Nunes, N. Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools-A review. *Wiley Interdiscip. Rev. Data Min. Know. Discov.* **2018**, 1–17, doi:10.1002/widm.1265.
41. Gevrey, M.; I., D.; S., L. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Modell.* **2003**, *160*, 249–264, doi:10.1016/S0304-3800(02)00257-0.
42. Peng, C.; Lin, G.; Zhai, S.; Ding, Y.; He, G. Non-Intrusive Load Monitoring via Deep Learning Based User Model and Appliance Group Model. *Energies* **2020**, *13*, 5629.
43. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
44. Gao, B.; Pavel, L. On the properties of the softmax function with application in game theory and reinforcement learning. *arXiv* **2017**, arXiv:1704.00805.
45. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv* **2015**, arXiv:1502.03167.
46. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
47. Pereira, L.; Nunes, N. A comparison of performance metrics for event classification in Non-Intrusive Load Monitoring. In Proceedings of the 2017 IEEE International Conference on Smart Grid Communications (SmartGridComm), Dresden, Germany, 23–27 October 2017; pp. 159–164.
48. Xue, J.; Titterton, M. Do unbalanced data have a negative effect on Linear Discriminant Analysis. *Pattern Recognit.* **2008**, *41*, 1558–1571.
49. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*; John Wiley and Sons Inc.: Hoboken, NJ, USA, 2001.
50. Theodoridis, S.; Koutroumbas, K. *Pattern Recognition*, 2nd ed.; Academic Press: Cambridge, MA, USA, 2003.
51. Lee, Y.; Lin, Y.; Wahba, G. Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. *J. Am. Stat. Assoc.* **2004**, *99*, 67–81.
52. Basu, K.; Debusschere, V.; Bacha, S.; Maulik, U.; Bondyopadhyay, S. Non intrusive load monitoring: A temporal multi-label classification approach. *IEEE Trans. Indust. Inform.* **2015**, *11*, 262–270.
53. Wu, X.; Gao, Y.; Jiao, D. Multi-Label Classification Based on Random Forest Algorithm for Non-Intrusive Load Monitoring System. *Processes* **2019**, *7*, 337.
54. Murata, H.; Onoda, T. Applying Kernel Based Subspace Classification to a Non-intrusive Monitoring for Household Electric Appliances. In Proceedings of the International Conference on Artificial Neural Networks (ICANN), Vienna, Austria, 21–25 August 2001.
55. Makonin, S.; Popowich, F. Non-intrusive load monitoring performance evaluation. *Energy Effic.* **2015**, *8*, 809–814.
56. Saitoh, T.; Osaki, T.; Konishi, R.; Sugahara, K. Current sensor based home appliance and state of appliance recognition. *SICE J. Control Meas. Sys. Integrat.* **2010**, *3*, 86–93.
57. Mignot, R.; Peeters, G. An Analysis of the Effect of Data Augmentation Methods: Experiments for a Musical Genre Classification Task. *Trans. Int. Soc. Mus. Inform. Ret.* **2019**, doi:10.5334/tismir.26.
58. De Baets, L.; Ruyssinck, J.; Develder, C.; Dhaene, T.; Deschrijver, D. Appliance classification using VI trajectories and convolutional neural networks. *Energy Build.* **2018**, *158*, 32–36.