Comprehensive Feature Selection for
Appliance Classification in NILM

N. Sadeghianpourhamami, J. Ruyssinck, D. Deschrijver, T. Dhaene,
C. Develder

Ghent University – imec, IDLab, Dept. of Information Technology,
Technologiepark Zwijnaarde 15, 9052 Ghent, Belgium

Abstract

Since the inception of non-intrusive appliance load monitoring (NILM), extensive research has focused on identifying an effective set of features that allows to form a unique appliance signature to discriminate various loads. Although an abundance of features are reported in literature, most works use only a limited subset of them. A systematic comparison and combination of the available features in terms of their effectiveness is still missing. This paper, as its first contribution, offers a concise and updated review of the features reported in literature for the purpose of load identification. As a second contribution, a systematic feature elimination process is proposed to identify the most effective feature set. The analysis is validated on a large benchmark dataset and shows that the proposed feature elimination process improves the appliance classification accuracy for all the appliances in the dataset compared to using all the features or randomly chosen subsets of features.

Keywords: Non-intrusive Load Monitoring, Feature Extraction, Feature
Selection

Email addresses: `nasrin.sadeghianpourhamami@UGent.be` (N. Sadeghianpourhamami),
`joeri.ruyssinck@UGent.be` (J. Ruyssinck), `dirk.deschrijver@UGent.be` (D. Deschrijver),
`tom.dhaene@UGent.be` (T. Dhaene), `chris.develder@UGent.be` (C. Develder)

1. Introduction

A breakdown of energy consumption at the appliance level is not only an essential requirement for energy providers, in designing practical demand response algorithms (e.g., taking into account human behavioral uncertainties, or targeting specific user and appliance groups), but it also benefits residential customers by providing them the necessary information for improving their energy consumption efficiency [1]. Non-intrusive (appliance) load monitoring (NILM) techniques are cost-effective solutions to obtain such information. The general framework of NILM starts from input measurements of total electricity consumption to eventually disaggregate it into the individual contributions of each load.

A crucial step in NILM is *feature extraction*, which applies signal processing techniques to extract features from voltage (V) and current (I) measurements. The ultimate goal of the feature extraction step is to derive a signature (using a feature or combination of features) that can uniquely identify the individual appliances. The performance of any NILM system depends on the uniqueness of the appliance signature compared to that of other devices. Hence, identification of such signature is crucial in improving the load discrimination capability of a NILM system. Although NILM has been the subject of research for over two decades, so far a systematic selection of the various electrical features proposed for effective discrimination of loads has not yet been presented. Identifying the most meaningful set of electrical parameters to distinguish all appliances still remains one of the major challenges in NILM [42]. In this paper, we tackle this issue and contribute with:

1. A concise and up-to-date review of the features reported in recent NILM literature (Section 2) and

2. A systematic signature identification algorithm based on a comprehensive dataset with diverse appliances and various households (Section 3.1).

3

2. State of the art on feature extraction

The seminal work by Hart *et al.* [2], has inspired extensive research on extracting features and developing discriminating algorithms for NILM purposes. Zeifman *et al.* [3] and Zoha *et al.* [4] provide an extensive overview of features and algorithms that were proposed before 2012. In this section, a concise and updated review is provided that incorporates the latest developments in the state-of-the-art on feature extraction.

The type of features that can be extracted from voltage (V) and current (I) measurements depends on the sampling rate [3] (e.g., step changes in power measured at the fundamental grid frequency, 50Hz in EU, 60Hz in US; or harmonics and transient based features from higher sampling rates). The features are also categorized into steady and transient state features, depending on the state of the measured waveform they represent [4]. Table 1 gives a summary of the proposed (combination of) features used for NILM in the literature. In this table, we categorize the state-of-the-art based on the combination of features employed for load discrimination into 13 categories, listed in its first column. Note that Table 1 does not compare their for following reasons: (1) the load discriminating algorithms differ among the papers, (2) the datasets used for performance evaluations are not the same, and (3) the performance measures differ. Below, we further comment on each category.

P-Q plane: step changes in real power (P) and reactive power (Q) are the first and the most commonly used steady state signatures in NILM. They can identify ON/OFF and high power appliances. However, appliances with low power consumptions are more challenging to discriminate using only P-Q features, because they exhibit overlap in the corresponding space.

P-Q plane and macroscopic transient features: this combination of features is specifically suited to identify appliances with a relatively long transient time [5], having significant spikes in their power draw, followed by slower changing variations [6][7] (e.g., heat pumps, electric loads in industrial settings). Such transients are characterized using edges and slopes [6][7] or by their power pro-

4

files [5]. Although macroscopic transient features are relatively cheaper to obtain (using low frequency measurements), they do not allow to discriminate non-linear[1] or multi-state[2] loads. Additionally, overlapping transient events could challenge the identification.

P as the sole feature: although real power measurements are cheaper to obtain than those of reactive power, appliances with similar power consumptions are harder to discriminate by relying solely on real power measurements. Hence, this feature has typically been complemented by time and frequency of the appliance usage. Powers *et al.* [8] record the time and frequency of the occurrence of the large changes in the power draws from a sparsely sampled (15 min) dataset. Farinaccio *et al.* [9] use appliance specific decision rules, whereas Marceau *et al.* [10] extend [9] and use signal filtering (smoothing) along with usage duration statistics as the complementary information. Finally, Baranski *et al.* [11][12][13] complement the real power data with a histogram of frequency of changes in power values and consider only the more frequent power changes thereof.

Low freq P-Q, I and V based features: combination of current and voltage based features (i.e., I_{rms}, I_{max}, V_{max}, V_{peak}, power factor (PF), and phase shift) with P-Q features shows good performance in identifying the ON/OFF kitchen appliances using Real Time Recognition and Profiling of Appliances (RECAP) [14]. However, it was acknowledged that multi-state appliance detection requires features beyond steady state and macroscopic features [14]. *PF* was also combined with P-Q features in [15].

P-Q plane, macroscopic transient and harmonics: proliferation of nonlinear loads in residential and commercial buildings has motivated the use of the harmonic contents of the waveforms as a discriminating feature. Sultanem [16] is

---

[1]The impedance of a nonlinear load changes with the applied voltage. The current drawn by the nonlinear load will not be sinusoidal due to the changing impedance, even when it is connected to a sinusoidal voltage, resulting in current harmonics. Examples are many electronic devices such as laptops and TVs.

[2]multi-state loads have many distinct power consumption values depending on their states, e.g., the various cycles in a washing machine's program.

the first to combine the current harmonics with low frequency based features. The research was done in parallel, but independently of the research by Hart *et al.* [2] and the subsequent extensions of his work [5]. Sultanem also proposed similar P-Q features and macroscopic transients, however the algorithm is different from Hart's work. The paper does not provide further details regarding the performance of the approach.

P-Q plane and harmonics: harmonic contents of current or power waveforms (obtained by Fast Fourier Transform (FFT) of high frequency measurements) are usually combined with P-Q features. Srinivasan *et al.* [17] use the magnitude and phase angle of the first 8 harmonics from both transient and steady state. Laughman *et al.* [18] used the harmonics of the transient signal to complement P-Q features and found that the 3rd harmonic can improve the discrimination between computers and incandescent bulbs. Berges *et al.* [19][20] use the regression coefficient of the nonlinear fit to FFT to complement the P-Q features. Dong *et al.* [21] use total harmonic distortion (THD) of current waveforms along with P and Q for load discrimination.

Spectral envelope: spectral envelopes (i.e., vectors of the first several coefficients of Short Time Fourier Transform (STFT) [22]) are natural extensions to the use of harmonics. Unlike FFT, in which the timing information is lost when transforming the signal into the frequency domain, STFT uses a fixed window to transform a small section of the signal at a time, hence, preserving the timing information by mapping the signal to a two-dimensional function of time and frequency. Therefore, spectral envelopes pave the way for identification of nonlinear and variable-load appliances [23][24]. However, the use of a fixed window size for all frequencies restricts the flexibility of STFT.

Wavelets: Compared to STFT, a wavelet transform is a more flexible approach to representing a variable signal. A wavelet transform decomposes a signal into time and scale using wavelets with adaptable scale properties. This way, one can use longer windows where more precise low-frequency information is required and shorter regions where high-frequency information is needed. Chan *et al.* [25] have taken the first step in constructing a load signature using

6

continuous wavelet transforms (CWT) of the current waveform. The load signature is constructed using a 4 level wavelet transform and uses Daubechies (DB) wavelet as the mother wavelet. Duarte *et al.* [26] compared STFT and CWT in decomposing switching voltage transients and recommended CWT as a promising approach to extract transient features in NILM. To avoid the computational complexity of CWT, a discrete wavelet transform (DWT) has been used as an alternative to extract features for NILM applications. Su *et al.* [27] used DWT of the turn-on current transient as a feature set and showed its advantages over STFT in transient analysis of the loads. Chang *et al.* [28][29] complement the P-Q features with transient energy and transient response time of power waveforms (calculated based on the DWT coefficients with DB3 wavelets) and demonstrate that the combination of features improves the accuracy and the training time of the NILM algorithm. Gray *et al.* [30] use the energy of the obtained wavelet coefficients at the detail levels as a feature set and compared the classification accuracy using features contracted by various orders of DB wavelets. They showed that higher order DB wavelets (and DB5 in particular) exhibit higher classification accuracy. Tabatabaei *et al.* [31] also calculate the energy of the wavelet coefficients in each DWT level (obtained using Haar wavelets) and use them as a feature set instead of the wavelet coefficients. Finally, Gillis *et al.* [32][33] propose a new wavelet specifically designed for NILM application rather than using DB wavelets (which are designed typically for other applications such as communications and image processing). However, the improvement achieved by the newly designed filter is found to be small compared to DB wavelets.

Although harmonic and wavelet coefficients are very effective in representing nonlinear waveforms, they require a high frequency sampling rate (at least twice the frequency of the highest harmonic intended for extraction) and complex signal processing. Hence, features based on the shape of waveforms are also considered in the literature since they can be obtained using relatively smaller sampling rates and simpler processing.

Raw waveforms: To eliminate the need for signal processing in the feature extraction phase, Suzuki *et al.* [34] have proposed to use the unprocessed cur-

7

rent waveforms for load discrimination. Although current waveforms of various nonlinear appliances differ, the feature based approaches are still shown to be more robust compared to the raw waveforms [3].

VI trajectories: in order to take into account the shapes of the waveforms in a more robust approach, Lam *et al.* [35] introduced shape features based on the two-dimensional VI trajectories. These features have been used to construct a taxonomy of household appliances. Hasan *et al.* [36] have illustrated further improvements in the load categorization by extracting additional features from VI trajectories. To make the shape based feature extraction computationally efficient, Du *et al.* [37] first map the VI trajectories to a grid of cells, each of which is assigned a binary value from which shape features are extracted. Finally, Gao *et al.* [38] use the aforementioned binary image of VI trajectory as a feature set.

Inactive current: with increasing use of nonlinear loads and the non-sinusoidal current draws thereof, more advanced power theories are considered for decomposition of the apparent power. A well-known example is the decomposition of current into active $(i_a(t))$ and nonactive current $(i_f(t))$ proposed by Fryze [39]. Huang *et al.* [40] show (using experimental results) that the similarity between nonactive current draws of various appliances (with similar power levels) are lower compared to their instantaneous current and power waveforms. However, the effectiveness of features based on nonactive current in improving the performance of NILM algorithms has not been tested using a load discriminating algorithm. Teshome *et al.* [41] leverage the dissimilarity of the nonactive current draws of various appliances and propose voltage-nonactive current VI_f trajectories as appliance feature. Their detailed analysis shows that the nonactive current waveform and the VI_f trajectories are more different than the VI trajectories of individual and combined appliances.

Unconventional features: Patel *et al.* [42] have used FFT of voltage noise to define appliance signatures and later extended their work to consider Fourier features of the electromagnetic interference signals in the 36-500 kHz range [43]. Although these features achieve high discrimination accuracy, they are sensitive

8

to the electrical household wiring. Additionally, it requires the appliances to be equipped with switch mode power supply, which is not the case for all of the household appliances. Kim *et al.* [44] combine non-electrical features such as ON/OFF duration distribution, frequency of appliance usage, and correlation between the usage of various appliances with the real power feature to improve the load differentiability. Wang *et al.* [45] represent the shape features of the real power using two unit shapes: rectangles and triangles. In a similar spirit, Koutitas *et al.* [46] replace the time series of a continuous power signal with a set of discrete pulses. Pulses are computed according to the first derivative of the flattened power signal. They further add human behavior information (i.e., time of use probability, look-for-neighbor pulses probability, duration of a pulse, sequence of operation and external conditions) to the pulses. Kong *et al.* [47] use the frequency and amplitude of the dominant peaks in the smoothed cepstrum of the voltage signal as appliance features to distinguish ON/OFF appliances. The cepstrum is defined as the inverse Fourier transform of the logarithm of the spectrum of a signal [47].

Combination of features: the simultaneous use of various features from afore-mentioned categorizes was first proposed by Liang *et al.* [48] and more recently by Lin *et al.* [49] and Gao *et al.* [38]. Although ad-hoc combination of features can improve the load discrimination capability, a systematic approach of opti-mally combining features (e.g., using the lowest possible number of features to obtain the best performance) has not been explored.

Table 1: Review and catagorization of state-of-the-art on feature extraction for NILM

Feature category	References	Low frequency			High frequency			
		Steady		Transient	Steady		Transient	
		$\Delta P \Delta Q$	I, V	macro	FFT	shape	STFT FFT	WT
P-Q plane	[2] [50]	✓						
P-Q plane & macroscopic transient	[6] [7]	✓		Shape				
P-Q plane & macroscopic transient & harmonics	[16]	✓		dura-tion	Har			
	[5]	✓		Shape	Har			
P-Q plane & harmonics	[17]	✓			Har		Har	
	[18]	✓					Har	
	[19] [20]	✓					Har	
	[21]	✓			THD			
Real power only	[8] [15] [10] [11] [12] [13]	ΔP						
Low frequency P, I & V based features	[14]	ΔP	RMS, max					
	[15]	✓PF						
Spectral envelope (SE)	[22] [23] [24]						SE	
Wavelets	[25] [26] [30] [32]							W_{coef}
	[33]							
	[31]							W_e
P-Q plane & Wavelets	[28]	✓			THD			W_{coef}, t_r, W_e
	[29]	✓						W_{coef}, W_e
	[27]	✓						W_{coef}
Shape features	[35] [36] [37]					VI–traj		
Raw waveforms	[34]					$i(t)$		
Nonactive current	[40] [41]					$i_f(t)$ VIf–traj		
Combination	[48]	✓			Har	$i(t), p(t),$ VI–traj		
	[38]	✓			Har	$i(t),$ VI–traj		
	[49]	✓	I_{cr}		Har, THD	$p(t),$ VI–traj	Har	

ΔP: step changes in real power; ΔQ: step changes in reactive power; **RMS**: root-mean-square of current and voltage; **max**: Maximum value of current and voltage waveforms; I_{cr}: current crest factor; **PF**: power factor, the ratio of real to apparent power ($\frac{P}{S}$); **Har**: coefficients of fast Fourier transform (FFT); **THD**: Total harmonic distortion; **VI–traj**: voltage–current trajectory; **VIf–traj**: voltage–nonactive current trajectory; $i(t)$: instantaneous current waveform; $i_f(t)$: instantaneous nonactive current waveform; $p(t)$: instantaneous power waveform; W_{coef}: coefficients of wavelet transform; W_e: energy of wavelet transform coefficients; t_r: transient response time based on wavelet coefficients.

225

10

3. Methodology

In this section, a systematic approach is presented to identify and combine features in such a way that the ability to distinguish various loads is maximized.

A comprehensive list of steady and transient state features, including references to their extraction steps is outlined in Table 2 and the Appendix.

The feature selection process starts from all features and iteratively eliminates the least important ones until a reduced subset is obtained. Note that the goal is not to reduce number of features to a set amount but rather, to choose a subset of features in such a way that the discriminative performance is either better or compared to using of all of them. In other words, the iterative feature selection terminates once further feature elimination yields no improvements or reduces the appliance identification accuracy.

3.1. Feature Selection Algorithm

Feature selection is an essential step in machine learning in which a subset of relevant features or variables is identified and selected to be used in model construction. In contrast to other dimensionality reduction techniques, feature selection does not change the original representation of features and therefore allows further interpretation by a domain expert after model construction. Feature selection methods are traditionally categorized in three types: filter methods, wrapper methods and embedded methods. Filter methods perform feature selection by only looking at the intrinsic properties of the data, most often by calculating a feature relevance score and removing features with low scores. Filter methods are computationally fast but are usually less accurate because the feature selection process is decoupled from the computational model. In wrapper methods, feature subsets are evaluated by the accuracy that a specific model (classifier or regression model) obtains by using them. Building a model for each possible feature subset is often computationally intangible and therefore a heuristic search procedure is wrapped around the model building. Embedded methods extract feature importance knowledge, obtained automatically by

11

Table 2: Notation and description of extracted features

State	Notation	Description
steady	P	real power
	Pnorm	normalized real power [2]
	Q	reactive power
	Qf	reactive power based on Fryze's formula [51]
	S	apparent power
	I-rms	current root mean square
	If-rms	nonactive current root mean square
	I-har(j)	j^{th} current harmonic coefficient
	If-har(j)	j^{th} nonactive current harmonic coefficient
	V-har(j)	j^{th} voltage harmonic coefficient
	I-THD	total harmonic distortion of current [21]
	If-THD	total harmonic distortion of nonactive current
	V-THD	total harmonic distortion of voltage
	Asymmetry	measure of asymmetry in VI trajectory [35]
	Intersections	number of intersections in VI trajectory [35]
	Area	enclosed area by VI trajectory with consideration of trajectory direction [35]
	Net area	net area enclosed by VI trajectory without consideration of its direction
	Curvature	measure of distortion of mean line of VI trajectory from a straight line [35]
	Slope	slope of the middle segment of VI trajectory [35]
transient	Wd(i)	energy of detail wavelet coefficients at i^{th} scale [31][30] (see Appendix)
	Wa	energy of approximate wavelet coefficients [31][30] (see Appendix)
	W-max-idx	index of the maximum energy wavelet coefficient [27]
	W-max	maximum value of the wavelet coefficient [27]
	I-max-tr	maximum value of the transient current
	I-max-tr-idx	location of maximum transient current (index)
	I-min-tr	minimum value of the transient current
	I-min-tr-idx	location of minimum transient current (index)
	diff-I-tr	difference between maximum and minimum values of transient current
	P-max-tr	maximum value of the transient power
	P-max-tr-idx	location of maximum transient power (index)
	P-min-tr	minimum value of the transient power
	P-min-tr-idx	location of minimum transient power (index)
	I-peak-num-tr	no. of local maximums of transient current

training a model. These last type of methods have the advantage that they are coupled to the model being used, but do not require as much computational effort as wrapper methods.

For the remainder of this paper, a Random Forest model is used as the main classification model. It was found that other model types gave a similar or worse performance, which is consistent with the findings of Gao *et al.* [38].

In the first step of this feature selection process, a well-known method, called Recursive Feature Elimination (RFE)[52], is explored. RFE starts by constructing a model using all features. In a second step, a ranking of the features according to their importance is extracted from the model. The least x important variables are then removed from the model, where x is user-defined. This process is then repeated until a stopping criterion is reached, e.g. an increase in validation error or a lower bound on the amount of retained features.

Results obtained using the RFE procedure (Section 4.2) were not satisfactory due to high correlation among features. Therefore a second approach is explored. More concretely, a process is started that iteratively trains the random forest algorithm using the retained features from step 1 and calculates the *permutation importance* of each feature. This importance measure can be calculated for each feature in turn by removing the association between that feature and the target in the validation phase. This is achieved by randomly permuting the values of the features several times and by measuring the average increase in the relative error of the model. Often this measure provides better results in assessing the feature importance (than e.g., the Gini Importance [53]) in the case where the included features are highly correlated. Intuitively, features with lesser influence in the load discrimination accuracy of the model will cause negligible changes in the model performance when permuted. This procedure is repeated several times and in each iteration the least important features are removed.

13

Table 3: List of appliances and number of retained records per appliance from PLAID dataset

Appliance	No. of Records
Hairdryer	137
Microwave	136
Compact Fluorescent Lamp (CFL)	125
Fan	112
Incandescent Light Bulb	111
Air Conditioner (AC)	63
Vacuum	38
Heater	34
Fridge	26
Washing Machine	23
Laptop	16

4. Results and Discussion

In this section, we demonstrate our systematic feature selection process and present the per appliance based analysis using a comprehensive dataset explained next.

4.1. Dataset Description

We base our analysis on the PLAID dataset [54] consisting of 1094 current and voltage records from 11 different appliance types collected at 30 kHz present in 56 households. Since we consider both transient and steady state features per appliance, we remove the records for which transient state of the device is not measured and hence, we are left with 821 records. Table 3 summarizes the list of appliances and the number of retained records per appliance. These records are then used to calculate 55 steady and 23 transient state features as listed in Table 2. Evaluations on this dataset commonly employ a leave-on-out cross validation scheme using a single house as atomic unit as opposed to a single trace [38]. The same procedure is also adopted in this paper.

4.2. Demonstration of Feature Selection Algorithm

In this section, we demonstrate, using PLAID dataset, the process of systematic feature selection.

Figure 1 shows the performance of the random forest algorithm in terms of cross validation score (percentage of correct classifications) vs. the number of
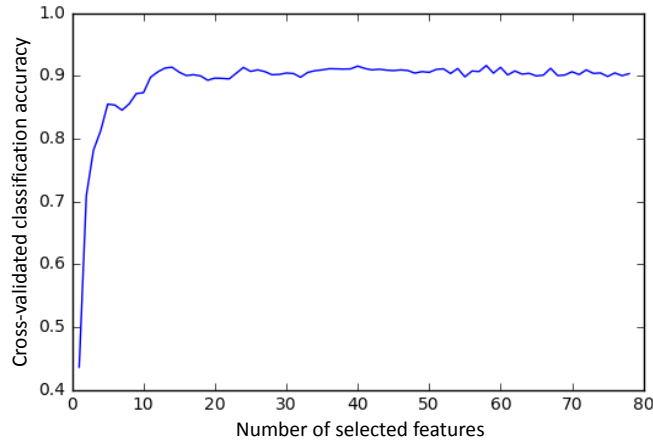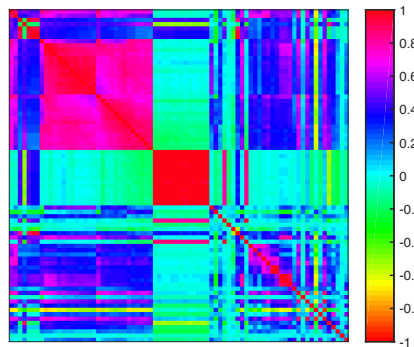
Fig. 1: Recursive feature elimination

Fig. 2: Heatmap indicating the correlation coefficient values among various features for PLAID dataset

selected features using the RFE scheme. The accuracy reaches a plateau quickly at around 20 features and only marginally increases in accuracy. This result seems to indicate that RFE using the Gini Importance is not a good approach to select an optimal subset, caused by the presence of highly correlated features in the dataset as depicted in Fig. 2. Due to the inherent randomness of the algorithm and small differences in accuracy measured in the plateau region, the

15
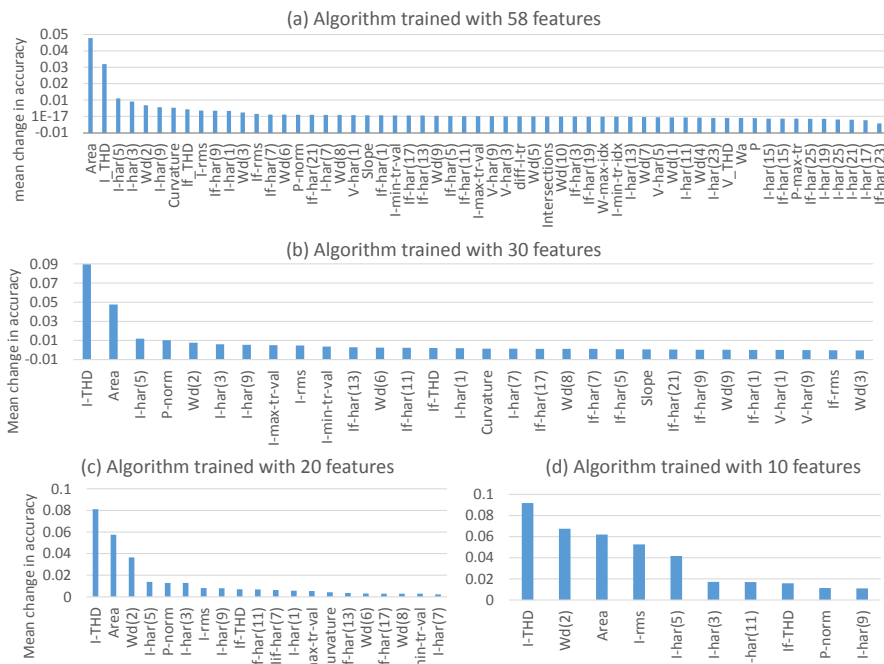
Fig. 3: Mean change in accuracy when permuting a feature at various iterations in step 2 of the feature elimination process

choice of threshold is not clear-cut. As such we opt for a conservative approach by only removing the last 20 features, which correspond to the model that reached the highest accuracy in repeated runs. Amongst the eliminated features are the higher order voltage harmonics, reactive and apparent powers from steady state features and locational features (i.e., indexes of max/min current and power) from the transient state features. All of the current harmonics, nonactive current harmonics and wavelet coefficients energies are retained in this step (see horizontal axis labels in Fig. 3(a) for the list of 58 retained features).

To investigate the possibility of a further reduction in the number of features, we use the 58 remaining features from step one and retrain the random forest algorithm. Once trained, we randomly permute each feature at a time during validation, and take the average accuracy (over 50 random permutation per feature). The resulting changes in accuracy for each feature permutation are

reported in Fig. 3(a). As seen from Fig. 3(a), less important features will have a negligible change in the accuracy when permuted. However, for the least importance features, no clear threshold appears as many features have approximately the same measure. Again, using a conservative approach we reduce the amount of features from 58 to 30. We repeat the process of feature importance ranking using random permutation to further reduce them to 20 and 10. The importance rankings using random permutation are depicted in Fig. 3(b), (c) and (d) for 30, 20 and 10 retained features respectively. Also note that as the number of the retained features decreases in each iteration, the mean change in accuracy due to a feature permutation in a subsequent iteration increases compared to the previous one. This indicates the effectiveness of the proposed scheme in retaining the influential features. Finally, the mean change in accuracy when the algorithm is trained with 10 features is significantly worse than the previous iteration indicating that all the retained features are influential and should not be disregarded. As for the list of retained features (indicated in the horizontal axis of plots in Fig. 3), all of the voltage harmonics are eliminated after two rounds of feature selection (from 58 to 20 features). This is because voltage variations in the grid are minimal with connection/disconnection of loads while the connecting loads affect the current draws. All of the power based features from transient state are also eliminated after two rounds. It is evident from Fig. 3 that wavelet coefficient energies and values of maximum and minimum currents are more influential than the other features. Finally, I-THD and Area from steady state and Wd(2) from transient state are ranked as the most important features in all of the feature elimination rounds.

The systematic reduction of features results in a significant increase in the classification accuracy as opposed to the model which uses all features. The classification accuracy is maximal when the combination of 20 features are used as appliance signature as listed in Fig. 3(c). However, the accuracy decreases from its peak when more features are eliminated. Additionally, systematic combination of various feature categories results in higher accuracy compared to using each category of features individually.
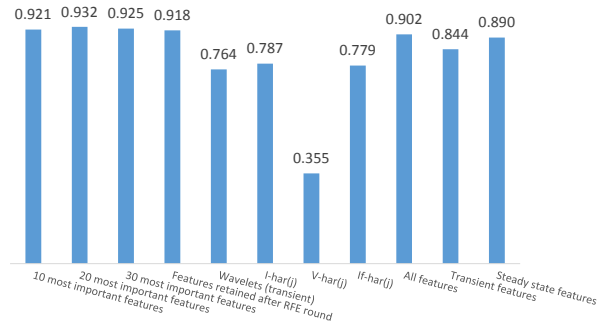
Fig. 4: Average (over all the houses) accuracy of the random forest algorithm with various catagories of features as input

We also test the stability of the algorithm by running the same algorithm on the same dataset multiple times. This is to ensure the deviation caused by the inherent randomness of the algorithm is kept within acceptable bounds. The standard deviation of the mean and the median are 0.0007 and 0 accordingly which ensures the algorithm's stability.

4.3. Per Appliance Analysis

In this section, we analyze the algorithm's performance from per-appliance point of view. Figure 5 indicates the F-score for each of the appliances in PLAID dataset with retained features from each iteration of the feature elimination process. The F-scores in Fig. 5 indicate that the algorithm's performance globally improves (i.e., improvement is across all appliances) after each round of feature elimination from its initial F-score (i.e., with all the features included as input). Appliances with high F-scores have less improvement due to systematic feature selection than the ones with smaller F-score. Since the small values of F-scores are typically associated with small number of records, we conclude that feature selection is more influential in algorithm performance when limited number of measurements are available.

To have further insights on which appliances are misclassified, we show the confusion matrices in different iterations of the feature selection process in Fig. 6. with numbers in each cell of the matrices (except diagonal elements) indicating
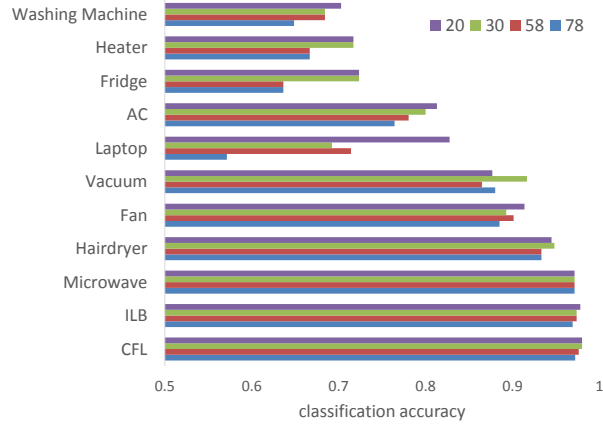
18

Fig. 5: F-score of the appliances when using combination of retained features in each iteration of proposed feature elimination process (numbers in legend correspond to number of retained features in each iteration and 78 corresponds to all features)
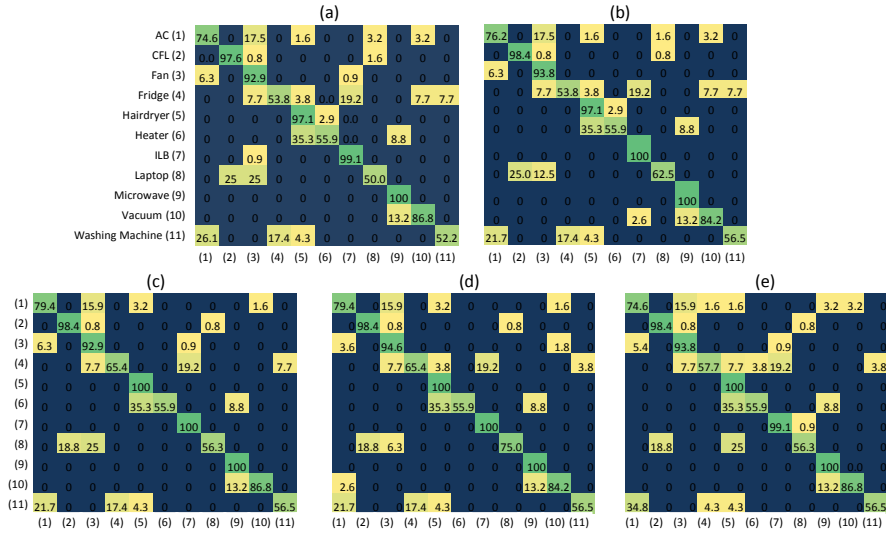

Fig. 6: Confusion matrix at each iteration of the proposed feature elimination process (a) All features, (b) 58, (c) 30, (d) 20, and (e) 10 retained features

the misclassification percentages. As observed from Fig. 6, the feature elimination process typically improves the misclassification rates and in majority of the

appliances, does not create new misclassification. Amongst all the appliances, 'heater' has the highest misclassification rate (misclassified as 'hairdryer' 35.3% of the time) and it is the only appliance for which misclassification does not improve with the feature elimination process. This is mainly due to the similarities in the power levels and the electrical components in the circuitry of these appliances. However, 'hairdryer' is classified correctly 100% in the improved algorithm. This is partially due to the higher number of hairdryer records in the database which makes the algorithm more biased to label the data as 'hairdryer' than 'heater'.

5. Conclusion

The effectiveness of a NILM algorithm to distinguish between appliances largely depends on determining a set of discriminative features. Various research has focused on suggesting and extracting such features to classify appliances. As a first contribution, we provided a systematic listing and comparison between features that have been proposed.

As a second contribution, we constructed and suggested an optimal subset of features to be used in a computational model which can achieve top performance in classifying appliances. It is well-known that the inclusion of irrelevant or redundant features affects on the accuracy of a computational model. Adding extra features that provide little or no extra value increases the dimensionality and such the complexity of the model. In addition, spurious correlations between the feature and the class labels introduces extra noise and increases the risk of overfitting. Therefore, to achieve top performance, we presented a systematic feature elimination process and have shown that by removing irrelevant features we were able to substantially boost model accuracy. Our approach consisted of iteratively removing sets of features that had low feature importance scores. These scores were determined by how much the performance of the model changed when the features values are randomly shuffled (permutation feature importance). Other approaches using RFE in combination with

20

the Gini importance measure, proved unsuccessful due to the large amount of highly correlated features in the data. Lastly, we have also shown that not only the overall model performance increased but that our feature selection process also globally improves F-scores across all appliances.

6. Future Work

The majority of the proposed NILM solutions are tested on private datasets with a limited number of appliances. Additionally, the existing literature categorizes appliances into three categories based on their operational characteristics: ON/OFF appliances, multiple state appliances, and variable load appliances. However, when it comes to identifying a unique signature for each appliance, such categorization is not very effective because it ignores the front-end circuit topology of the appliances which affects their current draws. Since features are extracted from current and voltage measurements, it is more practical to categorize loads considering their electrical operations and their front-end circuitry that connects them to the power grid as elaborated by He *et al.* [55]. They categorize the appliances into seven classes: resistive loads, reactive predominant loads, electronic loads with a power factor correction circuit, electronic loads without a power factor correction circuit, linear power supply using transformer to boost voltage, phase angle controllable loads, and complex structures. They further demonstrate that optimized features can be obtained from such categorization, which drives a much simpler and more feasible solution in differentiating the subtle differences between similar loads. However, their analyses are done on small dataset which is not publicly available.

The currently available public datasets do not include appliances from the aforementioned seven categories and hence, lack the required diversity. Having such diverse, comprehensive and publicly available NILM datasets paves the way for comparable, practical and generalizable NILM solutions. As our next step, we will create a comprehensive and diverse dataset (having various categories of appliances across different households) to perform our feature selection process
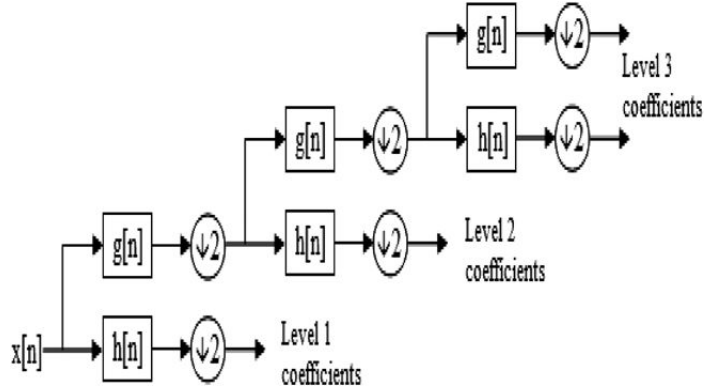
21

Fig. 7: Three decomposed DWT levels [29]

⁴³⁵ and identify the correlation of various appliance categories with such features.

Appendix

In this appendix, we explain how to obtain two of the listed features in Table 2: energy of wavelet coefficients in i^{th} level and nonactive current based on Fryze's formula.

⁴⁴⁰ *Energy of wavelet coefficients*: DWT is a multi-resolution analysis where a complex signal is decomposed into a set of approximate and detail coefficients using sets of high pass and low pass filters as shown in Fig. 7. Detail coefficients are obtained by applying a high pass filter (i.e., wavelet) to the sampled signal. They express the higher frequency components in the signal. Approximate ⁴⁴⁵ coefficients are obtained at the output of the last low pass filter. They express the signal at a lower resolution. Sum of the square of the coefficients in each level is the energy of the wavelet coefficients.

22

Nonactive current: According to Fryze [39], current waveform is decomposed into active and nonactive components:

$$i(t) = i_a(t) + i_f(t) \tag{1}$$

$i_a(t)$ (active current) is a current of the same wave-shape and phase angle as the voltage. It is responsible for the transference of average energy to the load and is obtained by:

$$i_a(t) = \frac{P}{V_{rms}^2} v(t) \tag{2}$$

where P is the active power and V_{rms} is the voltage root mean square.

$i_f(t)$ (nonactive current) is the orthogonal to the voltage (and hence to the active current) and is responsible for oscillation and disturbances in the active power without transferring any energy to the load. Due to orthogonality of active and nonactive currents:

$$I_{rms} = I_{a_{rms}} - I_{f_{rms}} \tag{3}$$

Where I_{rms} is the current root-mean-square and $I_{a_{rms}}$ and $I_{f_{rms}}$ are root mean square of the active and nonactive currents respectively.

References

[1] K. C. Armel, A. Gupta, G. Shrimali, A. Albert, Is disaggregation the holy grail of energy efficiency? the case of electricity, Energy Policy 52 (2013) 213 – 234, special Section: Transition Pathways to a Low Carbon Economy. `doi:10.1016/j.enpol.2012.08.062`.

[2] G. W. Hart, Nonintrusive appliance load monitoring, Proceedings of the IEEE 80 (12) (1992) 1870–1891. `doi:10.1109/5.192069`.

[3] M. Zeifman, K. Roth, Nonintrusive appliance load monitoring: Review and outlook, IEEE Transactions on Consumer Electronics 57 (1) (2011) 76–84. `doi:10.1109/TCE.2011.5735484`.

[4] A. Zoha, A. Gluhak, M. A. Imran, S. Rajasegarar, Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey, Sensors 12 (12) (2012) 16838–16866. `doi:10.3390/s121216838`.

[5] L. K. Norford, S. B. Leeb, Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms, Energy and Buildings 24 (1) (1996) 51 – 64. `doi:10.1016/0378-7788(95)00958-2`.

[6] A. I. Cole, A. Albicki, Data extraction for effective non-intrusive identification of residential power loads, in: IMTC/98 Conference Proceedings. IEEE Instrumentation and Measurement Technology Conference. Where Instrumentation is Going (Cat. No.98CH36222), Vol. 2, 1998, pp. 812–815 vol.2. `doi:10.1109/IMTC.1998.676838`.

[7] A. I. Cole, A. Albicki, Algorithm for nonintrusive identification of residential appliances, in: Circuits and Systems, 1998. ISCAS '98. Proceedings of the 1998 IEEE International Symposium on, Vol. 3, 1998, pp. 338–341 vol.3. `doi:10.1109/ISCAS.1998.704019`.

[8] J. T. Powers, B. Margossian, B. A. Smith, Using a rule-based algorithm to disaggregate end-use load profiles from premise-level data, IEEE Computer Applications in Power 4 (2) (1991) 42–47. `doi:10.1109/67.75875`.

[9] L. Farinaccio, R. Zmeureanu, Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses, Energy and Buildings 30 (3) (1999) 245 – 259. `doi:10.1016/S0378-7788(99)00007-9`.

[10] M. Marceau, R. Zmeureanu, Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings, Energy Conversion and Management 41 (13) (2000) 1389 – 1403. `doi:10.1016/S0196-8904(99)00173-9`.

24

[11] M. Baranski, V. J, Nonintrusive appliance load monitoring based on an optical sensor, in: 2003 IEEE Bologna Power Tech Conference Proceedings,, Vol. 4, 2003, pp. 8 pp. Vol.4–. `doi:10.1109/PTC.2003.1304732`.

[12] M. Baranski, J. Voss, Detecting patterns of appliances from total load data using a dynamic programming approach, in: Data Mining, 2004. ICDM '04. Fourth IEEE International Conference on, 2004, pp. 327–330. `doi: 10.1109/ICDM.2004.10003`.

[13] M. Baranski, J. Voss, Genetic algorithm for pattern detection in nialm systems, in: Systems, man and cybernetics, 2004 ieee international conference on, Vol. 4, IEEE, 2004, pp. 3462–3468.

[14] A. G. Ruzzelli, C. Nicolas, A. Schoofs, G. M. P. O'Hare, Real-time recognition and profiling of appliances through a single electricity sensor, in: 2010 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2010, pp. 1–9. `doi:10.1109/SECON.2010.5508244`.

[15] M. B. Figueiredo, A. de Almeida, B. Ribeiro, An Experimental Study on Electrical Signature Identification of Non-Intrusive Load Monitoring (NILM) Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 31–40. `doi:10.1007/978-3-642-20267-4_4`.

[16] F. Sultanem, Using appliance signatures for monitoring residential loads at meter panel level, IEEE Transactions on Power Delivery 6 (4) (1991) 1380–1385. `doi:10.1109/61.97667`.

[17] D. Srinivasan, W. S. Ng, A. C. Liew, Neural-network-based signature recognition for harmonic source identification, IEEE Transactions on Power Delivery 21 (1) (2006) 398–405. `doi:10.1109/TPWRD.2005.852370`.

[18] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, P. Armstrong, Power signature analysis, IEEE Power and Energy Magazine 1 (2) (2003) 56–63. `doi:10.1109/MPAE.2003.1192027`.

25

[19] M. Berges, E. Goldman, H. S. Matthews, L. Soibelman, Learning systems for electric consumption of buildings, in: ASCE International Workshop on Computing in Civil Engineering, ASCE, Austin, Texas, USA, 2009.

[20] M. E. Berges, E. Goldman, H. S. Matthews, L. Soibelman, Enhancing electricity audits in residential buildings with nonintrusive load monitoring, Journal of Industrial Ecology 14 (5) (2010) 844–858. `doi:10.1111/j.1530-9290.2010.00280.x.`

[21] M. Dong, P. C. M. Meira, W. Xu, C. Y. Chung, Non-intrusive signature extraction for major residential loads, IEEE Transactions on Smart Grid 4 (3) (2013) 1421–1430. `doi:10.1109/TSG.2013.2245926.`

[22] S. B. Leeb, S. R. Shaw, J. L. Kirtley, Transient event detection in spectral envelope estimates for nonintrusive load monitoring, IEEE Transactions on Power Delivery 10 (3) (1995) 1200–1210. `doi:10.1109/61.400897.`

[23] S. R. Shaw, S. B. Leeb, L. K. Norford, R. W. Cox, Nonintrusive load monitoring and diagnostics in power systems, IEEE Transactions on Instrumentation and Measurement 57 (7) (2008) 1445–1454. `doi:10.1109/TIM.2008.917179.`

[24] K. D. Lee, S. B. Leeb, L. K. Norford, P. R. Armstrong, J. Holloway, S. R. Shaw, Estimation of variable-speed-drive power consumption from harmonic content, IEEE Transactions on Energy Conversion 20 (3) (2005) 566–574. `doi:10.1109/TEC.2005.852963.`

[25] W. L. Chan, A. T. P. So, L. L. Lai, Harmonics load signature recognition by wavelets transforms, in: DRPT2000. International Conference on Electric Utility Deregulation and Restructuring and Power Technologies. Proceedings (Cat. No.00EX382), 2000, pp. 666–671. `doi:10.1109/DRPT.2000.855745.`

[26] C. Duarte, P. Delmar, K. W. Goossen, K. Barner, E. Gomez-Luna, Non-intrusive load monitoring based on switching voltage transients and wavelet

transforms, in: 2012 Future of Instrumentation International Workshop (FIIW) Proceedings, 2012, pp. 1–4. `doi:10.1109/FIIW.2012.6378333`.

[27] Y. C. Su, K. L. Lian, H. H. Chang, Feature selection of non-intrusive load monitoring system using stft and wavelet transform, in: 2011 IEEE 8th International Conference on e-Business Engineering, 2011, pp. 293–298. `doi:10.1109/ICEBE.2011.49`.

[28] H.-H. Chang, Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses, Energies 5 (11) (2012) 4569–4589. `doi:10.3390/en5114569`.

[29] H. H. Chang, K. L. Chen, Y. P. Tsai, W. J. Lee, A new measurement method for power signatures of nonintrusive demand monitoring and load identification, IEEE Transactions on Industry Applications 48 (2) (2012) 764–771. `doi:10.1109/TIA.2011.2180497`.

[30] M. Gray, W. G. Morsi, Application of wavelet-based classification in non-intrusive load monitoring, in: 2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE), 2015, pp. 41–45. `doi:10.1109/CCECE.2015.7129157`.

[31] S. M. Tabatabaei, S. Dick, W. Xu, Toward non-intrusive load monitoring via multi-label classification, IEEE Transactions on Smart Grid 8 (1) (2017) 26–40. `doi:10.1109/TSG.2016.2584581`.

[32] J. M. Gillis, S. M. Alshareef, W. G. Morsi, Nonintrusive load monitoring using wavelet design and machine learning, IEEE Transactions on Smart Grid 7 (1) (2016) 320–328. `doi:10.1109/TSG.2015.2428706`.

[33] J. M. Gillis, W. G. Morsi, Non-intrusive load monitoring using semi-supervised machine learning and wavelet design, IEEE Transactions on Smart Grid PP (99) (2016) 1–8. `doi:10.1109/TSG.2016.2532885`.

[34] K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura, K. Ito, Nonintrusive appliance load monitoring based on integer programming, in: 2008 SICE Annual Conference, 2008, pp. 2742–2747. `doi:10.1109/SICE.2008.4655131`.

[35] H. Y. Lam, G. S. K. Fung, W. K. Lee, A novel method to construct taxonomy electrical appliances based on load signatures, IEEE Transactions on Consumer Electronics 53 (2) (2007) 653–660. `doi:10.1109/TCE.2007.381742`.

[36] T. Hassan, F. Javed, N. Arshad, An empirical investigation of v-i trajectory based load signatures for non-intrusive load monitoring, IEEE Transactions on Smart Grid 5 (2) (2014) 870–878. `doi:10.1109/TSG.2013.2271282`.

[37] L. Du, D. He, R. G. Harley, T. G. Habetler, Electric load classification by binary voltage current trajectory mapping, IEEE Transactions on Smart Grid 7 (1) (2016) 358–365. `doi:10.1109/TSG.2015.2442225`.

[38] J. Gao, E. C. Kara, S. Giri, M. Bergs, A feasibility study of automated plug-load identification from high-frequency measurements, in: 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2015, pp. 220–224. `doi:10.1109/GlobalSIP.2015.7418189`.

[39] S. Fryze, Wirk- blind und scheinleistung in elektrischen stromkreisen mit nichtsinusoidalem verlauf von strom und spannung, Elektrotech. Z. 53 (25) (1932) 596–599.

[40] T. D. Huang, W. S. Wang, K. L. Lian, A new power signature for nonintrusive appliance load monitoring, IEEE Transactions on Smart Grid 6 (4) (2015) 1994–1995. `doi:10.1109/TSG.2015.2415456`.

[41] D. F. Teshome, T. D. Huang, K. L. Lian, Distinctive load feature extraction based on fryzes time-domain power theory, IEEE Power and Energy Technology Systems Journal 3 (2) (2016) 60–70. `doi:10.1109/JPETS.2016.2559507`.

28

[42] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds, G. D. Abowd, At the flick of a switch: Detecting and classifying unique electrical events on the residential power line, in: Proceedings of the 9th International Conference on Ubiquitous Computing, UbiComp '07, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 271–288.

[43] S. Gupta, M. S. Reynolds, S. N. Patel, Electrisense: single-point sensing using emi for electrical event detection and classification in the home, in: Proceedings of the 12th ACM international conference on Ubiquitous computing, ACM, 2010, pp. 139–148.

[44] H. Kim, M. Marwah, M. Arlitt, G. Lyon, J. Han, Unsupervised disaggregation of low frequency power measurements, in: Proceedings of the 2011 SIAM International Conference on Data Mining, 2011, pp. 747–758. `doi:10.1137/1.9781611972818.64`.

[45] Z. Wang, G. Zheng, Residential appliances identification and monitoring by a nonintrusive method, IEEE Transactions on Smart Grid 3 (1) (2012) 80–92. `doi:10.1109/TSG.2011.2163950`.

[46] G. C. Koutitas, L. Tassiulas, Low cost disaggregation of smart meter sensor data, IEEE Sensors Journal 16 (6) (2016) 1665–1673. `doi:10.1109/JSEN.2015.2501422`.

[47] S. Kong, Y. Kim, R. Ko, S. K. Joo, Home appliance load disaggregation using cepstrum-smoothing-based method, IEEE Transactions on Consumer Electronics 61 (1) (2015) 24–30. `doi:10.1109/TCE.2015.7064107`.

[48] J. Liang, S. K. Ng, G. Kendall, J. W. Cheng, Load signature studypart i: Basic concept, structure, and methodology, IEEE transactions on power Delivery 25 (2) (2010) 551–560.

[49] S. Lin, L. Zhao, F. Li, Q. Liu, D. Li, Y. Fu, A nonintrusive load identification method for residential applications based on quadratic program-

29

₆₃₀   ming, Electric Power Systems Research 133 (2016) 241 – 248. `doi:`
`10.1016/j.epsr.2015.12.014`.

[50] S. Drenker, A. Kader, Nonintrusive monitoring of electric loads, IEEE Computer Applications in Power 12 (4) (1999) 47–51. `doi:10.1109/67.795138`.

[51] S. Svensson, Power Measurement Techniques for Nonsinusoidal Conditions:
₆₃₅   The Significance of Harmonics for the Measurement of Power and Other AC Quantities, Chalmers tekniska högsk., 1999.

[52] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Machine Learning 46 (1) (2002) 389–422. `doi:10.1023/A:1012487302797`.

₆₄₀ [53] G. Louppe, Understanding random forests: From theory to practice, PhD thesis, University of Liege.

[54] J. Gao, S. Giri, E. C. Kara, M. Bergés, Plaid: A public dataset of high-resolution electrical appliance measurements for load identification research: Demo abstract, in: Proceedings of the 1st ACM Conference on Em-
₆₄₅   bedded Systems for Energy-Efficient Buildings, BuildSys '14, ACM, New York, NY, USA, 2014, pp. 198–199. `doi:10.1145/2674061.2675032`.

[55] D. He, L. Du, Y. Yang, R. Harley, T. Habetler, Front-end electronic circuit topology analysis for model-driven classification and monitoring of appliance loads in smart buildings, IEEE Transactions on Smart Grid 3 (4)
₆₅₀   (2012) 2286–2293. `doi:10.1109/TSG.2012.2219327`.