A Comprehensive Feature Study for Appliance Recognition on High Frequency Energy Data

Matthias Kahl
matthias.kahl@in.tum.de
Chair for Application and Middleware Systems
Technische Universität München
Germany

Anwar Ul Haq
haqa@in.tum.de
Chair for Application and Middleware Systems
Technische Universität München
Germany

Thomas Kriechbaumer
thomas.kriechbaumer@in.tum.de
Chair for Application and Middleware Systems
Technische Universität München
Germany

Hans-Arno Jacobsen
jacobsen@in.tum.de
Chair for Application and Middleware Systems
Technische Universität München
Germany

ABSTRACT

Awareness about the energy consumption of appliances can help to save energy in households. Non-intrusive Load Monitoring (NILM) is a feasible approach to provide consumption feedback at appliance level. In this paper, we evaluate a broad set of features for electrical appliance recognition, extracted from high frequency start-up events. These evaluations were applied on several existing high frequency energy datasets. To examine clean signatures, we ran all experiments on two datasets that are based on isolated appliance events; more realistic results were retrieved from two real household datasets. Our feature set consists of 36 signatures from related work including novel approaches, and from other research fields. The results of this work include a stand-alone feature ranking, promising feature combinations for appliance recognition in general and appliance-wise performances.

CCS CONCEPTS

•**Computing methodologies** → *Machine learning;* •**Information systems** → *Clustering and classification;*

KEYWORDS

NILM; NIALM; Load Information Retrieval; Appliance Recognition; High Frequency; Feature Study

ACM Reference format:
Matthias Kahl, Anwar Ul Haq, Thomas Kriechbaumer, and Hans-Arno Jacobsen. 2017. A Comprehensive Feature Study for Appliance Recognition on High Frequency Energy Data. In *Proceedings of e-Energy '17, Shatin, Hong Kong, May 16-19, 2017,* 11 pages.
DOI: http://dx.doi.org/10.1145/3077839.3077845

e-Energy '17, Shatin, Hong Kong

DOI: http://dx.doi.org/10.1145/3077839.3077845

1 INTRODUCTION

Achieving energy consumption reduction in household and industrial areas, non-intrusive load monitoring (NILM) has become an active area of research in the last two decades. It stands beside intrusive load monitoring that considers measurements on appliance level. Each concept has its pros and cons and its feasibility needs to be reappraised for the considered use case. The NILM research field provides a broad set of techniques to obtain detailed information about energy consumption and load. Considerable savings can be achieved by sharing fine-grained consumption information with consumers, especially at appliance level [8].

NILM techniques provide cost-effective approaches by using a single sensor to monitor the appliance level energy consumption of a building. They are able to extract consumption information on appliance level from aggregated loads by utilizing machine learning and pattern recognition approaches to learn individual appliance load signatures. Each appliance forms its own unique current signature and can be seen as an individual fingerprint. The primary task of the machine learning algorithms is to learn numerous appliance signatures to distinguish and identify a captured signature on demand. In the machine learning context, these signatures are called features.

Since its inception in the 1980's, NILM has evolved considerably to develop new techniques, metrics and algorithms for extracting information out of the appliance signatures. Although extensive literature on different disaggregation approaches is available, a comparison of these approaches and results is difficult. To the best of our knowledge, no comprehensive overview study about features and their combinations in terms of their contribution to machine learning approaches in NILM has been published.

One challenge in power disaggregation is finding an efficient set of features for distinguishing appliances with respect to the environment they are used in. Since computing power correlates to energy consumption, a high energy spending disaggregation system would put the energy saving idea ad absurdum. Each set of appliances and their environment (e.g., household, factory, etc.) in which they are used in demand a different set of features to receive optimal disaggregation results. One possibility would be to equip a disaggregation system modularly with the most relevant

features, regarding the appliance set and the environment. This work presents a wide set of features and shows which appliance characteristics can be covered with each individual feature. We evaluated all features and several of their combinations on four common, publicly available high frequency energy datasets to get a priori results on isolated and aggregated household environments. This will be an important step to realize such an efficient disaggregation system that eventually helps consumer in saving energy. Many countries are restricted in terms of privacy to protect home owners from being tracked and monitored by energy providers. Systems that allow a high frequency monitoring and a detailed consumption analysis are targeted directly at the end user.

To streamline the appliance recognition process, we composed an extensive list of 36 different appliance features derived from the current and voltage measurements of appliance start-up events. Since there is no ideal feature combination to disaggregate all appliances, we conducted extensive experiments to identify promising general feature combinations that perform best on all considered appliance types. The best combinations are identified using a forward selection and a 5-fold cross validation. In this work, we also derived new features based on specific appliance characteristics that may increase the appliance recognition performance.

The rest of the paper is organized as follows: In Section 2, we talk about recent work, Section 3 introduces all considered features, Section 4 describes the experimental setup and Section 5 the results. Our conclusions and a brief interpretation of study results are presented in Section 6.

2 RELATED WORK

Machine learning and pattern recognition is an efficient and widely used approach for many NILM research questions, especially for appliance identification. The appliance identification problem is often tackled with a classification task, using supervised training on existing, with ground truth-labeled appliance measurements. These types of machine learning approaches were already used in 1994 by Roos et al. [29] with neural networks and are still commonly used in more recent papers [3, 5, 19, 25].

Since appliances have individual characteristics, electrical power quantities including *Active Power*, *Reactive Power*, and *Apparent Power* may not be sufficient for all kinds of appliances. In 1992, NILM pioneer George Hart [14] described further signatures including harmonics and transient features for different appliance types. Several signal processing metrics of the temporal and spectral domain have also been applied to NILM over the years.

Further approaches, like spectral analysis using a wavelet transform, were implemented in 1995 by Leeb et al. [21] to build a prototype detector that "performs remarkably well" and was able to identify four appliance types. The *V-I Trajectory* was first used in 2005 by Ting et al. [33] and in 2007 by Lam et al. [20] for appliance classification and taxonomy purposes. Waveform-based and general signal processing metrics, including *Total Harmonic Distortion* and *Crest Factor*, which "provide a tremendously improved recognition capacity" were used in 2007 by Yang et al. [34]. In 2011, Lin et al. [23] used the *Crest Factor* feature among others to reach identification rates of "higher than 93%" for three different appliances. Gupta et al. [13] used electromagnetic interferences in the MHz range to identify appliances with an accuracy of around 94% for 7 to 20 appliances.

The state of the art in NILM comprises numerous approaches, features, experiments and results. A comparison between these approaches is either difficult, due to differences in datasets, data acquisition equipment, appliance models and other factors. The authors Armel et al. [2] review numerous disaggregation algorithms and requirements for smart meters. Their feature related comparison focuses on different sampling frequencies rather than on the individual feature performances.

The main goal of the energy disaggregation framework NILMTK from Batra et al. [4] is to allow a consistent comparison of different disaggregation strategies. As of now, NILMTK unfortunately appears to support only low frequency disaggregation. Due to this limitation, NILMTK cannot be used to evaluate approaches for high frequency energy data.

The work of Froehlich et al. [10] compares some high and low frequency features for disaggregation based on several criteria including installation, costs, sensing technology and ease of calibration. The evaluation criteria focus on the usefulness of the features in terms of environmental situations rather than performance in appliance recognition.

Gao et al. [12] evaluate several features for PLAID on 5 classifiers. The best results were achieved with a random forests classifier using the VI image feature. The authors state that the combination of features improves the classification performance, which motivates to take a deeper look into feature combinations. Since the authors compare only the most common features, several strong features such as *Wavelet Analysis* are not evaluated.

With the increasing amount of NILM approaches, the necessity for comparability of these works is also growing. The aim of this paper is to give a wide overview of established and novel appliance features, including their stand-alone and combined classification performances for appliance recognition. This work orients on the related work in terms of identification system architecture, typical appliances from common datasets and a wide set of established features. Furthermore it provides a contribution to a better comparison of NILM studies.

3 NILM FEATURES

In this work, we considered features that allow us to distinguish appliances based on their start-up events. Many of the discussed features are well documented in the literature and are used frequently by the NILM community. Some features were modified, renamed, or adapted for NILM purposes. All features were extracted from the region of interest I_{ROI}, which is 500ms of the start-up current and voltage.

Only appliance steady-state features satisfy the Feature Addition Criterion *FAC* [34], necessary for superimposing concurrently running appliances. In this work, all events are handled as isolated events, ignoring the chance of concurrently running appliances in the household datasets which is known as the switch continuity principle [14]. All features extract information from current and voltage signatures. Voltage-only features do not contain any relevant contribution to the current features since appliance-related fluctuations in the voltage signal are dependent on the current.

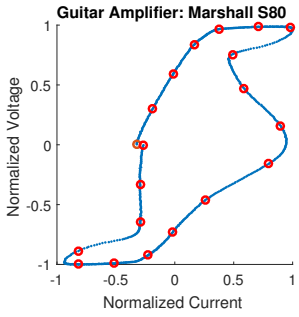

Figure 1: V-I Trajectory with 20 sampling points from a guitar amplifier.

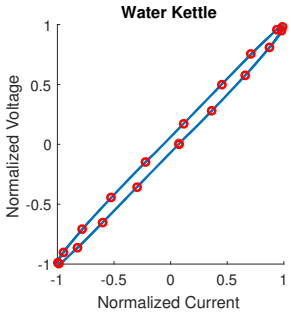

Figure 2: V-I Trajectory with 20 sampling points from a water kettle.

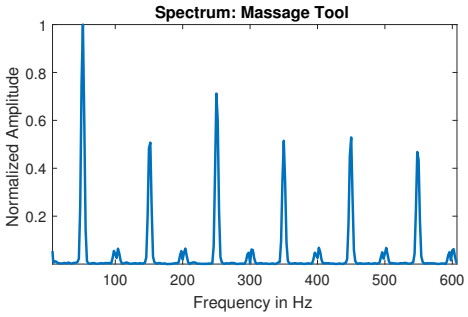

Figure 3: The spectrum of a motor equipped tool with a strong even odd harmonics imbalance.

3.1 Established Features

Early publications on NILM focused mostly on simple-to-compute features. The most used features include electrical power quantities like **Active Power** P, **Reactive Power** Q, and **Apparent Power** $|S|$ [5, 12, 14, 32] which can be considered a standard set for NILM purposes. The **Phase Shift** is the phase angle difference between voltage and current in degrees.

$$P = rms(I) \cdot rms(U) \cdot \cos(\phi) \qquad S = rms(I) \cdot rms(U)$$
$$Q = rms(I) \cdot rms(U) \cdot \sin(\phi) \qquad cos(\phi) = \frac{P}{S}$$

The mutual locus of the normalized instantaneous voltage and current waveforms is called **V-I Trajectory** [15] and is a common feature to visualize signal deformations. Linear (resistive) loads (boiler, heater, kettle, iron, toaster, stove etc.) draw an almost straight line from -1 to 1 due to their waveform equality of current and voltage. The resulting path is sampled by 20 time-equidistant points, creating a vector with 20 coordinates. The belly formed shape of a guitar amplifier and the calculated points are shown in Figure 1. In contrast, the V-I shape of a linear load, such as that from a kettle can be seen in Figure 2.

Many works on NILM focus on signatures based on harmonics [9, 22, 28]. Non-sinusoidal currents cause harmonic characteristics, which can be retrieved from a Fast Fourier Transform. Motor equipped appliances such as those depicted in Figure 3 show usually strong harmonics compared to resistive appliances that show almost none. The relative **Harmonics Energy Distribution** *HED* can be obtained by taking the amplitude of the first 20 harmonics $x_{f_1} \dots x_{f_{20}}$ in a ratio to the mains frequency amplitude x_{f_0}.

$$HED = \frac{1}{x_{f_0}} \cdot \left[x_{f_1}, x_{f_2}, \dots, x_{f_{20}}\right]$$

Wavelet Transformation is another method for retrieving spectral information. Wavelets can be helpful in handling the uncertainty principle in signal processing. The output is a frequency dependent size of the time window with an – often wanted – increased time resolution for higher frequency at the expense of a lower frequency resolution. It therefore results in an increased frequency resolution for lower frequencies at the expense of a lower time resolution [7]. The purpose of this feature is to get a non-linear spectral frequency distribution of 50 variable sized frequency bands.

The **Spectral Flatness** *SPF* is a measure for the energy distribution in the frequency spectrum. A theoretical flatness of 1.0 means that all frequencies show the same amplitude, which is defined as white noise. The closer the flatness comes to 0, the stronger are individual frequencies. Linear loads like a toaster show a relative low spectral flatness, whereas show appliances equipped with a switching mode power supply (SMPS) a relative high spectral flatness. The *SPF* is computed by the ratio of the geometric mean to the arithmetic mean of the energy spectrum [27].

$$SPF = \frac{\sqrt[N]{\prod_{f \in f_{bins}} x_f}}{\frac{1}{N} \sum_{f \in f_{bins}} x_f}$$

Many appliances such as motors or SMPS-equipped appliances form strong odd and only modest even harmonics. This imbalance can vary strongly between different appliance types. This makes the **Odd-Even Harmonics Ratio** *OER* [27] feature a useful characteristic for appliance recognition. Figure 3 shows the harmonics imbalance of a motor-equipped massage appliance.

$$OER = \frac{mean(x_{f_1}, x_{f_3}, \dots, x_{f_{19}})}{mean(x_{f_2}, x_{f_4}, \dots, x_{f_{20}})}$$

A 3-dimensional feature called **Tristimulus** [27] extracts the energy of different harmonic groups. This feature is an audio timbre equivalent to the color attributes in vision [27] and gives a 3-dimensional spectral energy distribution metric. It extracts the intensity for the lower, medium, and higher harmonics, which is different for all appliances.

$$T1 = \frac{x_{f_1}}{\sum_h x_{f_h}} \qquad T2 = \frac{x_{f_2} + x_{f_3} + x_{f_4}}{\sum_h x_{f_h}}$$
$$T3 = \frac{x_{f_5} + x_{f_6} + \dots + x_{f_{10}}}{\sum_h x_{f_h}}$$

The following scalar features focus on the waveform of a signal: **Form Factor** *FF* [26], **Crest Factor** *CF* [30], and logarithmic **Total Harmonic Distortion** *THD* [31]. These signatures correlate if the

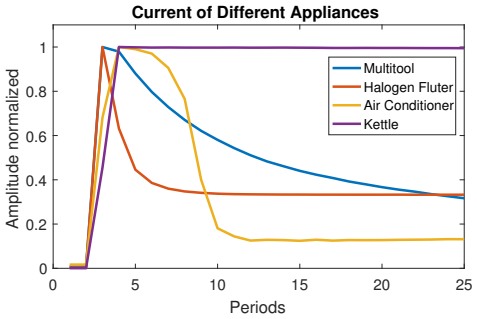

Figure 4: Different current draws over time on period level. The amplitudes are normalized to 1.

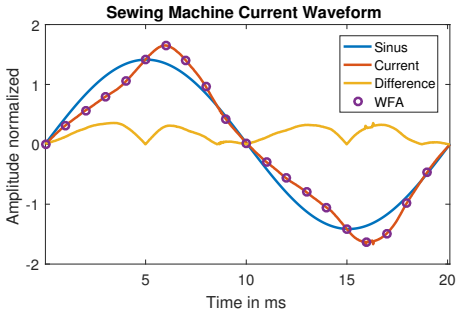

Figure 5: Current waveform of a motor-equipped sewing machine that deforms the current.

waveform is sinusoidal, a compound of strong harmonics or noise. SMPS-equipped appliances produce strong noise in the current waveform which results in a high form factor.

$$FF = \frac{rms(I_{ROI})}{mean(|I_{ROI}|)} \qquad CF = \frac{max(|I_{ROI}|)}{rms(I_{ROI})}$$

$$THD = 10 \cdot log_{10}\left(\frac{1}{x_{f0}} \cdot \sum_{n=1}^{5} x_{fn}\right)$$

Voltage, and therefore also current, can easily drop by a few percent due to rapid changes in the power grid outside the observed circuit. The **Resistance** R is not influenced by voltage fluctuations. Four metrics can be defined using the common Root Mean Square (RMS) and a derived function based on the median (less spike-affected computation). The reciprocal of the resistance is called **Admittance** Y [14] and can be calculated using the quadratic mean and quadratic median accordingly.

$$R_{mean} = \frac{\sqrt{\frac{1}{N} \cdot \sum {U_{ROI}}^2}}{\sqrt{\frac{1}{N} \cdot \sum {I_{ROI}}^2}} \qquad R_{median} = \frac{\sqrt{median\left({U_{ROI}}^2\right)}}{\sqrt{median\left({I_{ROI}}^2\right)}}$$

The Moving Pictures Experts Group (MPEG) published many audio description schemes in the MPEG-7 ISO standard [24]. These descriptors yield to the signal envelope and harmonic characteristics. Both aspects can also be found in energy data. While the signal envelope for musical instruments has an attack, decay, sustain and release state, an electrical appliance has a start-up, decay, steady state and turn off. Musical instrument onsets need to be found in the same way as appliance events. Different instruments draw different harmonic characteristics – similar to electrical appliances. Since high frequency energy data has similarities to audio data, a subset of these descriptors can be adapted for electricity purposes.

The **LogAttackTime** [27] is an envelope feature describing the logarithmic amount of milliseconds until a sound such as a musical instrument reaches its maximum intensity. For NILM purposes the LogAttackTime depicts the time $ln(t_A)$ until the current reaches its maximum $t_A = max(I_{ROI})$. This feature will often result in a low value, although some appliances, like a power drill, can have an increasing current characteristic due to its speed control.

The **Temporal Centroid** C_t describes the temporal balancing point of the current energy in the region of interest [27]. Appliances with a strong start-up current, such as a vacuum cleaner have a significantly different centroid due to higher consumption in the beginning of the event than appliances with a steady current (toaster).

$$I_{W(k)} = [\text{current sample vector}] \quad ; k^{th}\text{period}$$

$$I_{P(k)} = rms(I_{W(k)}) \quad ; k^{th}\text{period}$$

$$C_t = \frac{1}{f_0} \cdot \frac{\sum_{k=1}^{N} I_{P(k)} \cdot k}{\sum_{k=1}^{N} I_{P(k)}}$$

The **Spectral Centroid** C_f [27] defines the balancing point of the given spectrum. Appliances forming a non-linear current (e.g. SMPS-equipped appliances) have a significant higher spectral balancing point than linear loads due to the presence of high order frequencies.

$$C_f = \frac{\sum_{f \in f_{bins}} x_f \cdot f}{\sum_{f \in f_{bins}} x_f}$$

The **Harmonic Spectral Centroid** C_h [27] defines the balancing point based on the first 50 harmonics of the mains frequency. This feature covers similar characteristics as the spectral centroid but ignores non-harmonic noise.

$$C_h = \frac{\sum_{k=1}^{50} x_{f_k} \cdot k}{\sum_{k=1}^{50} x_{f_k}}$$

3.2 Developed Features

Since some appliances characteristics of the investigated datasets motivate further approaches, we developed features to improve the classification of these appliances. These features are retrieved from individual appliance observations, modifications of known features or ideas from existing literature.

The feature **Signal to Signal Mean Ratio** *SSMR* extracts information from the spectrum of the current signal and puts the strongest frequency amplitude into a ratio with the spectral mean. The *SSMR* is a variant of the Signal Noise Ratio *SNR*. Pure resistive appliances such as heaters, water kettles, and ovens are

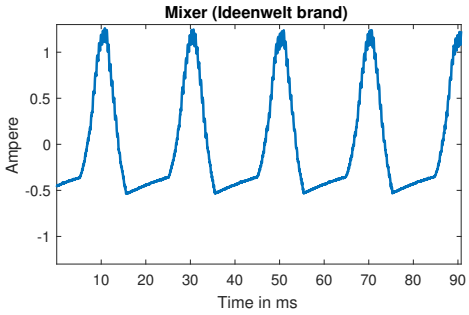

Figure 6: Current of a mixer with cut negative half waves.

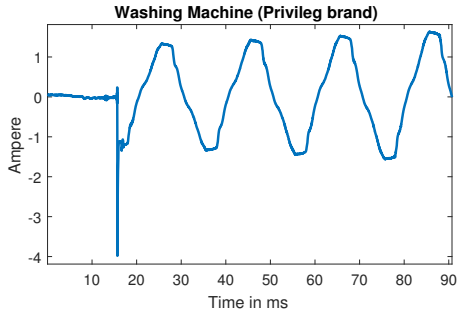

Figure 7: Current of a washing machine with a strong spike at the start-up.

known to have very weak harmonics compared to non-pure resistive loads [32]. Strong harmonics would result in a higher spectral mean. Thus, the *SSMR* can be seen as a scalar representation of the dominance of the strongest frequency in the current spectrum:

$$SSMR = \frac{max(spec)}{mean(spec)}$$

where *spec* is the absolute left-sided frequency spectrum of the whole current sample I_{ROI}.

Appliances that use a SMPS or incandescent light bulbs can have a strong and short current spike in the first period after their start-up (Figure 7). The amplitude of the first period differs in these cases significantly to the periods of the steady state. The feature **Inrush Current Ratio** *ICR* retrieves the RMS of the first period $I_{W(1)}$, and the last period $I_{W(N)}$ of the current region of interest I_{ROI}:

$$ICR = \frac{I_{P(1)}}{I_{P(N)}}$$

Some household appliances show different characteristics in the positive and negative half cycle of their current, which can be seen in the current waveform of a mixer (Figure 6). This behavior is usually caused by dimmers or motor speed controllers, which are widely used to reduce the voltage and therefore also the current. Some of these circuits affect only one half of the current cycle. This can be captured by comparing the RMS of 10 averaged positive and negative current half cycles in the **Positive-Negative Half Cycle Ratio** *PNR*. The mixer in Figure 6 has an imbalance ratio of around 0.6 while most other balanced appliance have a ratio close to 1.0

$$PNR = \begin{cases} \frac{I_{P_{pos}}}{I_{P_{neg}}} & \text{if } I_{P_{neg}} \geq I_{P_{pos}} \\ \frac{I_{P_{neg}}}{I_{P_{pos}}} & \text{if } I_{P_{neg}} < I_{P_{pos}} \end{cases}$$

The **Max-Min Ratio** *MAMI* is an alternative way to cover one-sided waveform characteristics. It puts the maximum and absolute minimum peak current in ratio. In comparison to *PNR*, this feature focuses on the peak values of each half wave that could cover one-sided spikes. The mixer of Figure 6 shows with around 0.3 a significant lower Max-Min Ratio than a theoretical ideal balanced appliance with a ratio of 1.0.

$$MAMI = \begin{cases} \frac{|min(I_{ROI})|}{|max(I_{ROI})|} & \text{if } |max(I_{ROI})| \geq |min(I_{ROI})| \\ \frac{|max(I_{ROI})|}{|min(I_{ROI})|} & \text{if } |max(I_{ROI})| < |min(I_{ROI})| \end{cases}$$

To determine if an appliance has a pure sine current or spikes from switching-artifacts like polling, the absolute maximum peak can be put into a ratio with the absolute mean current to result in a **Peak-Mean Ratio** *PMR*. Again, linear loads show equal values whereas appliances with strong start-up currents, like an incandescent light bulb, are recognizable with this feature.

$$PMR = \frac{max(|I_{ROI}|)}{mean(|I_{ROI}|)}$$

Appliances equipped with a SMPS or compact fluorescent lights (CFL) show very short (sub-period) peaks and spikes, which won't be covered when focusing only on the power of the whole period. Therefore the RMS current of this period is put into ratio to the maximum peak in the first period. This yields the **Max Inrush Ratio** *MIR* feature, a normalized indication of the peak steepness. A linear load has a theoretical *MIR* of $\frac{1}{\sqrt{2}} \approx 0.707$ while a CFL, for example, can have an *MIR* of around 0.3 due to its non-sinusoidal waveform.

$$MIR = \frac{I_{P(1)}}{max(|I_{W(1)}|)}$$

An indicator of the current steadiness can be retrieved by putting the variance and mean of the absolute current into a ratio. A short and high current spike increases the absolute variance to a higher degree than the absolute mean. Our experiments show that linear loads have a relatively low **Mean-Variance Ratio** *MVR* and appliances with a short peak (e.g. light bulb) a relatively high *MVR*.

$$MVR = \frac{mean(|I_{ROI}|)}{var(|I_{ROI}|)}$$

The mains voltage follows a relative sinusoidal waveform. Various non-linear appliances distort the current signal by adding individual non-sinusoidal characteristics. Figure 5 shows the waveform of a sewing machine compared to a generated sine wave. The goal of the **Waveform Distortion** *WFD* feature is to obtain a distortion metric of the current waveform compared to a single

period of a sine wave with the same energy Y_{sin}. Therefore, the first 10 post-start-up periods are averaged, normalized (with the RMS of itself), and aligned to the rising zero crossing. The absolute current wave is subtracted from the equivalent absolute sine wave and the differences are summed up. The smaller the value, the more sinusoidal and similar are the half waves of the current waveform.

$$I_{WM} = mean(I_{W(1)}, \ldots, I_{W(10)})$$
$$WFD = sum(|Y_{sin}| - |I_{WM}|)$$

A higher degree of information can be extracted from the waveform itself (e.g. square, saw-tooth, single-pulse). By taking the mean of the first 10 periods point-by-point and down-sampling this vector to 20 points as in Figure 5, we get the multi-dimensional **Waveform Approximation** *WFA* feature.

$$WFA = downsample(I_{WM}, \left\lfloor \frac{\text{length}(I_{WM})}{20} \right\rfloor)$$

Many appliances, including vacuum cleaners, light bulbs, and motor-based devices, do not have a steady but rather a decreasing power consumption over the initial start-up phase. Due to these variations in the period-by-period current, combining the RMS currents of each period from I_{ROI} into a multi-dimensional vector gives us a **Current Over Time** *COT* distribution. The **Admittance Over Time** *AOT* is less influenced by voltage fluctuation and is calculated by dividing current $I_{P(k)}$ and voltage $U_{P(k)}$. Figure 4 shows the current over time of four appliances; the maximum current is normalized to 1.0.

$$COT = \left[I_{P(1)}, I_{P(2)}, \ldots, I_{P(25)}\right]$$
$$AOT = \left[\frac{I_{P(1)}}{U_{P(1)}}, \frac{I_{P(2)}}{U_{P(2)}}, \ldots, \frac{I_{P(25)}}{U_{P(25)}}\right]$$

The inrush current is higher than the steady state current in many appliances. This inrush current significantly differs in time throughout various appliance types. The index of the first period, after the initial current increase levels off, can be used as the **Periods to Steady State Current** *PSS* feature. The steady state is reached when $I_{P(k)}$ falls below a pre-calculated limit above the median of *COT*.

$$L = \frac{1}{8} \cdot (max(COT) - median(COT)) + median(COT)$$
$$PSS = k \quad ; \text{ first period, where: } I_{P(k)} < L$$

To obtain the advantages of higher sampling frequencies, one can shift the main focus in the spectrum by applying a high-pass filter with a pass frequency of 5 kHz resulting in the following new features: **High Frequency Spectral Centroid** *HFSPC* and **High Frequency Spectral Flatness** *HFSPF*.

Using the high-pass-filtered spectrum one can use the **High Frequency Spectral Mean** *HFSPM* as an indication of the appliance impact on high frequency regions.

$$HFSPM = mean(x_f \text{ for } f \geq 5\,\text{kHz}, f \in f_{bins})$$

4 EXPERIMENTAL METHODOLOGY

We set up a machine learning test bed, to evaluate the classification performance of the above defined features. We used start-up events of PLAID [11], WHITED [17], UK-DALE [18], and BLUED [1] to evaluate various features and their combinations with different classifiers.

PLAID and WHITED focus on isolated single appliance measurements. From WHITED, a typical household subset is used (one appliance model per type, low inner-class diversity, 280 events of 27 appliance types). PLAID provides multiple appliance models per type (e.g. 7 heaters, 33 laptops, etc.). All appliances and events are used in our experiments (high inner-class diversity with 1074 events of 11 appliance types).

UK-DALE and BLUED provide real household measurements. Start-up events were extracted based on timestamps and event labels from the provided low frequency single appliance measurements. We defined an event as a spontaneous, single and significant rise in the current consumption for a couple of mains periods.

Our machine learning experiments follow a typical pattern recognition approach. Figure 8 shows the appliance classification pipeline with supervised learning embedded in the evaluation system. We define the classification problem as follows: Classify the appliance type based on a start-up event and represent each appliance type as an individual class.

For reliable results we used a stratified 5-fold cross-validation with four classifiers: *1-Nearest Neighbour* (KNN), *Binary Decision Tree* (BDT), *Linear Discriminant Analysis* (LDA), and *Support Vector Machines* (SVM). All computations are performed with a combination of Matlab (with LIBSVM [6] and various toolboxes), and Python (with common scientific computing packages). To remove unwanted feature weighting caused by different ranges of values, we implemented a feature variance normalization, based on the training data in each cross-validation step.

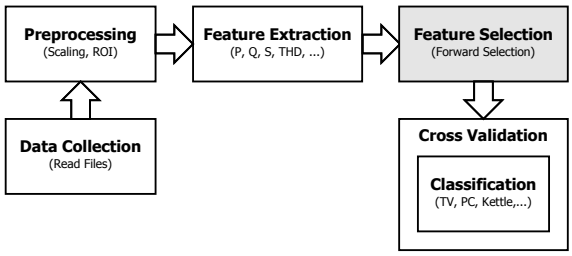

Figure 8: The simplified steps of our evaluation system based on pattern recognition and cross-validation.

Due to space limitations we present only the most relevant interpretations of a confusion matrix, where TP is the amount of True Positives, FP of False Positives, TN of True Negatives, and FN of False Negatives:

F1 Score: $$F1 = 2 \cdot \frac{Pr \cdot Re}{Pr + Re}$$

Precision: $$Pr = \frac{TP}{TP + FP}$$

Recall: $$Re = \frac{TP}{TP + FN}$$

Accuracy: $$Ac = \frac{TP + TN}{TP + FP + FN + TN}$$

These classification performance metrics are non-weighted averages over all appliance classes. This ensures that a class with a large amount of samples contributes in equal parts to the average as a class with a small amount of samples. Since F1 Score is the most common used metric, all features and feature combinations are ranked based on the F1 Score; the higher the score, the better the classification performance.

5 EXPERIMENTAL RESULTS

All experiments for isolated measurements were performed on PLAID and the WHITED subset. The selected appliances can be seen in Table 7. To cover real world scenarios, we applied our experiments on two high frequency household datasets: BLUED and UK-DALE. With these datasets we rank all stand-alone features, find the best 2-dimensional feature combination, and compute the best forward-selected combination. Therefore, we try to find the best performing features for the following questions:

What is the highest achievable classification performance …

- …for data with a high inner-class diversity? (PLAID)
- …for data with a low inner-class diversity? (WHITED)
- …for real household measurements? (BLUED, UK-DALE)

5.1 Stand-Alone Feature Ranking

The goal of this ranking is to compute the individual classification performance of each feature and rank them in a table. To be able to compare all features against one another, it is necessary to compute the stand-alone classification performance of each feature by itself.

The evaluation was performed over all datasets with a *1-Nearest Neighbor* classifier for comparability, traceability, and reproducibility (Table 1). It is important to differentiate between one- and multi-dimensional features, since multi-dimensional features have a significant higher potential in classification performance than scalar features.

For isolated measurements (PLAID and WHITED), the *Phase Shift* with an average F1 Score of 0.74 has the highest discriminating quality. This means that the most relevant one dimensional metric to recognize appliances lies in the individual voltage and current phase difference of each appliance. The best multidimensional feature is the *Wavelet Analysis* with an average F1 Score of 0.95. This on the other hand, means that, due to individual nonlinearities inside the appliances, complex spectral characteristics offer the highest discriminating quality for isolated events.

For real household measurements (BLUED and UK-DALE), the best scalar feature are the *Admittance* and the *Resistance* with an average F1 Score of 0.30. Note that classification based on a single scalar feature is difficult for any classifier. Since the household data contains more noise than isolated measurements, the recognition quality is significantly lower. The best multidimensional feature is *Admittance Over Time* with an average F1 Score of 0.69. For aggregated measurements, the individual temporal appliance energy consumption shows the highest robustness against unwanted interferences.

When ranking the features of all datasets and environments, the best scalar feature is *Phase Shift*, and the best multi-dimensional feature is *Current Over Time*.

5.2 2-Dimensional Feature Combination

Combining multiple features usually improves the classification performance. However, each additional feature increases computational complexity. Therefore, this experiment focuses on evaluating 2-dimensional feature combinations for each dataset and classifier, while keeping the complexity to a minimum. One can use a 2-dimensional feature space to visualize and examine the class boarders (Figure 9). A common visualization is the *P-Q* plane [12, 14, 32] where *Active Power* and *Reactive Power* form a 2 dimensional scatter plot.

We evaluated all possible combinations of scalar features, resulting in 406 possible pairs. The *PF-THD* plane shows an interesting sample distribution in terms of appliance clusters (Figure 10). The advantage of this plane is the high intra-class variance of low-power appliances, compared to the closely grouped clusters in the *P-Q* plane (Figure 9).

The best 2-dimensional feature combination for each classifier can be seen in Tables 2, 3, 4, and 5. For the WHITED subset, the best 2-dimensional feature combination (using KNN) is *Phase Shift* and *Admittance*, compared to *Active Power* and *Reactive Power* on PLAID. For BLUED, the best 2-dimensional feature combination (using KNN) is *Apparent Power* and *Phase Shift*, compared to *Active Power* and *Inrush-Steady State Ratio* on UK-DALE (using BDT). These results suggest that consistent high classification performance can be achieved, by including one of the power indicators (*Active*, *Reactive*, *Apparent Power*, etc.) in the feature combination.

5.3 Feature Forward Selection

As shown in Table 1, many features reach an already high stand-alone F1 Score for the appliance classification. The combination of several features usually improves the classification performance up to a limit. Computation time and classification aggravations anomalies due to large feature spaces like the Hughes phenomenon [16] motivate to keep the amount of features to a minimum. We chose a forward selection algorithm (Figure 11) to find the best compromise between classification performance and a smaller number of features. An evaluation that considers all possible 2^{36} combinations is not feasible in terms of computational effort and resources. The algorithm starts by selecting the best stand-alone feature, then computes all possible 2-pairs and selects the best one as starting point for the next iteration. This is repeated until the classification performance stops improving.

Table 6 shows the resulting feature combinations for all datasets computed by the feature forward selection algorithm. Since we used a randomized stratified 5-fold cross validation, the performance results vary slightly for each run. Stratified means in this context that the amount of training samples are balanced for each class. We achieved an F1 Score of 1.0, due to the intrinsic properties of WHITED (isolated measurements). PLAID showed comparably high classification performance, however, with a different set of features. BLUED and UK-DALE scored slightly below 0.80 due to noise and concurrently running appliances.

Feature	Time Window	Domain	Signal	Source	FAC	Dim	W[1]	P[2]	B[3]	U[4]	ø
Active Power	whole event	time	I, U	electronics	✓	1	0.71	0.49	0.32	0.24	0.44
Admittance	whole event	time	I, U	electronics	✓	1	0.75	0.52	**0.34**	**0.26**	0.47
Admittance (median)	whole event	time	I, U	electronics	✓	1	0.83	0.58	0.28	0.18	0.47
Apparent Power	whole event	time	I, U	electronics	✓	1	0.77	0.52	0.32	0.26	0.47
Crest Factor	whole event	time	I	electronics	✓	1	0.33	0.33	0.23	0.21	0.28
Even-Odd Harmonics Ratio	whole event	spectral	I	audio	✓	1	0.32	0.24	0.17	0.17	0.22
Form Factor	whole event	time	I	electronics	✓	1	0.62	0.45	0.15	0.14	0.34
Harmonic Spectral Centroid	whole event	spectral	I	MPEG 7	✓	1	0.54	0.32	0.16	0.09	0.28
Inrush Current Ratio	first, last period	time	I	developed		1	0.45	0.28	0.27	0.17	0.29
Log Attack Time	first 25 periods	time	I	MPEG 7		1	0.26	0.35	0.16	0.19	0.24
Max Inrush Ratio	first period, max sample	time	I	developed		1	0.27	0.29	0.15	0.07	0.20
Mean-Variance Ratio	whole event	time	I	developed	✓	1	0.28	0.30	0.18	0.14	0.23
Min-Max Ratio	min, max sample	time	I	developed	✓	1	0.17	0.28	0.20	0.19	0.21
Peak-Mean Ratio	whole event	time	I	developed		1	0.32	0.33	0.25	0.20	0.28
Periods to Steady State Current	first 25 periods	time	I	MPEG 7		1	0.18	0.22	0.11	0.15	0.17
Positive-Negative Half Cycle Ratio	mean first 10 periods	time	I	developed	✓	1	0.11	0.15	0.11	0.07	0.11
Phase Shift	whole event	spectral	I, U	electronics	✓	1	0.83	**0.64**	0.27	0.22	**0.49**
Reactive Power	whole event	time	I, U	electronics	✓	1	0.83	0.54	0.26	0.22	0.46
Resistance	whole event	time	I, U	electronics	✓	1	0.77	0.54	**0.34**	**0.26**	0.48
Resistance (median)	whole event	time	I, U	electronics	✓	1	**0.84**	0.57	0.30	0.17	0.47
Signal-Signal Mean Ratio	whole event	spectral	I	developed	✓	1	0.66	0.39	0.16	0.15	0.34
Spectral Centroid	whole event	spectral	I	MPEG 7	✓	1	0.75	0.42	0.15	0.09	0.35
Spectral Centroid HF	whole event	spectral	I	developed	✓	1	0.38	0.23	0.07	0.07	0.19
Spectral Flatness	whole event	spectral	I, U	audio	✓	1	0.68	0.37	0.16	0.08	0.32
Spectral Flatness HF	whole event	spectral	I, U	developed	✓	1	0.27	0.25	0.09	0.07	0.17
Spectral Mean HF	whole event	spectral	I, U	developed	✓	1	0.73	0.41	0.16	0.10	0.35
Temporal Centroid	first 25 periods	time	I	MPEG 7		1	0.50	0.38	0.22	0.17	0.32
Total Harmonic Distortion	whole event	spectral	I	audio	✓	1	0.73	0.50	0.27	0.21	0.43
Waveform Distortion	mean first 10 periods	time	I	developed	✓	1	0.46	0.40	0.13	0.09	0.27
Tristimulus	whole event	spectral	I	audio	✓	3	0.84	0.47	0.28	0.25	0.46
Harmonics	whole event	spectral	I	audio	✓	20	**0.99**	0.90	0.54	0.60	0.76
Waveform Approximation	mean first 10 periods	time	I	developed	✓	20	0.85	0.82	0.49	0.49	0.66
Admittance Over Time	first 25 periods	time	I, U	developed		25	0.98	0.82	**0.61**	**0.76**	0.79
Current Over Time	first 25 periods	time	I	developed		25	0.97	0.87	0.60	0.75	**0.80**
VI-Trajectory	mean first 10 periods	time	I, U	electronics	✓	40	0.93	0.88	0.39	0.45	0.66
Wavelet Analysis	whole event	spectral	I	audio	✓	50	**0.99**	**0.91**	0.56	0.67	0.78

[1] F1 Score of WHITED [2] F1 Score of PLAID [3] F1 Score of BLUED [4] F1 Score of UK-DALE

Table 1: Feature Overview: Features are extracted from different temporal regions of an event, which can be found in column *Time Window*. *Source* describes roughly the usual field of research or standard from which the feature originates. The value *developed* represents features that are newly developed or adapted from existing features in this work. The features with the highest F1 Score – scalar and multi-dimensional – are in bold and the overall winners are highlighted in gray.

	Features	F1	Pr	Re	Ac
KNN	**Phase Shift** **Admittance**	**0.98**	**0.98**	**0.98**	**0.99**
LDA	Reactive Power Signal-Signal Mean Ratio	0.92	0.94	0.93	0.99
SVM	Total Harmonic Distortion Harmonic Spectral Centroid	0.98	0.98	0.98	0.99
BDT	Spectral Mean (HF) Temporal Centroid	0.97	0.97	0.97	0.99

Table 2: The best 2-dimensional feature combination for each classifier with WHITED: *Phase Shift* and *Admittance* show promising results. Spectral indicators reach similar results.

	Features	F1	Pr	Re	Ac
KNN	**Active Power** **Reactive Power**	**0.89**	**0.91**	**0.88**	**0.99**
LDA	Phase Shift Temporal Centroid	0.54	0.55	0.58	0.95
SVM	Phase Shift Total Harmonic Distortion	0.86	0.90	0.84	0.98
BDT	Phase Shift Total Harmonic Distortion	0.82	0.84	0.80	0.98

Table 3: The best 2-dimensional feature combination for each classifier with PLAID: *Active Power* and *Reactive Power*

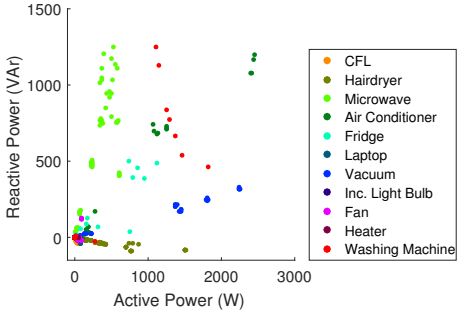

Figure 9: PLAID in P-Q plane

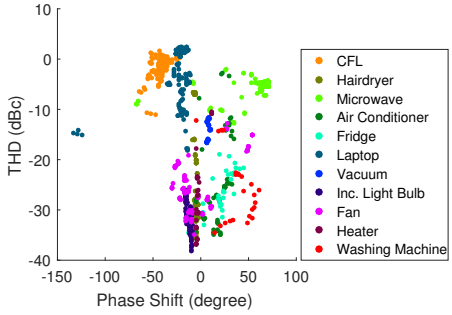

Figure 10: PLAID in PS-THD plane

	Features	F1	Pr	Re	Ac
KNN	**Apparent Power** **Phase Shift**	**0.59**	0.60	**0.59**	**0.97**
LDA	Active Power Signal-Signal Mean Ratio	0.36	0.36	0.38	0.96
SVM	Resistance (med) Temporal Centroid	0.57	**0.67**	0.57	0.97
BDT	Active Power Temporal Centroid	0.57	0.59	0.57	0.97

Table 4: The best 2-dimensional feature combination for each classifier with BLUED: The performance metrics show that the *Phase Shift* performs well, even for noisy household data.

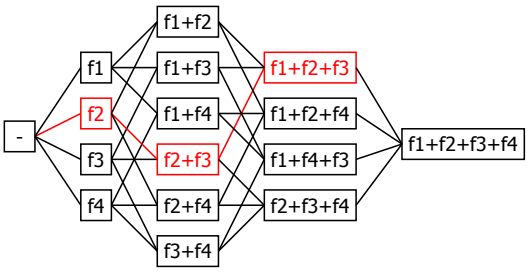

Figure 11: Feature Forward Selection: In this example, *f2* is selected as winner of the first iteration, [*f2, f3*] in the second, and [*f1, f2, f3*] as final winner, since the combination of all four features does not further improve the performance.

	Features	F1	Pr	Re	Ac
KNN	Inrush-Steady State Ratio Admittance	0.59	0.59	0.59	0.99
LDA	Reactive Power Total Harmonic Distortion	0.31	0.32	0.35	0.98
SVM	Active Power Inrush-Steady State Ratio	0.56	0.61	0.56	0.99
BDT	**Active Power** **Inrush-Steady State Ratio**	**0.60**	**0.61**	**0.60**	**0.99**

Table 5: The best 2-dimensional feature combination for each classifier with UK-DALE: In this case, the Binary Decision BDT reaches the highest classification performance with *Active Power* and *Inrush-Steady State Ratio*.

5.4 Individual Appliance Performance

Table 7 shows that the classification performance is not identical for each appliance type. In the case of WHITED, there are no misclassifications since the dataset consists of isolated measurements and only one model per appliance type, which lowers the inner-class variance and therefore improves the classification performance significantly.

Aside from some outliers, most appliance types show a relatively high classification performance. The appliance type *Lights* shows a notably lower classification performance in both datasets. We believe that the main reason for this result is a too broad definition of lights, which may include a mix of different lighting types (e.g. incandescent, LED, and CFL) and low power lights. A similar problem applies to *Desktop PC's*, which are usually equipped with a SMPS. The total classification fail of the boiler cannot be explained completely. One reason can be that resistive loads have almost no recognizable current signatures that can easily cause a misclassification to another resistive load of similar consumption.

5.5 Discussion

What is the best general feature combination for appliance recognition? First of all, there is no one-size-fits-all set of features. The individual composition of appliances in the environment determines the set of features. However, the best feature sets always contain information about a spectral energy distribution (e.g. *Wavelet Analysis*, *Harmonics*, etc.) and time-related power information (e.g. *Current and Admittance Over Time*). The results show that the main information to discriminate between appliances lies in the spectral distribution and the unsteadiness in the power consumption over time.

A poor result of a feature in Table 1 does not mean that it is worthless. Some features are based on characteristics that are formed only by a rare amount of appliances like the feature *PNR*. The waveform imbalance covered by this feature was only observed in the *Ideenwelt*-mixer of WHITED. The feature may have a poor general

	Features	F1	Pr	Re	Ac
W	Wavelet Analysis, Spectral Mean (HF), VI-Trajectory, Current Over Time	1.00	1.00	1.00	1.00
P	Wavelet Analysis, VI-Trajectory, Admittance Over Time, Form Factor, Phase Shift, Log Attack Time, Current Over Time, Max-Inrush Ratio, Resistance (med)	0.96	0.97	0.95	0.99
B	Admittance Over Time, Resistance (med), Total Harmonic Distortion, Spectral Mean (HF), Temporal Centroid, Phase Shift, Admittance, Admittance (med), Log Attack Time, Spectral Flatness, Sinus Difference Sum, Harmonic Spectral Centroid	0.76	0.78	0.75	0.98
U	Admittance Over Time, Resistance, Phase Shift, Temporal Centroid, Admittance, Active Power, Even-Odd Harmonics Ratio, Spectral Flatness	0.79	0.80	0.79	0.99

Table 6: The results of the forward selection for each dataset. Even with only four features, the optimal classification performance for WHITED is achievable. The results show that not all features are necessary to get the optimal classification performance.

discrimination quality but might contribute in a feature combination to distinguish the mixer from a similar appliance such as a multitool.

When focusing on the algorithms, the 1-nearest neighbor classifier leads in performance quality. It seems to be the optimal classifier for this task due to its simplicity and computation performance for small to medium feature spaces. Since the SVM classifier has a strong need for an intense parameter-search, a rough parameter search was used for the 2-dimensional feature combination due to computing time.

6 CONCLUSION

We performed an appliance recognition evaluation on 4 different high frequency datasets to identify the best features and feature combinations out of 36 implemented signatures from different research areas. The phase angle difference between voltage and current has the highest scalar performance across all datasets, while the multi-dimensional feature *Current Over Time* shows the most promising results in general.

According to our findings, the *Wavelet Analysis* discriminates best for isolated environments while the *Current Over Time* scores best for aggregated environments. Dataset-related results for two or more features from a large number of feature combination were empirically retrieved. To gain the best feature combination, a forward selection algorithm was applied for each datasets. Beside the features itself, the results of the classification performance for each individual appliance type shows that almost all appliances can be recognized with the introduced set of features. Only the lights and the boiler in the UK-DALE dataset could not be recognized.

Appliance	W KNN	P KNN	B SVM	U SVM	Ø
Air Conditioner	1.00	0.93			0.97
Air Compressor			1.00		1.00
Boiler				0.00	0.00
Breadmaker				0.97	0.97
Charger	1.00				1.00
Coffee Machine	1.00			0.88	0.94
CFL	1.00	1.00			1.00
Desktop PC	1.00			[1]0.63	0.82
Dishwasher				0.94	0.94
Drilling Machine	1.00				1.00
Fan	1.00	0.97			0.98
Fridge	1.00	0.75	0.95	0.99	0.92
Game Console	1.00				1.00
Garage Door			0.82		0.82
Hair Dryer	1.00	0.98		0.83	0.94
Heater		0.97			0.97
HiFi	1.00		0.86		0.93
Inc. Light Bulb	1.00	0.97			0.99
Iron	1.00		1.00	0.96	0.99
Juice Maker	1.00				1.00
Kettle	1.00			0.97	0.98
Kitchen Hood	1.00				1.00
Laptop	1.00	0.99			1.00
Lights			[2]0.60	0.04	0.32
Microwave	1.00	0.99		0.95	0.98
Mixer	1.00				1.00
Monitor			0.74		0.74
Printer	1.00		1.00		1.00
Rice Cooker	1.00				1.00
Sandwich Maker	1.00				1.00
Straighteners	1.00			0.86	0.93
Toaster	1.00			0.92	0.96
TV	1.00		0.85	0.83	0.90
Vacuum Cleaner	1.00	1.00		0.97	0.99
Washing Machine	1.00	0.94		0.93	0.96

[1] Mean of HTPC and Office PC [2] Mean of all eight lights

Table 7: F1 Score of each individual appliance in its corresponding dataset. The values are based on forward selected feature combinations and the best performing classifier.

ACKNOWLEDGEMENTS

This research was partially funded by the Alexander von Humboldt Foundation established by the government of the Federal Republic of Germany and was supported by the Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag.

REFERENCES

[1] Kyle Anderson, Adrian Ocneanu, Diego Benitez, Derrick Carlson, Anthony Rowe, and Mario Berges. 2012. BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research. In *Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD)* (2012-08). ACM.

[2] Kathleen Carrie Armel, Abhay Gupta, Gireesh Shrimali, and Adrian Albert. 2013. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy* 52 (jan 2013), 213–234. DOI: http://dx.doi.org/10.1016/j.enpol.2012.08.062

[3] Karim Said Barsim, Lukas Mauch, and Bin Yang. 2016. Neural Network Ensembles to Real-time Identification of Plug-level Appliance Measurements. *Signature* 2

(2016), 11.

[4] Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. 2014. NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. In *Proceedings of the 5th international conference on Future energy systems*. ACM, New York, NY, USA, 265–276. DOI: http://dx.doi.org/10.1145/2602044.2602051

[5] Mario Berges, Ethan Goldman, H Scott Matthews, and Lucio Soibelman. 2009. Learning Systems for Electric Consumption of Buildings. In *ASCI international workshop on computing in civil engineering* (2009), Vol. 38. American Society of Civil Engineers (ASCE). DOI: http://dx.doi.org/10.1061/41052(346)1

[6] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology* 2, 3 (April 2011), 271–2727. Issue 3. DOI: http://dx.doi.org/10.1145/1961189.1961199 Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[7] Ingrid Daubechies and others. 1992. *Ten Lectures on Wavelets*. Vol. 61. SIAM. DOI: http://dx.doi.org/10.1137/1.9781611970104

[8] Karen Ehrhardt-Martinez, Kat A Donnelly, and others. 2010. *Advanced Metering Initiatives and Residential Feedback Programs: A Meta-Review for Household Electricity-Saving Opportunities*. resreport. ACEEE.

[9] Frank Englert, Till Schmitt, Sebastian Kößler, Andreas Reinhardt, and Ralf Steinmetz. 2013. How to Auto-Configure Your Smart Home? High-Resolution Power Measurements to the Rescue. In *Proceedings of the fourth international conference on Future energy systems* (2013). ACM, Association for Computing Machinery (ACM), 215–224. DOI: http://dx.doi.org/10.1145/2487166.2487191

[10] Jon Froehlich, Eric Larson, Sidhant Gupta, Gabe Cohn, Matthew Reynolds, and Shwetak Patel. 2011. Disaggregated End-Use Energy Sensing for the Smart Grid. *IEEE Pervasive Computing* 10, 1 (jan 2011), 28–39. DOI: http://dx.doi.org/10.1109/mprv.2010.74

[11] Jingkun Gao, Suman Giri, Emre Can Kara, and Mario Bergés. 2014. PLAID: A Public Dataset of High-Resolution Electrical Appliance Measurements for Load Identification Research. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings* (2014). ACM, Association for Computing Machinery (ACM), 198–199. DOI: http://dx.doi.org/10.1145/2674061.2675032

[12] Jingkun Gao, Emre Can Kara, Suman Giri, Mario Berg, and others. 2015. A feasibility study of automated plug-load identification from high-frequency measurements. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (2015). IEEE, Institute of Electrical and Electronics Engineers (IEEE), 220–224. DOI: http://dx.doi.org/10.1109/globalsip.2015.7418189

[13] Sidhant Gupta, Matthew S Reynolds, and Shwetak N Patel. 2010. 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* (2010). ACM, ACM Press, 139–148. DOI: http://dx.doi.org/10.1145/1864349.1864375

[14] George William Hart. 1992. Nonintrusive Appliance Load Monitoring. *Proc. IEEE* 80, 12 (1992), 1870–1891.

[15] Taha Hassan, Fahad Javed, and Naveed Arshad. 2014. An Empirical Investigation of V-I Trajectory Based Load Signatures for Non-Intrusive Load Monitoring. *IEEE Transactions on Smart Grid* 5, 2 (mar 2014), 870–878. DOI: http://dx.doi.org/10.1109/tsg.2013.2271282

[16] G. Hughes. 1968. On the Mean Accuracy of Statistical Pattern Recognizers. *IEEE Transactions on Information Theory* 14, 1 (01 1968), 55–63. DOI: http://dx.doi.org/10.1109/TIT.1968.1054102

[17] Matthias Kahl, Anwar Ul Haq, Thomas Kriechbaumer, and Hans-Arno Jacobsen. 2016. WHITED - A Worldwide Household and Industry Transient Energy Data Set. In *3rd International Workshop on Non-Intrusive Load Monitoring*.

[18] Jack Kelly and William Knottenbelt. 2015. The UK-DALE Dataset, Domestic Appliance-Level Electricity Demand and Whole-House Demand from Five UK Homes. *Scientific Data* 2 (mar 2015), 150007. DOI: http://dx.doi.org/10.1038/sdata.2015.7

[19] Jeremy Zico Kolter and Matthew J Johnson. 2011. REDD: A Public Data Set for Energy Disaggregation Research. In *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA* (2011), Vol. 25. Citeseer, ACM, 59–62.

[20] Hong Yin Lam, G.S.K. Fung, and W.K. Lee. 2007. A Novel Method to Construct Taxonomy Electrical Appliances Based on Load Signatures. *IEEE Transactions on Consumer Electronics* 53, 2 (05 2007), 653–660. DOI: http://dx.doi.org/10.1109/tce.2007.381742

[21] Steven B Leeb, Steven R Shaw, and James L Kirtley. 1995. Transient Event Detection in Spectral Envelope Estimates for Nonintrusive Load Monitoring. *IEEE Transactions on Power Delivery* 10, 3 (jul 1995), 1200–1210. DOI: http://dx.doi.org/10.1109/61.400897

[22] Jian Liang, Simon KK Ng, Gail Kendall, and John WM Cheng. 2010. Load signature study—Part I: Basic concept, structure, and methodology. *Power Delivery, IEEE Transactions on* 25, 2 (2010), 551–560.

[23] Y. H. Lin, M. S. Tsai, and C. S. Chen. 2011. Applications of Fuzzy Classification with Fuzzy CMeans Clustering and Optimization Strategies for Load Identification in NILM Systems. In *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)* (2011-06). Institute of Electrical and Electronics Engineers (IEEE), 859–866. DOI: http://dx.doi.org/10.1109/FUZZY.2011.6007393

[24] MPEG. 2005. MPEG-7 Audio Descriptions. (2005). http://mpeg.chiariglione.org/standards/mpeg-7/audio

[25] Shwetak N Patel, Thomas Robertson, Julie A Kientz, Matthew S Reynolds, and Gregory D Abowd. 2007. At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line. *Lecture Notes in Computer Science* 4717 (2007), 271–288.

[26] Rupert Patzelt and Herbert Schweinzer. 2013. *Elektrische Meßtechnik*. Springer-Verlag. DOI: http://dx.doi.org/10.1007/978-3-7091-6557-7

[27] G. Peeters. 2004. *A large set of audio features for sound description (similarity and classification) in the CUIDADO project*. Tech. Rep. IRCAM.

[28] Andreas Reinhardt, Dominic Burkhardt, Manzil Zaheer, and Ralf Steinmetz. 2012. Electric Appliance Classification Based on Distributed High Resolution Current Sensing. In *37th Annual IEEE Conference on Local Computer Networks, Workshop Proceedings*. IEEE, 999–1005. DOI: http://dx.doi.org/10.1109/lcnw.2012.6424093

[29] JG Roos, IE Lane, EC Botha, and Gerhard P Hancke. 1994. Using Neural Networks for Non-intrusive Monitoring of Industrial Electrical Loads. In *Instrumentation and Measurement Technolgy Conference*. IEEE, IEEE, 1115–1118. DOI: http://dx.doi.org/10.1109/imtc.1994.351862

[30] Tony J Rouphael. 2009. *RF and digital signal processing for software-defined radio: a multi-standard multi-mode approach*. Newnes.

[31] Doron Shmilovitz. 2005. On the Definition of Total Harmonic Distortion and Its Effect on Measurement Interpretation. *IEEE Transactions on Power Delivery* 20, 1 (jan 2005), 526–528. DOI: http://dx.doi.org/10.1109/tpwrd.2004.839744

[32] F Sultanem. 1991. USING APPLIANCE SIGNATURES FOR MONITORING RESIDENTIAL LOADS AT METER PANEL LEVEL. *Power Delivery, IEEE Transactions on* 6, 4 (1991), 1380–1385. DOI: http://dx.doi.org/10.1109/61.97667

[33] KH Ting, Mark Lucente, George SK Fung, WK Lee, and SYR Hui. 2005. A Taxonomy of Load Signatures for Single-Phase Electric Appliances. In *IEEE PESC (Power Electronics Specialist Conference)*. IEEE, 12–18.

[34] Hong-Tzer Yang, Hsueh-Hsien Chang, and Ching-Lung Lin. 2007. Design a Neural Network for Features Selection in Non-intrusive Monitoring of Industrial Electrical Loads. In *11th International Conference on Computer Supported Cooperative Work in Design* (2007). IEEE, 1022–1027.