Non-Intrusive Load Monitoring: A Review

Pascal A. Schirmer and Iosif Mporas

***Abstract*—The rapid development of technology in the electrical energy sector within the last 20 years has led to growing electric power needs through the increased number of electrical appliances and automation of tasks. In parallel the global climate protection goals, energy conservation and efficient energy management arise interest for reduction of the overall energy consumption. These requirements have led to the recent adoption of smart-meters and smart-grids, as well as to the rise of Load Monitoring (LM) using energy disaggregation, also referred to as Non-Intrusive Load Monitoring (NILM), which enables appliance-specific energy monitoring by only observing the aggregated energy consumption of a household. The real-time information on appliance level can be used to get deeper insights in the origin of energy consumption and to make optimization, strategic load scheduling and demand management feasible. The three main contributions are as follows: First, a generalized up-to-date review of NILM approaches including a high-level taxonomy of NILM methodologies is provided. Second, previously published results are grouped based on the experimental setup which allows direct comparison. Third, the article is accompanied by a software implementation of the described NILM approaches.**

***Index Terms*—Energy disaggregation, non-intrusive load monitoring (NILM), smart meter, smart grid.**

I. INTRODUCTION

WITH the development of technology and the increasing usage of electrical appliances and automated services, the electric energy needs have been growing steadily for the last century with an annual growth of approximately 3.4% per year in the last decade [1]. Nowadays residential and commercial buildings account already for roughly 36% of the total electrical demand in the USA and 25% in the EU [2], [3]. In parallel, studies indicate that detailed analysis and real-time feedback of energy consumption can lead up-to 20% savings in energy consumption through the detection of faulty devices and poor operational strategies [4]–[6]. Therefore, in the last few decades extensive research in smart grids, smart systems, and energy demand management was carried out and different optimization techniques have been developed to reduce residential energy consumption [7], [8]. To make use of those techniques accurate and fine-grained monitoring of electrical energy consumption is needed [9]. Specifically, the systematic review conducted by [10] states that even reporting the real-time aggregated consumption to households has an positive effect on their consumer behaviour, while according to [11] the largest improvements in terms of energy savings can be made when monitoring energy consumption on device level.

When only the aggregated consumption is measured and the device consumption is estimated through disaggregation algorithms, the task of LM is referred to as NILM [12], in contrast to Intrusive Load Monitoring (ILM) where multiple sensors are used, usually one per device. While NILM was initially proposed to minimizing the number of installed energy meters and thus to reduce the wiring harness and improve the retrofitting capabilities [12], [13], several other utility and non-utility applications have been investigated. As regards utility applications, energy consumption reduction for residential [10], [14] and industrial [15] areas is the most common application. Furthermore, NILM has been used in energy management of smart-grids to optimize load schedules as well as to increase customers' satisfaction [7], [16]. Moreover, NILM has been used to improve load forecasting utilizing specific appliance usage patterns [17]. In non-utility applications NILM has been used for fault detection and diagnostics in both the industrial [18] and residential sector [19]. Also, the privacy persevering nature of NILM has been exploited for human monitoring and assisted living [20]. Last, NILM was evaluated in terms of its ability to extract socio-economic information and consumer behaviour [21].

Several NILM surveys have been published within the last ten years, with each of them having a different focus on presenting the topic. In detail, [22] and [23] are the first ones describing the concept of NILM, but not covering the recent progress in utilizing deep learning methods in the NILM task. Similarly, [24] provides classification of appliances types and features, however it also does not include a review of the deep learning based NILM methods. Conversely, [25] mentions the rising significance of Deep Neural Network (DNN) based NILM approaches in 2016 and provides an thorough discussion on features, distinguishing steady-state from transient states. Furthermore, in [26] categorization of NILM methods is presented, while additionally the importance of benchmarking, thus choosing adequate performance metrics' and datasets, is pointed out. Latest reviews are [27] focusing on providing an overview on datasets, performance metrics and tool-kits for NILM as well as [28] providing a review of several NILM approaches that are using low-frequency DNN architectures.

The main contributions of the review are as follows: (i) A taxonomy of the NILM approaches is introduced, classifying them into three fundamentally different areas,

Manuscript received 11 January 2022; revised 24 April 2022; accepted 5 July 2022. Date of publication 8 July 2022; date of current version 22 December 2022. Paper no. TSG-00051-2022. *(Corresponding author: Pascal A. Schirmer.)*

Pascal A. Schirmer is with the School of Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, U.K., and also with the Department of Power Electronics, BMW AG, 80809 Munich, Germany (e-mail: p.schirmer@herts.ac.uk).

Iosif Mporas is with the School of Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, U.K. (e-mail: i.mporas@herts.ac.uk).

Digital Object Identifier 10.1109/TSG.2022.3189598

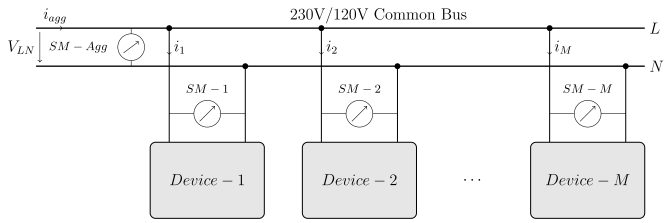

Fig. 1. Block diagram of the load monitoring task for M devices using $M+1$ SMs, one for each device and one for the aggregated signal.

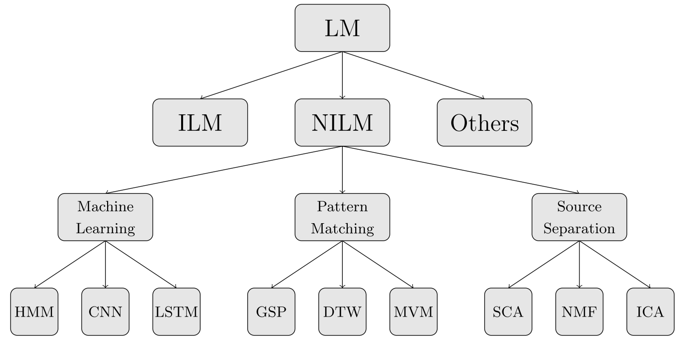

Fig. 2. Taxonomy of load monitoring techniques with focus on NILM.

and providing a framework for comparing different NILM techniques, while discussing the advantages and limitations of different approaches considering these three categories. (ii) Previously published results are grouped based on the experimental setup which allows direct comparison. (iii) The article is accompanied by a software framework implementation of the described NILM architectures[1] allowing the recalculation of them. (iv) The review article is the first review article to discuss the topic of transfer learning, introducing the effects of device signatures on the transfer capability. (v) A hardware realization is presented, including a standardized test scenario for future NILM architectures.

The remainder of this article is structured as follows: In Section II an overview of electric load monitoring is provided including a taxonomy of the most common NILM approaches as well as mathematical description of the NILM problem. In Section III generalized disaggregation architectures for data driven based NILM, pattern matching based NILM, and source separation based NILM are introduced. In Section IV data acquisition and publicly available datasets are presented, while in Section V performance metrics are presented and the performances of previously published NILM approaches are reported. The article is discussed in Section VI and concluded in Section VII, respectively.

II. Electric Load Monitoring

Load monitoring is the task of extracting electrical energy consumption at appliance level based on one or multiple measurements, in other words to detect the onsets (switch-on times) and offsets (switch-off times) of appliances. The concept of LM is illustrated in Fig. 1. As illustrated in Fig. 1 in LM the scenario of M devices being mounted on one common bus, e.g., one electrical phase, is considered. In detail, all devices share the same voltage, namely the line-to-neutral voltage (V_{LN}), while each device draws its individual current (i_m), where $i_{agg} = \sum_{m=1}^{M} i_m$ being the aggregated current. Similarly, the aggregated power (p_{agg}) and appliance power (p_m) are the product of current and voltage, while the device load and the aggregated load are monitored by Smart Meters (SMs). Fig. 2 presents the generalized taxonomy of the most widely used approaches for solving the LM problem and can be briefly split into NILM, ILM, and other methods.

The term ILM is used in the literature to define the measurement of electrical energy consumption with one sensor per device and thus there is no need for a disaggregation algorithm

[1] BaseNILM: https://github.com/pascme05/BaseNILM.

TABLE I
Feature Categorization According to Sampling Frequency

Low (0-100Hz)	**Mid (0.1-2kHz)**	**High (2-20kHz)**	**Ultra (>20kHz)**
· Active power · Reactive power · Phase angle · Power factor · V-I trajectory · Mean values · Variance · RMS values · Peak values · Frequency · $\delta/\delta\delta$ values	· $3^{rd}, 5^{th}, 7^{th}$ harmonics · DC component · Crest factor · Form factor	· Harmonic spectrum · HF conductance · HF susceptance · Wavelets · Slope time · Rise/fall time · THD · Transient energy	· Wavelet · EMI

in that case. However, the intrusive approach has the drawback of higher cost through wiring issues, data acquisition and wired/wireless communication caused by the increased number of sensors [29]. Furthermore, except being impractical due to the tens of energy meters that would have to be installed, the problematic retrofitting capability makes ILM an almost infeasible approach for existing households and buildings. In contrast to ILM, the term NILM is used to define the task of estimating the energy consumption of individual electrical appliances by using only the aggregated signal and applying disaggregation algorithms. These appliances are usually electric loads with different electrical properties (resistive, inductive, capacitive, or electronic [23], [30]) and working routines (always on, one-state, multi-state, non-linear and continuous [31]) operating at the same time. Other methods include approaches which are not exclusively limited to energy measures, but utilize other metering architectures (e.g., sound or light sensors) to identify electrical appliances [32]. The NILM approach can further be split into three main categories for appliance identification, namely Machine Learning (ML), Pattern Matching (PM), and single-channel Source Separation (SS). It should be noted, that previous works have also classified approaches using different taxonomies, e.g., according to sampling frequency describing steady-state and transient models [22], [23], which in this review has been integrated in the feature categorization shown in Table I. Furthermore, state-based and event-based approaches have been distinguished [26] as well, however this categorization is not applicable when considering appliances with continuous or non-linear behaviour due to their missing operational states.

As regards the NILM approaches based on machine learning, they are mainly based on the extraction of features, which will be used either to train a machine learning algorithm (e.g., Support Vector Machines (SVMs) [33], Artificial Neural Networks (ANNs) [34], Decision Trees (DTs) [35], Hidden Markov Models (HMMs) and their variants [36]–[38] and K-Nearest Neighbours (KNNs) [39]) or to define a set of rules or thresholds [40]–[43]. However, latest research in deep learning and big data has led to a significant increase in data-driven approaches using large scale datasets (e.g., AMPds [44]). Accordingly, approaches based on Convolutional Neural Networks (CNNs) [45]–[47], Recurrent Neural Networks (RNNs) [48], [49], Long Short Time Memory (LSTM) [48], [50] and Bidirectional LSTM (BiLSTM) [51] have been proposed in the literature, while in some papers denoising Auto Encoders (dAEs) [52] and Gate Recurrent Units (GRUs) [47] have been used as well. Additionally, latest research has focused on Generative Adversarial Networks (GANs) [53], [54] and bidirectional Transformers to incorporate self-attention mechanisms and to further improve the performance of the disaggregation algorithm [55]. Moreover, the appliance identification task mostly requires training of a classifier and hence can be categorized as a supervised or unsupervised approach (e.g., unsupervised machine learning algorithms as k-Means, fuzzy c-Means [56], Mean-Shift-Clustering [57] or dAEs utilizing a multi-environment event detector [58]).

In addition to the above mentioned machine-learning based NILM solutions, approaches using template matching have been proposed. More specifically, in [59] Dynamic Time Warping (DTW) was used to detect transient signatures for NILM and a weighted DTW was proposed and evaluated for different sampling frequencies. In [60] a hybrid detection approach utilizing Factorial Hidden Markov Models (FHMMs) and DTW-based iterative subsequence clustering was introduced for generating sub-sequences to refine initial estimates provided by the FHMM. In [61] load disaggregation was performed using subsequence searching by utilizing DTW and iteratively disaggregating one appliance at a time in order to decrease the overall energy consumption of the appliances after each iteration, i.e., with the appliance with the highest energy consumption being disaggregated first. In [62] a DTW-based pattern matching approach was proposed and its performance was compared to HMMs and DTs [63]. Furthermore, variants of DTW (elastic matching algorithms) have been evaluated and in [63] Multi-Variance-Matching (MVM) was found to be performing best. Moreover, Graph Signal Processing (GSP) has been used for energy disaggregation [64]–[67].

As regards NILM approaches based on single channel source separation, they formulate the NILM problem as an optimization procedure [68]. The assumption is based on the extraction of the individual power consumption signatures of the target appliances from the aggregated signal by utilizing constraints (e.g., sparseness [69], contextual figures [70] or probabilistic features [29]) on the optimization algorithm. The most widely used algorithms are Independent Component Analysis (ICA) [71], [72], Non-Negative Matrix Factorization (NMF), Non-Negative Tensor Factorization (NTF) [68], [73] and the Sparse Component Analysis (SCA) [74]–[76]. In contrast to appliance identification algorithms without source separation, these methods are unsupervised algorithms by their nature, but also require apriori knowledge due to the limitation of measuring only the aggregated signal, thus they can be considered as semi-unsupervised ones [29].

There are few other methods based on knowledge or rules that have been proposed in the NILM literature, e.g., utilizing frequency filters for detecting appliance signatures in the frequency domain or thresholds to detect appliance operation based on time domain signatures [40], [77]. Due to their lack of generality and their inferior performance compared to state-of-the-art ML, PM and SS techniques rule/knowledge-based methods have not been considered as an own category.

A. NILM Problem Formulation

NILM (energy disaggregation) can be formulated as the task of determining the power consumption on device level based on the measurements of one sensor, within a time window (frame). Specifically, for a set of M-1 known devices each consuming power $p_m \in \mathbb{R}^N$, with $1 \leq m \leq M-1$ and N being the total number of samples, the aggregated power $p_{agg} \in \mathbb{R}^N$, measured by the sensor will be:

$$p_{agg} = f(p_1, \ldots, p_{M-1}, e) = \sum_{m=1}^{M-1} p_m + e = \sum_{m=1}^{M} p_m \quad (1)$$

where $e = p_M \in \mathbb{R}^N$ is noise generated by one or more unknown devices (also referred to as ghost power [63]) and $f(\cdot)$ is the aggregation function. In NILM the goal is to find estimations $\hat{p}_m, \hat{e}$ of the power consumption of each device m using an estimation method $f^{-1}(\cdot)$ with minimal estimation error and $\hat{p}_M = \hat{e}$, i.e.,

$$\hat{P} = \left\{\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_{M-1}, \hat{e}\right\} = f^{-1}(p_{agg}) \quad (2)$$

As (2) is practically impossible to be solved analytically, most energy disaggregation methodologies are based on the segmentation of the aggregated signal into frames and estimation of the power consumption on device level within each frame utilizing either machine learning-based models, pattern matching or single-channel blind source separation. To provide more distinct information to the disaggregation function $f^{-1}(\cdot)$ usually each frame of the active power signal, p_{agg}^{τ}, is transferred to a higher dimensional feature representation, $X_{agg}^{\tau} = v(p_{agg}^{\tau})$, where $v(\cdot)$ is a feature mapping function. Accordingly, the disaggregation problem from (2) can be reformulated on frame level as in (3).

$$\hat{P}^{\tau} = \left\{\hat{p}_1^{\tau},\ \hat{p}_2^{\tau}, \ldots, \hat{p}_{M-1}^{\tau}, \hat{e}^{\tau}\right\} = f^{-1}\left(X_{agg}^{\tau}\right) \quad (3)$$

where $\hat{P}^{\tau}$ is the estimated power of the M-1 appliances and the noise for the τ^{th} frame.

B. Feature Extraction and Feature Selection

To efficiently disaggregate appliances from the aggregated consumption, the estimation of device signatures is essential. A device signature represents the operational nature of the

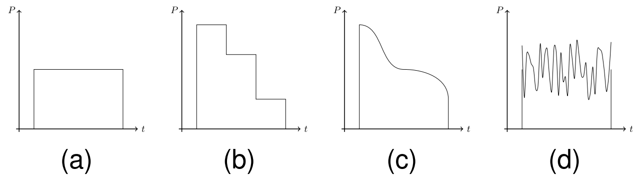

Fig. 3. Illustration of four device types with different operational behaviour. (a) One-State (b) Multi-State c) Continuous d) Non-Linear.

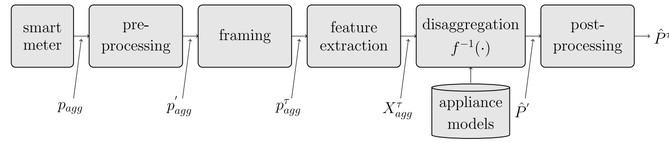

Fig. 4. Block diagram of a generalized data driven architecture for NILM.

device, and thus its electrical properties. According to [12] devices can be categorized into three categories:

- Single-State appliances that only have one state of operation, thus ON/OFF behaviour (e.g., lamps, kettles, etc.)
- Multi-State appliances that have a finite set of operating states and can be represented as a Finite State Machine (FSM) (e.g., washing machine, dryer, dishwasher, etc.)
- Continuous appliances [78] that have an infinite amount of power states and operate in a power range $\Delta p = [0, \ldots, p_{max}]$ (e.g., light dimmers, voltage source inverters, fully controlled air conditioning, etc.)

In addition to these three appliance categories, a fourth category has been introduced in [22] and is referred to as 'always-on'. These appliances stay always in their 'ON' state after initially being switched on (e.g., hard wired smoke detectors or hard wired phones), until they are switched off again. However, as these appliances usually have a low and constant power consumption, they can be treated as a constant offset to the disaggregation problem and are thus not further considered. Conversely, one other fourth category has been introduced in [79] and referred to as non-linear appliances characterized by their strong statistical variation of power consumption without fixed states (e.g., electronic devices like laptops, personal computers, LCD screens, etc.). An example of each of the four device categories is illustrated in Fig. 3.

To efficiently disaggregate the aggregated signal to appliance level the appliance signatures of one-dimensional power signal need to be transferred to a sequence of multi-dimensional feature vectors uniquely characterizing the appliance behaviour (except for proposed approaches performing disaggregation directly on the raw data [36]–[39]). The features extracted from the aggregated energy measurements strongly depend on the sampling frequency. Therefore, either low-frequency features (macroscopic also referred to steady-state) or high-frequency features (microscopic also referred to transient) can be used [80]. Steady-state features are mainly active and reactive power and a set of statistical low-frequency features computed from the active or reactive power (e.g., mean, median, variance or energy) [81], while also dimensionality reduction techniques like principal component analysis have been used to extract lower dimensional feature vectors and improve run time [82]. What regards transient features, they include current harmonics, Total Harmonic Distortion (THD) and transient energy [11], [40]. In this context, the importance of several low- and high-frequency features has been investigated. In detail, the approach in [79] evaluates statistical and electrical low-frequency features, while the approach in [83] evaluates high-frequency start-up events extracting 36 different features. Furthermore, [84] focuses on switching transients of up to 250 kHz, while [85] investigates the signatures of continuous and non-linear appliances especially considering switched mode power supplies, and [86], [87] evaluate wavelet based features for energy disaggregation. According to the above, features can briefly be grouped into low-, medium-, high- and ultra-high frequency ones as shown in Table I, where the most widely used features are listed.

III. Disaggregation Architectures

As described in Section II previously published NILM approaches can be grouped into three major categories, namely machine learning, pattern matching and source separation. In this Section a generalized architecture, which is based on previously published approaches in the literature, is presented aiming to serve as a baseline architecture for future research. It must be noted, that some approaches might not include all presented steps, for example some architectures do not have pre/post-processing steps or do not include a feature extraction stage when raw sample values are used directly, however for the sake of completeness of the generalized architecture these steps have been included. The generalized block diagram of the baseline NILM architecture commonly being utilized for all three approaches (ML, PM, and SS) is illustrated in Fig. 4. The disaggregation stage will differ in each of the three NILM approaches, with more detailed explanation of the disaggregation being provided in Sections III-A–III-C.

As illustrated in Fig. 4 the generalized NILM architecture consists of smart metering of the aggregated signal (p_{agg}), pre-processing (p'_{agg}), framing (p^{τ}_{agg}), feature extraction (X^{τ}_{agg}), disaggregation, and post-processing giving an estimation of the target appliances' power consumption for each frame ($\hat{P}^{\tau} = \{\hat{p}^{\tau}_1,\ \hat{p}^{\tau}_2, \ldots, \hat{p}^{\tau}_{M-1}, \hat{e}^{\tau}\}$). The smart meter acquires measurement samples of the active power p_{agg} (or of the raw current and voltage waveforms [88]) with a fixed sampling frequency f_s, resulting in a time series of data samples $p_{agg} \in \mathbb{R}^N : 1 \leq n \leq N$, where N denotes the total number of samples and n is the sample index. During pre-processing the raw samples p_{agg} are processed, removing outliers and faulty measurements, applying filtering operations to limit the frequency range or suppress certain harmonics [89], [90], resulting into the pre-processed time series of data p'_{agg}. At the framing stage the pre-processed measurements p'_{agg} are segmented in frames using a sliding time window of length L samples resulting into the frame blocked signal $p^{\tau}_{agg} \in \mathbb{R}^L : 1 \leq \tau \leq T$, where T denotes the total number of frames and τ is the frame index. Depending on the application, causal [91] approaches taking the L previous samples to form a frame and non-causal [92], [93] approaches

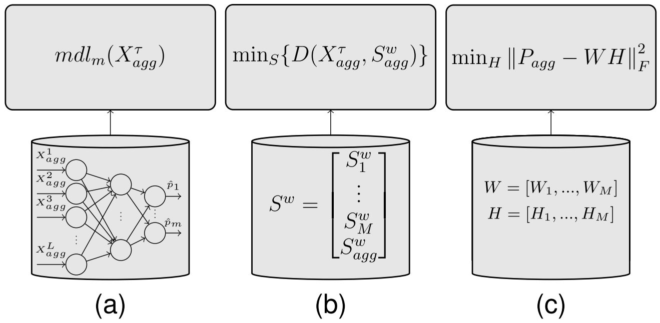

Fig. 5. Specific approaches for appliance disaggregation. (a) Machine Learning (b) Pattern Matching c) Source Separation.

choosing $\lfloor L/2 \rfloor$ to be the mid-point of the frame have been used to form frames, while the overlap between two consecutive frames can also vary, e.g., no overlap [94] resulting into a number of frames $T = \lfloor \frac{N}{L} \rfloor$ or one sample overlap [95], [96] resulting into a number of frames $T \approx N$. During feature extraction, for every frame $p^{\tau}_{agg} \in \mathbb{R}^L$ a feature vector $X^{\tau}_{agg} \in \mathbb{R}^F$ is calculated using one or more feature extraction algorithms, where F is the dimensionality of the feature vector, i.e., the number of calculated features. For an overview of possible NILM features, the reader is referred to Section II-B of this article. During the disaggregation step ML, PM, or SS is used to either determine the on/off states (classification or event-based NILM [27]) or the power consumption (regression or state-based NILM [27]) of the target appliance(s). An illustration of the modelling approaches in the disaggregation step for each of the three approaches is shown in Fig. 5 with descriptions in Sections III-A–III-C.

Finally, during the post-processing step several different approaches have been investigated to further increase performance, including state correction [97], filtering [98], or a second stage machine learning architecture [99], resulting in more accurate energy consumption estimations.

A. Machine Learning

In machine learning based approaches the relation between the aggregated signal X_{agg} and the correspondent appliance signals $P = \{p_1, p_2, \ldots, p_{M-1}, e\}$ is being learnt during the training of the ML model. Specifically, either a regression or a classification model is trained, with the model's output being the estimated appliance power consumption (regression) or the predicted appliance states (classification) respectively. Furthermore, architecture performing regression and classification combined have also been proposed [47]. The trained ML model is then used to estimate the appliance consumption using the incoming aggregated signal during testing/operation. As shown in Fig. 2 for machine learning-based approaches mainly HMM, LSTM and CNN based architectures have been evaluated in the literature, with research activities of HMM becoming less popular after 2016 (for details about HMMs in NILM the reader is referred to [26], [37], [38]). Conversely, both LSTM and CNN are currently widely used in the NILM task, with LSTM/RNN focusing on exploiting the temporal characteristics of the appliances signatures [51] using features capturing temporal information [100], while CNNs have been utilized as multi-dimensional feature extraction engines [88].

All three architectures have in common that they train a model based on the feature inputs X_{agg} and the ground-truth values of the appliance's consumptions P. In the case of Neural Network (NN) based approaches they are distinguished to one model per appliance used [88], i.e., $R_m(X_{agg}, p_m)$ where $R_m(\cdot)$ denotes the model of the m^{th} appliance and to one model commonly used for all appliances [101], i.e., $R(X_{agg}, P)$ where $R(\cdot)$ denotes a generalized model for all M appliances, with the two NN architectures only differing in the number of nodes in the output layer which affects performance and computational time [102], i.e., the single model offers lower performance but lower computational load comparing to the multiple device-dependent models. As an alternative of those two, also grouping of electrical appliances has been evaluated [103], [104]. For NN-based approaches the models $R_m(\cdot)$ are used as direct realization of the regression function $f^{-1}(\cdot)$ described in (3). Assuming the number neurons in the input layer being equal to the frame length L (or the feature size) the prediction of a feed-forward neural network with one hidden layer can be written as in (4) [105].

$$\hat{p}_m = R_m(X_{agg}) = \varphi\left(\sum_{j=1}^{J} w^{(2)}_{mj} h\left(\sum_{l=1}^{L} w^{(1)}_{jl} X^{i}_{agg}\right)\right) \tag{4}$$

where $\varphi(\cdot)$ is the activation function of the output layer and $h(\cdot)$ the activation function of the hidden layer, $\boldsymbol{w}$ are the weights of the first and second layer respectively and J are the number of neurons in the hidden layer. Similar formulations can be found for LSTM in [51] and for CNN in [106].

In general, two different optimization approaches have been used when trying to optimize the model performance. First, efforts have been made to find the optimal machine learning architecture, e.g., the number of layers, nodes and corresponding activation functions, using grid search or Bayesian optimization [107]. Second, the choice of input features was optimized to find accurate representations of the devices in the feature space and thus improving model performance [79], [108], [109]. Furthermore, for neural network-based approaches three different implementation techniques have been utilized, namely sequence-to-point learning (seq2point) [96], sequence-to-subsequence learning (seq2subseq) [110] and sequence-to-sequence learning (seq2seq) [111]. The concept of these three learning approaches is illustrated in Fig. 6.

While there are significant differences in terms of computation time due to different amounts of input frames, i.e., $T \approx N$ for seq2point, $T = \lfloor \frac{N}{K} \rfloor$ (with $1 < K < L$ and K denoting the length of the subsequence) for seq2subseq and $T = \lfloor \frac{N}{L} \rfloor$ for seq2seq, performance differences are less significant, with a clear advantage for seq2point based architectures due to the higher number of training instances.

B. Pattern Matching

As illustrated in Fig. 2 in pattern matching based approaches mainly GSP, DTW and MVM based architectures have been

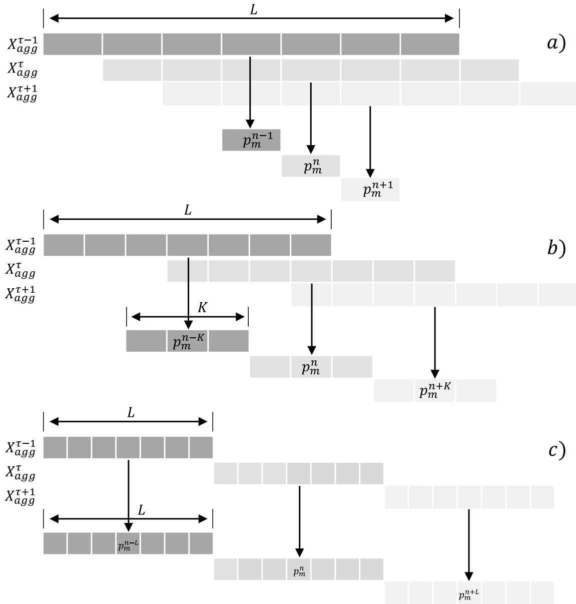

Fig. 6. Three most common learning approaches for the disaggregation stage of NN based learning strategies. (a) seq2point (b) seq2subseq c) seq2seq.

evaluated. In contrast to machine learning based methods, pattern matching based methods do not rely on model training, but on forming a dictionary matrix consisting of a set of reference signatures [63]. Depending on the approach, this dictionary might vary in its form, being a graph of vertices and edges detecting rising and falling edges as in GSP or being a matrix of characteristic reference signatures (see Fig. 5(b)) for each appliance as in DTW and MVM. The goal of these approaches is to find the closest match between an unknown observed signature, X^{τ}_{agg}, (i.e., a frame of the multidimensional aggregated feature vector) and the known signatures from a signature database $S : S_v,\ 1 \leq v \leq V$ including V known reference signatures for each of the M appliances as well as the aggregated signal [63]. The difference between the unknown signature and any reference signature in the reference signature database S is then calculated based on the local cost between signatures using a local cost function $D(\cdot)$ [59], e.g., Euclidean distance or Manhattan distance, an can be formulated as in (5):

$$D\left(X^{\tau}_{agg}(i), S^{v}_{agg}(j)\right) = \left\| X^{\tau}_{agg}(i) - S^{v}_{agg}(j)\right\|_p \qquad (5)$$

where $||\cdot||$ denotes the vector Lp-norm with $p \geq 1$ and i, j denote the sample indices of the unknown signature and the reference signature respectively. Based on the notation of unknown and reference signatures, the closest match between the unknown signature and the reference signature database can be written as in (6), while the estimates of the appliance power consumptions can be determined as in (7).

$$k(\tau) = \underset{S:S_v,\ 1\leq v\leq V}{argmin} \left\{ D\left(X^{\tau}_{agg}, S^{v}_{agg}\right)\right\} \qquad (6)$$

$$\hat{P}^{\tau} = \left\{ \frac{1}{L}\sum_{L} S^{1}_{k(\tau)}, \frac{1}{L}\sum_{L} S^{2}_{k(\tau)}, \ldots, \frac{1}{L}\sum_{L} S^{M}_{k(\tau)}\right\} \qquad (7)$$

where $k(\tau)$ is an index for selecting the closest signature out of the signature database. Eq. (6) and (7) represent the generic realization of the inverse of the aggregation function $f^{-1}(\cdot)$ as described in (3), while the specific realization of the optimization for finding the closest signatures depends on the approach, i.e., DTW or GSP. Moreover, other elastic matching methods have been used, like MVM [63] which (in contrast to DTW) allows skipping some elements of the target pattern.

C. Source Separation

As illustrated in Fig. 2 and described in Section II for source separation-based approaches mainly SCA and NMF/NTF based architectures have been evaluated. To turn (1) into the form of a source separation problem p_{agg} and p_m need to be reshaped into matrix form, such that matrices of size $P_{agg} \in \mathbb{R}^{L\times T}$ and $P_m \in \mathbb{R}^{L\times T}$ are created, where L is the frame-length and T is the number of frames. Using the above notation (1) can be rewritten as in (8).

$$P_{agg} = \sum_{m=1}^{M} P_m \qquad (8)$$

Based on the above notation, single-channel source separation tries to build a model for each signal P_m based on a set of bases (e.g., a dictionary) and activations (e.g., on/off states) during the training process, i.e.:

$$P_m \approx W_m H_m \qquad (9)$$

where $W_m \in \mathbb{R}^{L\times r}$ is a matrix of r bases vectors modelling the signal and $H_m \in \mathbb{R}^{r\times T}$ is the set of activations. Based on these dictionaries W_m and activations H_m the proposed NILM approaches try to find activations $\hat{H}_m$ and thus estimate the consumption values $\hat{P}_m$ for the aggregated signal P_{agg} with minimal estimation error as described in (10) and in (11).

$$\hat{H}_{1:M} = argmin_{H_{1:M}\geq 0} \left\| P_{agg} - [W_1, \ldots, W_M] \begin{bmatrix} H_1 \\ \vdots \\ H_M \end{bmatrix} \right\|_F^2 \qquad (10)$$

$$\hat{P}_m \approx W_m \hat{H}_m$$

$$s.t. \quad \sum_{m=1}^{M} \frac{1}{2}\left\| P_m - \hat{P}_m \right\|_F^2 \leq \epsilon \qquad (11)$$

where ϵ is an error margin. Eq. (10) and (11) represent the generic realization of the inverse of the aggregation function $f^{-1}(\cdot)$ as described in (3), while the specific realization of the optimization for finding the optimal dictionary and activation matrices depends on the approach, i.e., SCA, NMF or NTF.

IV. Datasets and Data Acquisition

To accurately evaluate the disaggregation architectures proposed in the literature as presented in Section III as well as to perform accurate comparison among NILM solutions, publicly available benchmark datasets, which are measured in close to real conditions using smart meters, have been developed.

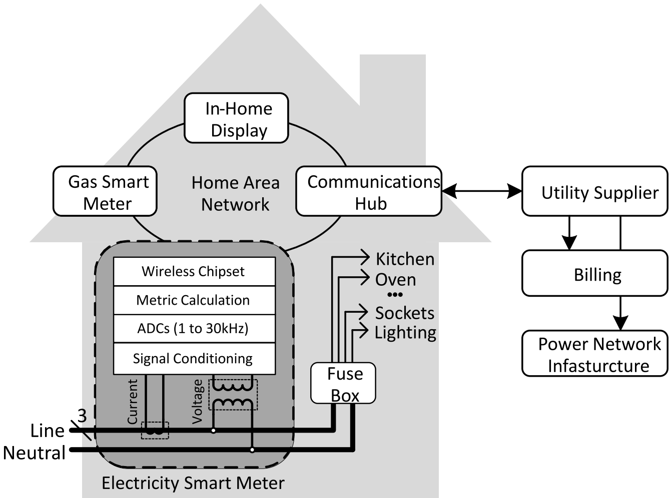

Fig. 7. Generalized smart meter architecture.

A. Smart Metering

To train and test the disaggregation architectures as described in Section III as well as to disaggregate households energy consumption, data acquired with high sampling frequencies in the order of seconds are needed [112]. Such data are usually acquired by the so-called smart meters. A smart meter is a device that can be used to measure the power consumption of a household or building, while being able to transmit the recorded data via a communication channel to the consumer's utility company or other energy monitoring/management services. The simplest smart meters may sample the energy consumption once per minute, whilst more advanced ones may sample at rates of up to 30 kHz [113]. Higher sampling frequencies are usually preferred, since they contain more detailed information about the energy consumption, however they increase the amount of data linearly and the cost of hardware exponentially [114]. With the sampling rate in the order of seconds data handling for several months/years becomes feasible and hardware costs are relatively low. Based on the measurements of raw current and voltage features, as described in Section II-B, can be directly calculated on the smart meter. A generalized architecture of a smart meter integrated in a consumer household is illustrated in Fig. 7.

Fig. 7 shows an example smart metering design, demonstrating the installation location at the home and the resulting aggregated power measurement that is made. To calculate the various features (listed in Section II-B) the voltage of the mains supply and the current being drawn from the supply are measured with the use of signal conditioning analogue electronics and high-speed Analogue to Digital Converters (ADCs) [115]. For the voltage measurement, isolated conversion from mains voltage to low voltage can be performed using a transformer. Alternatively, non-isolating direct voltage division is possible [84], [116]. For the current measurement, a current transformer, Hall effect sensor, or Rogowski coil can be used to provide isolated current to low signal voltage conversion [84], [116]. After ADC, the samples are processed to calculate the evaluation metrics, e.g., total energy consumed, maximum power, minimum power, etc., and then transmit them to the energy supplier or other service provider. The method of transmission depends on the country of installation. For example, in the U.K., mobile cellular networks are commonly used for the transmission, but other methods such as wireless ad hoc networks and ZigBee are also used. The transmission, storage, and access to data is heavily legislated, with user contracts being required for smart meter installation and operation in domestic environments [117]–[119].

B. Publicly Available Datasets

To ensure uniform comparison, standardization as well as comparability between different NILM methods, the definition of standardized performance metrics and the usage of public datasets is essential [120]. Further, the combination of public datasets and performance metrics, enables cross-comparison of proposed methods (e.g., filtering techniques, selected features, classifiers), aiming to advance the NILM task.

With rising interest in NILM techniques in the last 20 years, the monitoring of energy consumption for creation of publicly available datasets was only started within the last ten years [120]. Therefore, most existing datasets offer good representation of existing housing structures or less often of non-residential buildings, and thus are suitable for development and evaluation of new NILM methods. However, changes in electrical household appliances are apparent, e.g., more continuous devices (controlled air conditioning and switched power supplies) or strongly non-linear devices are getting more frequently found in households [121]. Therefore, current datasets might not be an accurate description of the housing structures of the next years. To give the reader an overview of the existing datasets that are suitable for NILM, Table II introduces the most widely used datasets with aggregated signal measurements and the corresponding ground truth, and is an extended/updated version of Table I from [27]. In detail, the columns in Table II list the year and country in which the database was recorded, the number of houses and target devices, the monitoring duration, the measurement approach, the measured features and the sampling resolution separately for the target devices and the aggregated data.

Table II provides a list of 29 databases with energy/power recordings from eleven countries, different sampling frequencies, number, and types of devices and monitoring duration lengths. Out of these 29 databases 18 databases (REDD [122], BLUED [123], ECO [124], UK-DALE [125], Dataport [126], Smart [127], RAE [128], iAWE [129], IHEPCDC [130], REFIT [131], AMPds [44], [132], COMBED [133], DRED [134], SustDataED [135], EEUD [136], SysD [137], LIFTED [138] and BLOND [116]) can be used for training NILM systems as they include both aggregated power consumption as well as consumption on device level, while five of them (RBSAM [139], HES [140], Tracebace [141], GREEND [142], ACS-F1/2 [81]) only include power consumption on appliance level hence are not suitable for testing NILM approaches. Furthermore, there is a set of six (PLAID [143], WHITED [144], LILAC [145], CREAM [146], HFED [147], COOLL [148]) additional

TABLE II
PUBLICLY AVAILABLE DATASETS AND THEIR PROPERTIES

Name	REF	Year	Country	#-houses	#-devices	Duration	Approach	Measured Features						Resolution	
								V	I	P	Q	S	O	Agg	App
REDD	[121]	2011	US	6	9-24	2-4 weeks	SB	✓	✓	✓	-	-	-	15 kHz	1/3 Hz
BLUED	[122]	2012	US	1	30	1 week	EB/SB	✓	✓	✓	✓	-	-	12 kHz	1 Hz
ECO	[123]	2014	Swiss	6	7-12	8 months	SB	✓	✓	✓	✓	-	φ	1 Hz	1 Hz
UK-DALE	[124]	2014	UK	5	5-40	655 days	SB	✓	✓	✓	✓	✓	-	16 kHz	6 sec
Dataport	[125]	2013	US	1400	70	4 years	SB	-	-	✓	-	-	-	1 min	1 min
Smart	[126]	2012	US	8	25	3 months	SB	-	-	✓	-	✓	-	1 Hz	1 Hz
RAE	[127]	2016	Canada	1	24	72 days	SB	✓	✓	✓	✓	✓	PF, f_s	1 Hz	1 Hz
iAWE	[128]	2013	India	1	33	74 days	SB	✓	✓	✓	✓	✓	f_s, φ	1 Hz	1 Hz
IHEPCDC	[129]	2013	France	1	9	4 years	SB	✓	✓	✓	✓	✓	-	1 min	1 min
REFIT	[130]	2014	UK	2	9	2 years	SB	-	-	✓	-	-	PF	8sec	8sec
AMPd	[44], [131]	2013	Canada	1	19	2 years	SB	✓	✓	✓	✓	✓	PF, Fs	1 min	1 min
COMBED	[132]	2014	India	1	200	1 month	SB	✓	-	✓	-	-	-	30 sec	30 sec
DRED	[133]	2015	Holland	1	9	6 months	SB	-	-	✓	-	-	-	1 Hz	1 Hz
SustDataED	[134]	2016	Portugal	-	17	10 days	EB/SB	✓	✓	✓	✓	-	-	12.8 kHz	0.5 Hz
EEUD	[135]	2017	Canada	23	-	1 year	SB	-	-	✓	-	-	-	1 min	1 min
SysD	[136]	2020	Austria	-	21	180 days	SB	-	-	✓	-	-	-	5 Hz	5 Hz
LIFTED	[137]	2019	US	1	15	1 week	SB	-	-	✓	-	-	-	50 Hz	50 Hz
BLOND	[115]	2018	Germany	1	53	50-230 days	SB	✓	✓	-	-	-	-	50-250 kHz	6.4-50 kHz
RBSAM	[138]	2014	US	101	6	27 months	EB	-	-	-	-	-	EE	-	15 min
HES	[139]	2012	UK	250	23	1-12 months	EB	-	-	-	-	-	EE	-	2-5 min
Tracebace	[140]	2012	Germany	158	43	6 years	SB	-	-	✓	-	-	-	-	0.1 Hz
GREEND	[141]	2014	Austria	8	9	6 months	SB	-	-	✓	-	-	-	-	1 Hz
ACS-F1/2	[81]	2011/13	Swiss	-	10 (15)	2 hours	SB	✓	✓	✓	✓	✓	φ	-	0.1 Hz
PLAID	[142]	2014	US	55	11	-	-	✓	✓	-	-	-	-	-	30 kHz
WHITED	[143]	2016	Germany	-	110	-	-	✓	✓	-	-	-	-	-	44.1 kHz
LILAC	[144]	2018	Germany	-	15	-	-	✓	✓	-	-	-	-	-	50 kHz
CREAM	[145]	2020	Germany	-	2	-	-	✓	✓	-	-	-	-	-	6.4 kHz
HFED	[146]	2015	India	-	15	-	-	-	-	-	-	-	EMI	-	5 MHz
COOLL	[147]	2016	France	-	12	-	-	✓	✓	-	-	-	-	-	100 kHz

SB='State-Based', EB='Event-Based', V='Voltage', I='Current', P='Active Power', Q='Reactive Power', S='Apparent Power', O='Optional', EE='Electric Energy', Agg='Aggregated', App='Appliance'

databases which consist of transient appliance signatures and can only be used for extracting features, create transient appliance models or design edge detectors. Especially, CREAM enables the extraction of internal operation states, which could potentially be used to improve the modelling of complex devices. Moreover, the databases can be categorized according to their high/low sampling frequencies with five of them, namely REDD, UK-DALE, BLUED, BLOND and SustDataED, having high frequency measures of raw current and voltage for the aggregated data. The BLOND database is, to the best of the authors' knowledge, the only database providing high frequency consumption measurements (above line frequency) for each individual appliance, hence providing high frequent ground truth data making it suitable for testing disaggregation approaches with different sampling frequencies. As regards the measured features all databases contain the active power as the main feature, with only exception the BLOND database, which contains voltage and current with high sampling frequency.

V. ENERGY DISAGGREGATION PERFORMANCE

To accurately compare NILM accuracy of any of the three major approaches presented in Section III it is crucial to evaluate results on a standardized setup as well as the same dataset(s) as described in Section IV, including the same pre-processing, selection of appliances, and data splits (data subset selection) as well as the same accuracy metrics.

A. Performance Metrics

As NILM was introduced roughly 30 years ago a wide variety of different performance metrics have been proposed in literature so far. Considering the latest trends in NILM only a few of these metrics have become widely accepted performance measures of recent NILM approaches and are listed in Table III. Specifically, Table III includes, next to the performance metric, a short description and the definition of variables for each metric. Furthermore, if the metric was introduced as part of the performance evaluation in NILM, the corresponding publication, where it was initially proposed, is reported. An extensive overview for all metrics used to measure performance in NILM can be found in [27].

The performance metrics in Table III can be split into the metrics for event-based NILM approaches (classification) including the classification Accuracy (ACC), the Inaccurate True-Positives (ITPs), Accurate True-Positives (ATP), the Precision (PR), the Recall (RC) and the F-Measure (F1) and the metrics for state-based approaches which perform energy disaggregation (regression) including Mean-Average-Error (MAE), Root-Mean-Square-Error (RMSE), Estimated Energy Fraction Index (EEFI), Actual Energy Fraction Index (AEFI) and Estimation Accuracy (E_{ACC}). The event-based NILM metrics, these can be further split into metrics for detection on device level (e.g., PR or RC), namely if a device is ON/OFF, and metrics for detection on state level to account for multi-state devices (e.g., ITP or ATP).

To evaluate event-based NILM approaches many researchers use ACC to measure how well an algorithm can predict ON/ OFF states of specific appliances. However, since there are appliances that either operate very rarely (e.g., washing machine, dishwasher or dryer) or are almost always ON (e.g., fridge, heater or lighting) using ACC as a metric can result into misleading performances [26], [149]. Therefore, the F-Measure is used to evaluate the prediction of these

TABLE III
MOST WIDELY USED PERFORMANCE METRIC TO COMPARE NILM SETUPS

Name	REF	Metric	Variables/Description
Event-based (Event Detection)			
Classification Accuracy	-	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$	Accuracy is the number of correctly assigned matches compared to all possible matches (True positives (TP), True negatives (TN), False Positives (FP), False Negatives (FN)).
Inaccurate True-Positives	[148]	$ITP = \frac{\sum_{t=1}^{T} \lvert\hat{p}_m^t - p_m^t\rvert}{N^m}$	To account for different number of states the inaccurate true positives normalize each appliance to its maximum number of states.
Accurate True-Positive	[148]	$ATP = 1 - ITP$	The accurate true positive are defined according to the inaccurate positives resulting to 100% ground-truth.
Precision	[148]	$PR = \frac{TP}{TP+FN}$	Precision is the proportion of relevant instances that were reported of being relevant against all the instances that were relevant.
Recall	[148]	$RC = \frac{TP}{TP+FN}$	Recall is the proportion of relevant instances that were reported as being relevant against the truly relevant instances.
F-Measure	-	$F_1 = 2 \cdot \frac{PR \cdot RC}{PR+RC}$	The F-measure is the harmonic mean of precision and recall.
State-based (Energy Disaggregation)			
Root-Mean-Square-Error	-	$RMSE = \sqrt{\frac{1}{T} \sum_{\tau=1}^{T} (\hat{p}_m^\tau - p_m^\tau)^2}$	The root-mean-square error gives the difference between the ground-truth and estimated power consumption.
Mean Absolute Error	-	$MAE = \frac{1}{T} \sum_{\tau=1}^{T} \lvert\hat{p}_m^\tau - p_m^\tau\rvert$	The disaggregation error is the normalized error between estimated power consumption and ground-truth.
Sum of Absolute Error	-	$SAE = \frac{\lvert E-\hat{E}\rvert}{E}$	The sum of the absolute error is the normalized difference between actual energy E and predicted energy $\hat{E}$.
Estimated Energy Fraction Index	[75]	$EEFI = \frac{\sum_{\tau=1}^{T} \hat{p}_m^\tau}{\sum_{\tau=1}^{T} \sum_{m=1}^{M} \hat{p}_m^\tau}$	The estimated energy fraction index provides the estimated fraction of energy consumed.
Actual Energy Fraction Index	[75]	$AEFI = \frac{\sum_{\tau=1}^{T} p_m^\tau}{\sum_{\tau=1}^{T} \sum_{m=1}^{M} p_m^\tau}$	The actual energy fraction index provides the actual fraction of energy consumed.
Estimation Accuracy	[121]	$E_{ACC} = 1 - \frac{\sum_{\tau=1}^{T} \sum_{m=1}^{M} \lvert\hat{p}_m^\tau - p_m^\tau\rvert}{2 \sum_{t=\tau}^{T} \sum_{m=1}^{M} \lvert p_m^\tau\rvert}$	The estimation accuracy is used to evaluate overall performance of the NILM disaggregation algorithm.
Estimation Accuracy (Device level)	[121]	$E_{ACC}^m = 1 - \frac{\sum_{\tau=1}^{T} \lvert\hat{p}_m^\tau - p_m^\tau\rvert}{2 \sum_{\tau=1}^{T} \lvert p_m^\tau\rvert}$	The estimation accuracy can be modified to measure individual device performance.
Disaggregation Error	[68]	$DE = \sum_{m=1}^{M} \frac{1}{2} \lVert p_m - \hat{p}_m\rVert_2^2$	Total disaggregation error as commonly used for source separation approaches.

appliances more accurately. However, the F-Measure is not rigorously defined for multi-state appliances and considers only ON/OFF states. Therefore, PR and RC have been redefined considering the number of states of each device through calculating the IPTs and ATPs, respectively. Moreover, to evaluate how well an NILM approach can disaggregate the aggregated signal on device level a set of state-based performance metrics have been proposed in the literature. RMSE is widely used in the NILM community taking the difference between the estimated power consumption $\hat{p}_m^\tau$ and the actual consumption p_m^τ for each time frame τ. However, RMSE makes the comparison between different datasets and approaches rather difficult as the measure is not normalized to the total energy being consumed [26]. To address this issue a set of normalized performance metrics have been proposed with E_{ACC} introduced in [122] being the most widely used NILM metric. Specifically, E_{ACC} measures how well the energy has been disaggregated (including a double counting for errors) and maps the difference between the estimated consumption and the ground-truth to a disaggregation accuracy making it a suitable choice for comparison of different NILM methods.

B. Benchmarking

To accurately estimate the performance of different NILM approaches performance evaluation must be carried out under conditions that can be easily reproduced by other researchers, thus publicly available databases and common performance metrics must be used as described in Section IV-B and Section V-A. In specific, comparability and reproducibility within NILM is assured through the wide acceptance of the datasets presented in Section IV-B and the accuracy metrics in Section V-A. Table IV lists the most recent NILM disaggregation results reported in the literature for the REDD, AMPds and UK-DALE database using the E_{ACC}, the SAE and the MAE metric. In detail, to assure one-to-one comparability the results have been clustered according to the three major NILM approaches, the utilized classifier, the dataset, as well as the evaluated appliances. It must be noted that Table IV only provides a selected sub-set of best performing approaches, which have been evaluated on the same experimental setup to assure comparability. A list of the low-frequency DNN approaches only, can be found in [28].

As can be seen in Table IV there are significant differences in terms of reported performance between the three main approaches (ML, PM, and SS) as well as within each category performances strongly vary according to the evaluated dataset and the selected appliances. However, not considering minor differences in the experimental setup and the exact choice of appliances the following trends can be observed. First, machine learning and pattern matching based approaches significantly outperform source separation-based approaches, with machine learning based approaches having average performances of 88.1% (80.7% for pattern matching) in terms of E_{ACC} when using the REDD dataset, while the source separation-based approaches average performance being 63.3% on similar setups. Second, pattern matching based approaches are more robust showing lower deviations when being evaluated on similar scenarios, e.g., for the REDD database reporting ±2.2%, while machine learning and source separation have deviations of ±5.0% and ±6.1%, respectively. Third, all approaches' performance is inversely proportional to the number of appliances to be disaggregated making NILM task challenging considering the increasing numbers of electrical appliances in nowadays and future households.

TABLE IV
BEST PERFORMING NILM APPROACHES FOR MACHINE LEARNING, PATTERN MATCHING AND SOURCE SEPARATION. IN THE COLUMN 'APPLIANCES', 'DEF' REFERS TO THE DEFERRABLE LOADS [37]

Approach	Classifier	Ref	Year	Dataset	Appliances	Metric	Perf.
Machine Learning	RNN	[149]	2020	REDD-1-4/6	all	E_{ACC}	83.2%
	RF	[150]	2020	REDD-1-4/6	all	E_{ACC}	80.5%
	HMM	[37]	2016	REDD-1-4/6	def	E_{ACC}	88.6%
	DNN	[151]	2013	REDD-1-4/6	def	E_{ACC}	87.4%
	CNN	[105]	2020	REDD-1-4/6	def	E_{ACC}	93.8%
	HMM	[98]	2016	REDD-2	def	E_{ACC}	86.4%
	RF	[99]	2020	REDD-2	def	E_{ACC}	93.4%
	HMM	[37]	2016	REDD-2	def	E_{ACC}	94.8%
	HMM	[98]	2013	REDD-2	def	E_{ACC}	84.8%
	CNN	[149]	2020	AMPds	all	E_{ACC}	89.3%
	CNN	[91]	2019	AMPds	all	E_{ACC}	88.4%
	BiLSTM	[106]	2019	AMPds	DR, DW, HO, WO	MAE	31.1
	FHMM	[36]	2016	AMPds	DR, DW, HO, WO	MAE	153.6
	GAN	[53]	2021	AMPds	DR, WO, HO	MAE	35.3
	GAN	[54]	2020	AMPds	DR, WO, HO	MAE	38.5
	LSTM	[106]	2019	AMPds	DR, WO, HO	MAE	38.6
	CNN	[153]	2018	AMPds	DR, WO, HO	MAE	63.4
	CNN	[88]	2021	UK-DALE-1	def	MAE	2.6
	CNN	[154]	2019	UK-DALE-1	def	MAE	3.6
Pattern Matching	matching	[17]	2019	REDD	all	E_{ACC}	84.0%
	matching	[155]	2016	REDD	all	E_{ACC}	81.0%
	clustering	[156]	2019	REDD	all	E_{ACC}	80.6%
	MVM	[63]	2020	REDD	all	E_{ACC}	80.9%
	GSP	[157]	2018	REDD-2	MW, KO, SO, FR, DW, LI	E_{ACC}	77.2%
	MVM	[63]	2020	REDD-2	MW, KO, SO, FR, DW, LI	E_{ACC}	80.6%
	DTW	[59]	2017	BLUED	3, 8, 11, 23, 27, 29, 31, 34, 47, 52, 55, 56, 58	F1	89.2%
	clustering	[158]	2017	BLUED	3, 8, 11, 23, 27, 29, 31, 34, 47, 52, 55, 56, 58	F1	83.3%
	DTW	[63]	2020	REDD-1,2,6	all	F1	89.2%
	DTW	[159]	2014	REDD-1,2,6	all	F1	86.2%
Source Separation	DDSC	[160]	2017	REDD-1-4/6	all	E_{ACC}	62.6%
	DDSC	[160]	2017	REDD-1-4/6	all	E_{ACC}	66.1%
	DDSC	[76]	2010	REDD-1-4/6	all	E_{ACC}	56.4%
	DDSC	[76]	2010	REDD-1-4/6	all	E_{ACC}	59.3%
	DL	[161]	2015	REDD-1-4/6	all	E_{ACC}	72.0%
	NTF	[73]	2014	REDD-1-4/6	all	DE	0.070
	DDSC	[73]	2014	REDD-1-4/6	all	DE	0.096
	NMF	[68]	2017	AMPds	all	DE	0.880
	DDSC	[68]	2017	AMPds	all	DE	1989
	Net	[68]	2017	AMPds	all	DE	1671
	NMF	[68]	2017	AMPds	def	SAE	0.150
	DDSC	[68]	2017	AMPds	def	SAE	0.303
	MNMF	[162]	2020	AMPds	def	SAE	0.116

VI. DISCUSSION

In the previous sections an up-to-date overview of existing NILM techniques, databases and performance evaluation has been presented. The following sub-sections provide a summary of the limitations for the three major NILM approaches. Furthermore, the topic of transfer learning as well as the challenges and future directions for hardware implementation of NILM are discussed.

A. Limitations

As discussed above each of the three approaches, namely the machine learning, the pattern matching and the single-channel source separation, has its own advantages. However, each of these three approaches also comes with limitations which are related to the nature of approach itself. First, machine learning based architectures suffer from high computational costs as well as the need for large volumes of training data to converge and give accurate disaggregation results. Even though this issue has been partially addressed in the state-of-the-art approaches as well as in the proposed architectures, computational load is still high, especially for the best performing architectures based on CNNs. Second, while pattern matching techniques do not rely on large datasets and do not have any trainable parameters, thus do not suffer from high computational complexity, they have high requirements in terms of storage for saving the set of reference signatures. Furthermore, the transferability capability of pattern matching based architectures is limited due to their fixed set of reference signatures. Third, while the NILM problem is intrinsically a single-channel source separation problem, the proposed architectures based on single channel source separation have not been able to come close to the performance of the machine learning or the pattern matching ones. Specifically, the lack of integration of temporal contextual information as well as the restrictions in terms of using multi-dimensional data are significant disadvantages of the single channel source separation approaches [68], [74]. A summary of the advantages and limitations of each method is provided in Table V.

TABLE V
COMPARISON OF ML, PM AND SS NILM ALGORITHMS

Properties	Accuracy	Runtime	Memory	Scaling	Transfer
ML	+	-	o	+	+
PM	o	o	-	o	o
SS	-	+	+	-	-

Performance: '+': better , 'o': average , '-': worse

Furthermore, limitations are not only of algorithmic nature, but often are related to the training and testing data or to the evaluation of the performance. Therefore, latest research works have been focusing on generating synthesized data to overcome the shortcomings of creating new datasets, i.e., the cost for hardware or the monitoring duration, using 1-D Wasserstein generative adversarial network [163] and data augmentation [164], while over-fitting through high frequent changes and distortions in datasets has been evaluated as well [165]. Moreover, the limitations in terms of comparability for NILM approaches has also been considered [166], proposing new accuracy metrics.

B. Transfer Learning

Additionally, to the investigation of possible solutions of the NILM problem on the same dataset, most recent evaluations have started investigating transferability approaches, where the training of the appliance models is done using one dataset, while testing is performed on other datasets (or the same dataset but on a different house). As transferability approaches are a relatively recent direction within the area of NILM, only few approaches have been presented in the literature [47], [94], [96], [110], [167], [168] and they investigate previously published architectures in terms of their transferability capability or evaluate their performance on cross domain learning, e.g., training on one appliance signature and testing on another [167]. However, to achieve high accuracies in transfer learning NILM, the architecture and input feature vectors must be specifically adapted to the NILM problem, to enable accurate cross domain learning. In specific, the architecture and the features should be able to capture the general behaviour/characteristics of appliance types, as models from different manufactures might present differently scaled signatures. Three different scaling operations must be considered and are illustrated in Fig. 8.

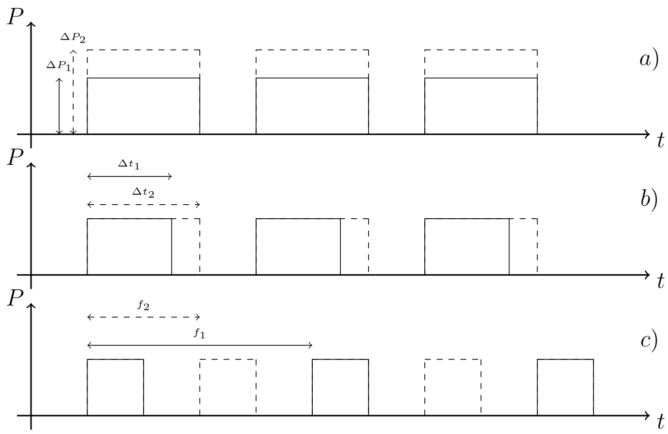

Fig. 8. Three differences in device signatures during transfer learning. (a) Y-scaling (b) X-scaling c) frequency-scaling.

TABLE VI
COMPARISON OF THE BEST PERFORMING NILM TRANSFER APPROACHES

Approach	Classifier	Ref	Year	Dataset	Appliances	Metric	Perf.
Transfer Learning	CNN	[96]	2018	REDD	MW,FR,DW,WM	MAE	23.69
	CNN	[47]	2019	REDD	MW,FR,DW	MAE	65.84
	GRU	[47]	2019	REDD	MW,FR,DW	MAE	85.00
	CNN	[167]	2019	REFIT	def	MAE	13.72
	CNN	[47]	2019	REFIT	MW,FR,DW,WM	MAE	49.70
	GRU	[47]	2019	REFIT	MW,FR,DW,WM	MAE	51.81
	CNN	[96]	2018	UK-DALE	def	MAE	15.47
	GAN	[109]	2020	UK-DALE	def	MAE	7.84
	HMM	[170]	2015	UK-DALE	FR, DW, WM	MAE	81.3
	CNN	[96]	2018	UK-DALE	FR, DW, WM	MAE	20.4
	CNN	[168]	2020	UK-DALE	FR, DW, WM	MAE	19.5

As can be seen in Fig. 8 three different scaling operations might appear when considering different models of the same device. First, scaling in Y-direction (amplitudes), i.e., different models of the same appliance have the same signature shape, but with a different maximum power consumption, which might be due to different sizes. Second, different models of the same appliance might have different operational times, i.e., the signature shape and the maximum power consumption might be the same, but the operational time might be shorter or longer (X-Scaling). Third, the frequency of operation might be different due to external influences, e.g., a fridge is switching on more frequently due to a higher ambient temperature or a washing machine is used at different intervals (Frequency-Scaling). New transfer learning architectures have to find those invariant features that capture and integrate these differences in model-independent device signatures, i.e., either through normalization and using frequency signatures in case of amplitude- scaling [17], through incorporating temporal information in case of temporal-scaling [93] or through incorporating state probabilities in case of frequency-scaling [169]. The state-of-the-art results are tabulated in Table VI.

As tabulated in Table VI most transferability approaches are based on CNN models to learn the activation profiles of each appliance within the convolutional layers as discussed in [167]. Furthermore, the work presented in [47] indicates that CNNs outperform architectures such as GRUs but result into longer training times due to a higher number of trainable parameters. Moreover, the approach presented in [110] proposes conditional GANs with seq2subseq learning overcoming convergence issues for long input sequences. These approaches have been evaluated on three/four appliances of the REDD dataset and report MAE scores ranging from 20.2 up to 85.0, which are still significantly lower than the performance scores of non-transfer learning approaches.

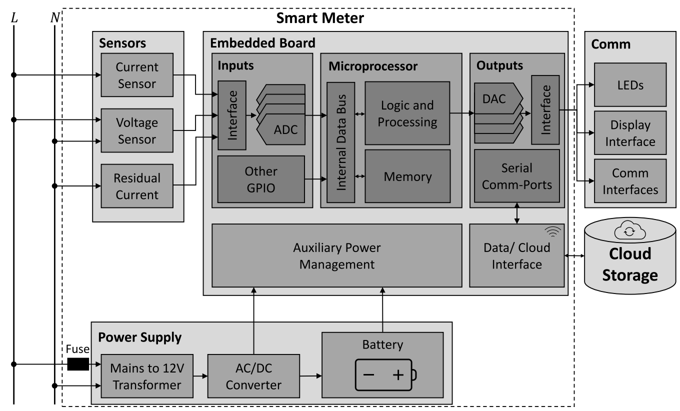

Fig. 9. A generalized smart meter architecture on block diagram level.

C. Hardware Realization

Up-to now NILM algorithms have been evaluated mostly on publicly available datasets as described in Section IV-B, while transfer learning approaches have been evaluated using different datasets as described in Section VI-B. However, only very few approaches [41], [106], [112], [171], [172] have been evaluated on actual hardware and have been tested using hardware implementations of smart meter architectures [40], [41], [106], [172]. A generalized hardware architecture of a smart meter is illustrated in Fig. 9.

As can be seen in Fig. 9 the smart meter consists of a set of sensors for measuring current and voltage, an evaluation board including signal conditioning, signal processing, data storage, signal outputs, as well as a power supply. A communication interface allows the smart meter to exchange information locally and through the cloud.

Due to testing of the NILM architecture on hardware, several additional complications arise. First, hardware implantations are by default based on transfer learning, as models need to be trained offline using collected dataset(s), while being tested on hardware afterwards [106]. Second, microprocessors in smart meters have significant restrictions in terms of computational power as well as on RAM and ROM for calculating features in real time and storing the disaggregated signatures [37]. Third, measurement errors due to the hardware circuitry, e.g., precision of current and voltage sensors or influence of switching's on the same power line as well as electromagnetic interference [85] appear. To address the above points, future proposed NILM architectures should be tested on a standardized test scenario, including:

- A standardized smart meter to assure to have the same computational power
- Wiring diagram (including interface definition, cable cross-section and cable length)

- The appliances to be tested (defined by supplier number)
- A standardized operational sequence of the appliances

VII. Conclusion

Non-intrusive load monitoring is an effective approach to monitor device operation and power consumption by only observing the aggregated power consumption signal. Compared to intrusive load monitoring, it has the advantage of reduced costs for hardware through the installation of only one sensor at the inlet of each household. Based on the disaggregated power consumption signals optimization approaches, i.e., load scheduling for reduction of energy cost or grid distortion as well as fault detection of electrical devices, can be developed. This review introduced a taxonomy of the NILM approaches, namely the ML, PM and SS ones, while an overview about NILM approaches, corresponding architectures, available datasets, and state-of-the-art benchmarking are discussed with respect to this taxonomy. It was demonstrated, that NILM approaches based on ML and PM have reached very high performances when being evaluated on the same data domain. Further studies, however, might be necessary to gain more knowledge about extraction of invariant features to improve the transfer capability of NILM architectures.

Four future directions and open challenges in energy disaggregation are discussed below. First, NILM is a time-series problem thus architectures and approaches that have considered time series information, either in the model or at the features, have shown significant performance advantages over other methods. Due to the increasing number of appliances and the continuous nature of new energy efficient appliances, including variable power electronics, it is crucial to exploit the time-series information to achieve higher NILM performances. Second, as transfer learning for NILM is expected to have significant impact as it is crucial for real-world implementation of NILM systems the extraction for time and appliance invariant features should be focused on reaching comparable performances of transfer and non-transfer NILM approaches. Furthermore, to quantitatively measure if the performance improvement for a transferability approach is actually related to the improved capturing of invariant appliance signatures or is attributed to a general improvement [173]. Third, the impact of NILM can only be fully evaluated if hardware implementations of NILM approaches in real world applications or hardware in the loop systems are also evaluated. Therefore, hardware oriented NILM approaches, e.g., AI on the edge, should be evaluated in more detail, considering issues like run-time, data transfer and security issues. Last, the authors hope that the provided baseline software implementation will promote new research in the area of NILM.

References

[1] B. Yu, Y. Tian, and J. Zhang, "A dynamic active energy demand management system for evaluating the effect of policy scheme on household energy consumption behavior," *Energy*, vol. 91, pp. 491–506, Nov. 2015.

[2] O. Elma and U. S. Selamoğullar, "A survey of a residential load profile for demand side management systems," in *Proc. 5th IEEE Int. Conf. Smart Energy Grid Eng. (SEGE)*, Aug. 2017 pp. 85–89.

[3] "Energy Statistics—An Overview." Eurostat. 2018. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_statistics_-_an_overview#Final_energy_consumption

[4] D. He, W. Lin, N. Liu, R. G. Harley, and T. G. Habetler, "Incorporating non-intrusive load monitoring into building level demand response," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1870–1877, Dec. 2013.

[5] D. Lee and C.-C. Cheng, "Energy savings by energy management systems: A review," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 760–777, Apr. 2016.

[6] M. Zeifman and K. Roth, "Viterbi algorithm with sparse transitions (VAST) for nonintrusive load monitoring," in *Proc. IEEE Symp. Comput. Intell. Appl. Smart Grid*, 2011, pp. 1–8.

[7] H. Çimen, N. Çetinkaya, J. C. Vasquez, and J. M. Guerrero, "A microgrid energy management system based on non-intrusive load monitoring via multitask learning," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 977–987, Mar. 2021.

[8] V. Indragandhi, R. Logesh, V. Subramaniyaswamy, V. Vijayakumar, P. Siarry, and L. Uden, "Multi-objective optimization and energy management in renewable based AC/DC microgrid," *Comput. Elect. Eng.*, vol. 70, pp. 179–198, Aug. 2018.

[9] A. Chiş, J. Rajasekharan, J. Lundén, and V. Koivunen, "Demand response for renewable energy integration and load balancing in smart grid communities," in *Proc. 24th Eur. Signal Process. Conf. (EUSIPCO)*, 2016, pp. 1423–1427.

[10] J. Kelly and W. Knottenbelt, "Does disaggregated electricity feedback reduce domestic electricity consumption? A systematic review of the literature," in *Proc. 3rd Int. NILM Workshop*, 2016. [Online]. Available: http://arxiv.org/pdf/1605.00962v2

[11] M. N. Meziane, P. Ravier, G. Lamarque, K. Abed-Meraim, J.-C. Le Bunetel, and Y. Raingeaud, "Modeling and estimation of transient current signals," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, 2015, pp. 1960–1964.

[12] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[13] J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. Reynolds, and S. Patel, "Disaggregated end-use energy sensing for the smart grid," *IEEE Pervasive Comput.*, vol. 10, no. 1, pp. 28–39, Jan.–Mar. 2011.

[14] Y. Du, L. Du, B. Lu, R. Harley, and T. Habetler, "A review of identification and monitoring methods for electric loads in commercial and residential buildings," in *Proc. IEEE Energy Convers. Congr. Expo (ECCE)*, Atlanta, GA, USA, Sep. 2010, pp. 4527–4533.

[15] D. García-Pérez, D. Pérez-López, I. Díaz-Blanco, A. González-Muñiz, M. Domínguez-González, and A. A. C. Vega, "Fully-convolutional denoising auto-encoders for NILM in large non-residential buildings," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2722–2731, May 2021.

[16] A. Ruano, A. Hernandez, J. Ureña, M. Ruano, and J. Garcia, "NILM techniques for intelligent home energy management and ambient assisted living: A review," *Energies*, vol. 12, no. 11, p. 2203, 2019.

[17] S. Welikala, C. Dinesh, M. P. B. Ekanayake, R. I. Godaliyadda, and J. Ekanayake, "Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 448–461, Jan. 2019.

[18] A. Aboulian *et al.*, "NILM dashboard: A power system monitor for electromechanical equipment diagnostics," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1405–1414, Mar. 2019.

[19] A. Rafati, H. R. Shaker, and S. Ghahghahzadeh, "Fault detection and efficiency assessment for HVAC systems using non-intrusive load monitoring: A review," *Energies*, vol. 15, no. 1, p. 341, 2022.

[20] Á. Hernández, A. Ruano, J. Ureña, M. G. Ruano, and J. J. Garcia, "Applications of NILM techniques to energy management and assisted living," *IFAC-PapersOnLine*, vol. 52, no. 11, pp. 164–171, 2019.

[21] P. A. Schirmer and I. Mporas, "On the non-intrusive extraction of residents' privacy- and security-sensitive information from energy smart meters," *Neural Comput. Appl.*, to be published.

[22] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 76–84, Feb. 2011.

[23] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838–16866, 2012.

[24] A. Ridi, C. Gisler, and J. Hennebert, "A survey on intrusive load monitoring for appliance recognition," in *Proc. 22nd Int. Conf. Pattern Recognit.*, 2014, pp. 3702–3707.

[25] C. Klemenjak and P. Goldsborough, "Non-intrusive load monitoring: A review and outlook," 2016, *arXiv:1610.01191*.

[26] A. Faustine, N. H. Mvungi, S. Kaijage, and K. Michael, "A survey on non-intrusive load monitoring methodies and techniques for energy disaggregation problem," 2017, *arXiv:1703.00785*.

[27] L. Pereira and N. Nunes, "Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—A review," *Wiley Interdiscipl. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 6, p. e1265, 2018.

[28] P. Huber, A. Calatroni, A. Rumsch, and A. Paice, "Review on deep neural networks applied to low-frequency NILM," *Energies*, vol. 14, no. 9, p. 2390, 2021.

[29] D. Egarter, V. P. Bhuvana, and W. Elmenreich, "PALDi: Online load disaggregation via particle filtering," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 2, pp. 467–477, Feb. 2015.

[30] Z. Wang and G. Zheng, "Residential appliances identification and monitoring by a nonintrusive method," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 80–92, Mar. 2012.

[31] S. R. Shaw, S. B. Leeb, L. K. Norford, and R. W. Cox, "Nonintrusive load monitoring and diagnostics in power systems," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 7, pp. 1445–1454, Jul. 2008.

[32] Y. F. Wong, Y. A. Şekercioğlu, T. Drummond, and V. S. Wong, "Recent approaches to non-intrusive load monitoring techniques in residential settings," in *Proc. IEEE Symp. Comput. Intell. Appl. Smart Grid (CIASG)*, Singapore, 2013, pp. 73–79.

[33] T. Hassan, F. Javed, and N. Arshad, "An empirical investigation of V-I trajectory based load signatures for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 870–878, Mar. 2014.

[34] Y.-H. Lin and M.-S. Tsai, "An advanced home energy management system facilitated by nonintrusive load monitoring with automated multiobjective power scheduling," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1839–1851, Jul. 2015.

[35] P. Bilski and W. Winiecki, "Generalized algorithm for the non-intrusive identification of electrical appliances in the household," in *Proc. IEEE 9th Int. Conf. Intell. Data Aquisition Adv. Comput. Syst. Technol. Appl. (IDAACS)*, Bucharest, Romania, Sep. 2017, pp. 730–735.

[36] L. Mauch and B. Yang, "A novel DNN-HMM-based approach for extracting single loads from aggregate power signals," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Shanghai, China, Mar. 2016, pp. 2384–2388.

[37] S. Makonin, F. Popowich, I. V. Bajić, B. Gill, and L. Bartram, "Exploiting HMM sparsity to perform online real-time nonintrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2575–2585, Nov. 2016.

[38] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," in *Proc. 15th Int. Conf. Artif. Intell. Stat.*, vol. 22. La Palma, Islands, 2012, pp. 1472–1482. [Online]. Available: http://proceedings.mlr.press/v22/zico12.html

[39] A. Ridi, C. Gisler, and J. Hennebert, "Automatic identification of electrical appliances using smart plugs," in *Proc. 8th Int. Workshop Syst. Signal Process. Appl. (WoSSPA)*, Algiers, Algeria, May 2013, pp. 301–305.

[40] P. Bilski and W. Winiecki, "The rule-based method for the non-intrusive electrical appliances identification," in *Proc. IEEE 8th Int. Conf. Intell. Data Acquisition Adv. Comput. Syst. Technol. Appl. (IDAACS)*, Warsaw, Poland, Sep. 2015, pp. 220–225.

[41] T. T. Tran *et al.*, "Identification of in-home appliances through analysis of current consumption," in *Proc. ACM IMCOM*, Danang, Vietnam, 2007, pp. 1–5.

[42] Y. Zhou, Q. Zhai, X. Li, and Y. Yang, "A method for recognizing electrical appliances based on active load demand in a house/office environment," in *Proc. Chin. Autom. Congr. (CAC)*, Jinan, China, Oct. 2017, pp. 3584–3589.

[43] R. S. Prasad and S. Semwal, "A simplified new procedure for identification of appliances in smart meter applications," in *Proc. IEEE Int. Syst. Conf. (SysCon)*, Orlando, FL, USA, 2013, pp. 339–344.

[44] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajić, "AMPds: A public dataset for load disaggregation and eco-feedback research," in *Proc. IEEE Elect. Power Energy Conf.*, 2013, pp. 1–6.

[45] K. S. Barsim and B. Yang, "On the feasibility of generic deep disaggregation for single-load extraction," 2018, *arXiv:1802.02139*.

[46] X. Wu, X. Han, and K. X. Liang, "Event-based non-intrusive load identification algorithm for residential loads combined with underdetermined decomposition and characteristic filtering," *IET Gener. Transm. Distrib.* vol. 13, no. 1, pp. 99–107, 2019.

[47] D. Murray, L. Stankovic, V. Stankovic, S. Lulic, and S. Sladojevic, "Transferability of neural network approaches for low-rate energy disaggregation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2019, pp. 8330–8334.

[48] W. He and Y. Chai, "An empirical study on energy disaggregation via deep learning," in *Proc. 2nd Int. Conf. Artif. Intell. Ind. Eng. (AIIE)*, Paris, France, 2016, pp. 338–342.

[49] I. Çavdar and V. Faryad, "New design of a supervised energy disaggregation model based on the deep neural network for a smart grid," *Energies*, vol. 12, no. 7, p. 1217, 2019.

[50] L. Mauch and B. Yang, "A new approach for supervised power disaggregation by using a deep recurrent LSTM network," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Orlando, FL, USA, Dec. 2015, pp. 63–67.

[51] M. Kaselimi, N. Doulamis, A. Voulodimos, E. Protopapadakis, and A. Doulamis, "Context aware energy disaggregation using adaptive bidirectional LSTM models," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3054–3067, Jul. 2020.

[52] F. C. C. Garcia, C. M. C. Creayla, and E. Q. B. Macabebe, "Development of an intelligent system for smart home energy disaggregation using stacked denoising Autoencoders," *Procedia Comput. Sci.*, vol. 105, pp. 248–255, Jan. 2017.

[53] M. Kaselimi, N. Doulamis, A. Voulodimos, A. Doulamis, and E. Protopapadakis, "EnerGAN++: A generative adversarial gated recurrent network for robust energy disaggregation," *IEEE Open J. Signal Process.*, vol. 2, pp. 1–16, 2021.

[54] M. Kaselimi, A. Voulodimos, E. Protopapadakis, N. Doulamis, and A. Doulamis, "EnerGAN: A generative adversarial network for energy disaggregation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2020, pp. 1578–1582.

[55] Z. Yue, C. R. Witzig, D. Jorde, and H.-A. Jacobsen, "BERT4NILM: A bidirectional transformer model for non-intrusive load monitoring," in *Proc. 5th Int. Workshop Non-Intrusive Load Monit.*, 2020, pp. 89–93.

[56] T. Y. Ji, L. Liu, T. S. Wang, W. B. Lin, M. S. Li, and Q. H. Wu, "Non-intrusive load monitoring using additive factorial approximate maximum a posteriori based on iterative fuzzy c -means," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6667–6677, Nov. 2019.

[57] M. Dong, P. C. M. Meira, W. Xu, and C. Y. Chung, "Non-intrusive signature extraction for major residential loads," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1421–1430, Sep. 2013.

[58] D. Jorde, M. Kahl, and H.-A. Jacobsen, "MEED: An unsupervised multi-environment event detector for non-intrusive load monitoring," in *Proc. IEEE Int. Conf. Commun. Control Comput. Technol. Smart Grids (SmartGridComm)*, 2019, pp. 1–6.

[59] B. Liu, W. Luan, and Y. Yu, "Dynamic time warping based non-intrusive load transient identification," *Appl. Energy*, vol. 195, pp. 634–645, Jun. 2017.

[60] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. E. Rizzoli, "A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring," *Appl. Energy*, vol. 185, pp. 331–344, Jan. 2017.

[61] H. Wang and W. Yang, "An iterative load disaggregation approach based on appliance consumption pattern," *Appl. Sci.*, vol. 8, no. 4, p. 542, 2018.

[62] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Power disaggregation for low-sampling rate data," in *Proc. 2nd Int. Non-Intrusive Appl. Load Monit. Workshop*, vol. 1. Austin, TX, USA, 2014, p. F1.

[63] P. A. Schirmer, I. Mporas, and M. Paraskevas, "Energy disaggregation using elastic matching algorithms," *Entropy*, vol. 22, no. 1, p. 71, 2020.

[64] A. Faustine, L. Pereira, and C. Klemenjak, "Adaptive weighted recurrence graphs for appliance recognition in non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 398–406, Jan. 2021.

[65] B. Zhao, L. Stankovic, and V. Stankovic, "On a training-less solution for non-intrusive appliance load monitoring using graph signal processing," *IEEE Access*, vol. 4, pp. 1784–1799, 2016.

[66] V. Stankovic, J. Liao, and L. Stankovic, "A graph-based signal processing approach for low-rate energy disaggregation," in *Proc. IEEE Symp. Comput. Intell. Eng. Solutions (CIES)*, Orlando, FL, USA, Dec. 2014, pp. 81–87.

[67] K. He, L. Stankovic, J. Liao, and V. Stankovic, "Non-intrusive load disaggregation using graph signal processing," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1739–1747, May 2018.

[68] A. Rahimpour, H. Qi, D. Fugate, and T. Kuruganti, "Non-intrusive energy disaggregation using non-negative matrix factorization with sum-to-k constraint," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4430–4441, Nov. 2017.

[69] N. Pathak, N. Roy, and A. Biswas, "Iterative signal separation assisted energy disaggregation," in *Proc. 6th Int. Green Sustain. Comput. Conf.*, Las Vegas NV, USA, Dec. 2015, pp. 1–8.

[70] M. Wytock and J. Z. Kolter, "Contextually supervised source separation with application to energy disaggregation," in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014, pp. 486–492.

[71] Y. Zhu and S. Lu, "Load profile disaggregation by blind source separation: A wavelets-assisted independent component analysis approach," in *Proc. IEEE PES General Meeting Conf. Expo.*, National Harbor, MD, USA, Jul. 2014, pp. 1–5.

[72] S. Semwal, D. Joshi, R. S. Prasad, and D. Raveendhra, "The practicability of ICA in home appliances load profile separation using current signature: A preliminary study," in *Proc. Int. Conf. Power Energy Control (ICPEC)*, Dindigul, India, Feb. 2013, pp. 756–759.

[73] M. Figueiredo, B. Ribeiro, and A. de Almeida, "Electrical signal source separation via nonnegative tensor factorization using on site measurements in a smart home," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 2, pp. 364–373, Feb. 2014.

[74] S. Singh and A. Majumdar, "Analysis co-sparse coding for energy disaggregation," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 462–470, Jan. 2019.

[75] D. Piga, A. Cominola, M. Giuliani, A. Castelletti, and A. E. Rizzoli, "Sparse optimization for automated energy end use disaggregation," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 3, pp. 1044–1051, May 2016.

[76] J. Z. Kolter, S. Batra, and A. Y. Ng, "Energy disaggregation via discriminative sparse coding," in *Proc. 23rd Int. Conf. Neural Inf. Process. Syst.*, 2010, pp. 1153–1161. [Online]. Available: http://papers.nips.cc/paper/4054-energy-disaggregation-via-discriminative-sparse-coding

[77] L. K. Norford and S. B. Leeb, "Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms," *Energy Build.*, vol. 24, no. 1, pp. 51–64, 1996.

[78] W. Wichakool, Z. Remscrim, U. A. Orji, and S. B. Leeb, "Smart metering of variable power loads," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 189–198, Jan. 2015.

[79] P. A. Schirmer and I. Mporas, "Statistical and electrical features evaluation for electrical appliances energy disaggregation," *Sustainability*, vol. 11, no. 11, p. 3222, 2019.

[80] H.-H. Chang, K.-L. Lian, Y.-C. Su, and W.-J. Lee, "Power-spectrum-based wavelet transform for nonintrusive demand monitoring and load identification," *IEEE Trans. Ind. Appl.*, vol. 50, no. 3, pp. 2081–2089, May/Jun. 2014.

[81] C. Gisler, A. Ridi, D. Zufferey, O. A. Khaled, and J. Hennebert, "Appliance consumption signature database and recognition test protocols," in *Proc. 8th Int. Workshop Syst. Signal Process. Appl. (WoSSPA)*, Algiers, Algeria, May 2013, pp. 336–341.

[82] R. Machlev, D. Tolkachov, Y. Levron, and Y. Beck, "Dimension reduction for NILM classification based on principle component analysis," *Electr. Power Syst. Res.*, vol. 187, Oct. 2020, Art. no. 106459.

[83] M. Kahl, A. Ul Haq, T. Kriechbaumer, and H.-A. Jacobsen, "A comprehensive feature study for appliance recognition on high frequency energy data," in *Proc. 8th Int. Conf. Future Energy Syst.*, 2017, pp. 121–131.

[84] A. U. Haq, T. Kriechbaumer, M. Kahl, and H.-A. Jacobsen, "CLEAR—A circuit level electric appliance radar for the electric cabinet," in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, 2017, pp. 1130–1135.

[85] M. Kahl, T. Kriechbaumer, D. Jorde, A. Ul Haq, and H.-A. Jacobsen, "Appliance event detection—A multivariate, supervised classification approach," in *Proc. 10th ACM Int. Conf. Future Energy Syst.*, 2019, pp. 373–375.

[86] J. M. Gillis and W. G. Morsi, "Non-intrusive load monitoring using semi-supervised machine learning and wavelet design," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2648–2655, Nov. 2017.

[87] J. M. Gillis, S. M. Alshareef, and W. G. Morsi, "Nonintrusive load monitoring using wavelet design and machine learning," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 320–328, Jan. 2016.

[88] P. A. Schirmer and I. Mporas, "Double Fourier integral analysis based convolutional neural network regression for high-frequency energy disaggregation," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 3, pp. 439–449, Jun. 2022.

[89] V. Andrean, X.-H. Zhao, D. F. Teshome, T.-D. Huang, and K.-L. Lian, "A hybrid method of cascade-filtering and committee decision mechanism for non-intrusive load monitoring," *IEEE Access*, vol. 6, pp. 41212–41223, 2018.

[90] X. Wu, X. Han, L. Liu, and B. Qi, "A load identification algorithm of frequency domain filtering under current underdetermined separation," *IEEE Access*, vol. 6, pp. 37094–37107, 2018.

[91] A. Harell, S. Makonin, and I. V. Bajić, "Wavenilm: A causal neural network for power disaggregation from the complex power signal," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2019, pp. 8335–8339.

[92] P. A. Schirmer and I. Mporas, "Integration of temporal contextual information for robust energy disaggregation," in *Proc. IEEE 38th Int. Perform. Comput. Commun. Conf. (IPCCC)*, 2019, pp. 1–6.

[93] P. A. Schirmer, I. Mporas, and A. Sheikh-Akbari, "Robust energy disaggregation using appliance-specific temporal contextual information," *EURASIP J. Adv. Signal Process.*, vol. 2020, no. 1, p. 394, 2020.

[94] Y. Liu, X. Wang, and W. You, "Non-intrusive load monitoring by voltage–current trajectory enabled transfer learning," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5609–5619, Sep. 2019.

[95] P. A. Schirmer and I. Mporas, "Low-frequency energy disaggregation based on active and reactive power signatures," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, 2021, pp. 1426–1430.

[96] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for nonintrusive load monitoring," 2016, *arXiv:1612.09106*.

[97] P. Ducange, F. Marcelloni, and M. Antonelli, "A novel approach based on finite-state machines with fuzzy transitions for nonintrusive home appliance monitoring," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1185–1197, May 2014.

[98] M. J. Johnson and A. S. Willsky, "Bayesian nonparametric hidden semi-Markov models," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 673–701, 2013. [Online]. Available: http://dl.acm.org/ft_gateway.cfm?id=2502602&type=pdf

[99] P. A. Schirmer, I. Mporas, and A. Sheikh-Akbari, "Energy disaggregation using two-stage fusion of binary device detectors," *Energies*, vol. 13, no. 9, p. 2148, 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/9/2148

[100] S. M. Tabatabaei, S. Dick, and W. Xu, "Toward non-intrusive load monitoring via multi-label classification," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 26–40, Jan. 2017.

[101] P. A. Schirmer and I. Mporas, "Binary versus multiclass deep learning modelling in energy disaggregation," in *Proc. 2nd ICESF*, 2021, pp. 45–51.

[102] J. A. Mueller and J. W. Kimball, "Accurate energy use estimation for nonintrusive load monitoring in systems of known devices," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2797–2808, Jul. 2018.

[103] Y. Liu, G. Geng, S. Gao, and W. Xu, "Non-intrusive energy use monitoring for a group of electrical appliances," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3801–3810, Jul. 2018.

[104] M. Svensén and C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2007.

[105] D. Yang, X. Gao, L. Kong, Y. Pang, and B. Zhou, "An event-driven convolutional neural architecture for non-intrusive load monitoring of residential appliance," *IEEE Trans. Consum. Electron.*, vol. 66, no. 2, pp. 173–182, May 2020.

[106] M. Kaselimi, N. Doulamis, A. Doulamis, A. Voulodimos, and E. Protopapadakis, "Bayesian-optimized bidirectional LSTM regression model for non-intrusive load monitoring," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2019, pp. 2747–2751.

[107] H.-H. Chang, "Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses," *Energies*, vol. 5, no. 12, pp. 4569–4589, 2012.

[108] P. A. Schirmer, I. Mporas, and M. Paraskevas, "Evaluation of regression algorithms and features on the energy disaggregation task," in *Proc. 10th Int. Conf. Inf. Intell. Syst. Appl. (IISA)*, 2019, pp. 1–4.

[109] Y. Pan, K. Liu, Z. Shen, X. Cai, and Z. Jia, "Sequence-to-subsequence learning with conditional Gan for power disaggregation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2020, pp. 3202–3206.

[110] T.-T.-H. Le, S. Heo, and H. Kim, "Toward load identification based on the Hilbert transform and sequence to sequence long short-term memory," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3252–3264, Jul. 2021.

[111] D. Benyoucef, P. Klein, and T. Bier, "Smart Meter with non-intrusive load monitoring for use in smart homes," in *Proc. IEEE Int. Energy Conf.*, 2010, pp. 96–101.

[112] J. Gao, E. C. Kara, S. Giri, and M. Bergés, "A feasibility study of automated plug-load identification from high-frequency measurements," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Orlando, FL, USA, 2015, pp. 220–224.

[113] G. C. Koutitas and L. Tassiulas, "Low cost disaggregation of smart meter sensor data," *IEEE Sensors J.*, vol. 16, no. 6, pp. 1665–1673, Mar. 2016.

[114] T. Sirojan, S. Lu, B. T. Phung, and E. Ambikairajah, "Embedded edge computing for real-time smart meter data analytics," in *Proc. Int. Conf. Smart Energy Syst. Technol. (SEST)*, 2019, pp. 1–5.
[115] T. Kriechbaumer and H.-A. Jacobsen, "BLOND, a building-level office environment dataset of typical electrical appliances," *Sci. Data*, vol. 5, Mar. 2018, Art. no. 180048. [Online]. Available: https://www.nature.com/articles/sdata201848.pdf
[116] J. Zhao, J. Liu, Z. Qin, and K. Ren, "Privacy protection scheme based on remote anonymous attestation for trusted smart meters," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3313–3320, Jul. 2018.
[117] H. Wang, J. Zhang, C. Lu, and C. Wu, "Privacy preserving in non-intrusive load monitoring: A differential privacy perspective," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2529–2543, May 2021.
[118] C.-C. Sun, D. J. S. Cardenas, A. Hahn, and C.-C. Liu, "Intrusion detection for cybersecurity of smart meters," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 612–622, Jan. 2021.
[119] N. Batra *et al.*, "NILMTK," in *Proc. 5th ACM Int. Conf. Future Energy Syst. (e-Energy)*, Cambridge, U.K., Jun. 2014, pp. 265–276.
[120] D. Srdjan, D. Marko, and L. Vanco, "A non-intrusive identification of home appliances using active power and harmonic current," *Electron. Energetics*, vol. 30, no. 2, pp. 199–208, 2017, doi: 10.2298/FUEE1702199D.
[121] J. Z. Kolter and M. J. Johnson, Eds., "REDD: A public data set for energy disaggregation research," in *Artificial Intelligence*. Detroit, MI, USA: Greenhaven Press, 2011.
[122] K. D. Anderson, M. E. Bergés, A. Ocneanu, D. Benitez, and J. M. F. Moura, "Event detection for non intrusive load monitoring," in *Proc. 38th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Montreal, QC, Canada, 2012, pp. 3312–3317.
[123] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, "The ECO data set and the performance of non-intrusive load monitoring algorithms," in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Build. (BuildSys)*, 2014, pp. 80–89.
[124] J. Kelly and W. Knottenbelt, "The U.K.-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five U.K. homes," *Sci. Data*, vol. 2, Mar. 2015, Art. no. 150007.
[125] C. Holcomb, Ed., *A Test-Bed for NILM*. Pittsburgh, PA, USA, 2012.
[126] S. Barker *et al.*, "Smart*: An open data set and tools for enabling research in sustainable homes," in *Proc. ACM SustKDD*, vol. 111, Aug. 2012, p. 108.
[127] S. Makonin, Z. Wang, and C. Tumpach, "RAE: The rainforest automation energy dataset for smart grid meter data analysis," *Data*, vol. 3, no. 1, p. 8, 2018.
[128] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava, "It's different," in *Proc. 5th ACM Workshop Embedded Syst. Energy-Efficient Build.*, 2013, pp. 1–8.
[129] G. Hebrail and A. Berard. "Individual Household Electric Power Consumption Data Set." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption# (Accessed: Jul. 12, 2022).
[130] D. Murray, L. Stankovic, and V. Stankovic, "An electrical load measurements dataset of united kingdom households from a two-year longitudinal study," *Sci. Data*, vol. 4, no. 1, pp. 1–12, 2017.
[131] S. Makonin, B. Ellert, I. V. Bajić, and F. Popowich, "Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014," *Sci. Data*, vol. 3, no. 1, pp. 1–12, 2016.
[132] N. Batra, O. Parson, M. Berges, A. Singh, and A. Rogers, "A comparison of non-intrusive load monitoring methods for commercial and residential buildings," 2014, *arXiv:1408.6595*.
[133] A. S. U. Nambi, A. R. Lua, and V. R. Prasad, "LocED," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Build. (BuildSys)*, Seoul, South Korea, Nov. 2015, pp. 45–54.
[134] M. Ribeiro, L. Pereira, F. Quintal, and N. Nunes, "SustDataED: A public dataset for electric energy disaggregation research," in *Proc. ICT Sustain.*, 2016, pp. 244–245.
[135] G. Johnson and I. Beausoleil-Morrison, "Electrical-end-use data from 23 houses sampled each minute for simulating micro-generation systems," *Appl. Thermal Eng.*, vol. 114, pp. 1449–1456, Mar. 2017.
[136] C. Klemenjak, C. Kovatsch, M. Herold, and W. Elmenreich, "A synthetic energy dataset for non-intrusive load monitoring in households," *Sci. Data*, vol. 7, no. 1, pp. 1–17, 2020.
[137] L. Yan, J. Han, R. Xu, and Z. Li, "LIFTED: Household appliance-level load dataset and data compression with lossless coding considering precision," in *Proc. IEEE Power Energy Soc. General Meeting (PESGM)*, 2020, pp. 1–5.
[138] N. E. E. Alliance, "Residential building stock assessment: Metering study," Ecotope Inc., Seattle, WA, USA, Rep. #E14-283, 2014.
[139] J.-P. Zimmermann *et al.*, *Household Electricity Survey: A Study of Domestic Electrical Product Usage*, Leatherhead, U.K., Intertek Test. Certification Ltd., 2012.
[140] A. Reinhardt *et al.*, "On the accuracy of appliance identification based on distributed load metering data," in *Proc. Sustain. Internet ICT Sustain. (SustainIT)*, 2012, pp. 1–9. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6388037
[141] A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro, and A. M. Tonello, "GREEND: An energy consumption dataset of households in Italy and Austria," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Venice, Italy, Nov. 2014, pp. 511–516.
[142] J. Gao, S. Giri, E. C. Kara, and M. Bergés, "PLAID: A public dataset of high-resoultion electrical appliance measurements for load identification research," in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Build. (BuildSys)*, 2014, pp. 198–199.
[143] M. Kahl, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen, "Whited—A worldwide household and industry transient energy data set," in *Proc. 3rd Int. Workshop Non-Intrusive Load Monitor.*, 2016, pp. 1–4.
[144] M. Kahl *et al.*, "Measurement system and dataset for in-depth analysis of appliance energy consumption in industrial environment," *Technisches Messen*, vol. 86, no. 1, pp. 1–13, 2019.
[145] D. Jorde, T. Kriechbaumer, T. Berger, S. Zitzlsperger, and H.-A. Jacobsen, "CREAM, a component level coffeemaker electrical activity measurement dataset," *Sci. Data*, vol. 7, no. 1, pp. 1–13, 2020.
[146] M. Gulati, S. S. Ram, and A. Singh, "An in depth study into using EMI signatures for appliance identification," in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Build. (BuildSys)*, 2014, pp. 70–79.
[147] T. Picon *et al.*, "COOLL: Controlled on/off loads library, a public dataset of high-sampled electrical signals for appliance identification," 2016, *arXiv:1611.05803*.
[148] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proc. SIAM Int. Conf. Data Min.*, Philadelphia, PA, USA, 2011, pp. 747–758.
[149] P. A. Schirmer and I. Mporas, "Energy disaggregation using fractional calculus," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2020, pp. 3257–3261.
[150] P. A. Schirmer and I. Mporas, "Energy disaggregation from low sampling frequency measurements using multi-layer zero crossing rate," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2020, pp. 3777–3781.
[151] H.-H. Chang, L.-S. Lin, N. Chen, and W.-J. Lee, "Particle-swarm-optimization-based nonintrusive demand monitoring and load identification in smart meters," *IEEE Trans. Ind. Appl.*, vol. 49, no. 5, pp. 2229–2236, Sep./Oct. 2013.
[152] W. Kong, Z. Y. Dong, J. Ma, D. J. Hill, J. Zhao, and F. Luo, "An extensible approach for non-intrusive load disaggregation with smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3362–3372, Jul. 2018.
[153] K. Chen, Q. Wang, Z. He, K. Chen, J. Hu, and J. He, "Convolutional sequence to sequence non-intrusive load monitoring," 2018, *arXiv:1806.02078*.
[154] H. Wang, W. Yang, T. Chen, and Q. Yang, "An optimal load disaggregation method based on power consumption pattern for low sampling data," *Sustainability*, vol. 11, no. 1, p. 251, 2019.
[155] C. Dinesh, B. W. Nettasinghe, R. I. Godaliyadda, M. P. B. Ekanayake, J. Ekanayake, and J. V. Wijayakulasooriya, "Residential appliance identification based on spectral information of low frequency smart meter measurements," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2781–2792, Nov. 2016.
[156] Q. Liu, K. M. Kamoto, X. Liu, M. Sun, and N. Linge, "Low-complexity non-intrusive load monitoring using unsupervised learning and generalized appliance models," *IEEE Trans. Consum. Electron.*, vol. 65, no. 1, pp. 28–37, Feb. 2019.
[157] B. Zhao, K. He, L. Stankovic, and V. Stankovic, "Improving event-based non-intrusive load monitoring using graph signal processing," *IEEE Access*, vol. 6, pp. 53944–53959, 2018.
[158] F. Jazizadeh, B. Becerik-Gerber, M. Berges, and L. Soibelman, "An unsupervised hierarchical clustering based heuristic algorithm for facilitated training of electricity consumption disaggregation systems," *Adv. Eng. Inform.*, vol. 28, no. 4, pp. 311–326, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474034614000913
[159] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Non-intrusive appliance load monitoring using low-resolution smart meter data," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)* Venice, Italy, Nov. 2014, pp. 535–540.

[160] S. Singh and A. Majumdar, "Deep sparse coding for non–intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4669–4678, Sep. 2018.
[161] E. Elhamifar and S. Sastry, "Energy disaggregation via learning 'powerlets' and sparse coding," in *Proc. 29th AAAI Conf. Artif.*, 2015, pp. 629–635. [Online]. Available: https://dl.acm.org/citation.cfm?id=2887007.2887095
[162] P. A. Schirmer and I. Mporas, "Multivariate non-negative matrix factorization with application to energy disaggregation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2021, pp. 3285–3289.
[163] A. Harell, R. Jones, S. Makonin, and I. V. Bajić, "TraceGAN: Synthesizing appliance power signatures using generative adversarial networks," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 4553–4563, Sep. 2021.
[164] H. Rafiq, X. Shi, H. Zhang, H. Li, M. K. Ochani, and A. A. Shah, "Generalizability improvement of deep learning-based non-intrusive load monitoring system using data augmentation," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3265–3277, Jul. 2021.
[165] R. Jones, C. Klemenjak, S. Makonin, and I. V. Bajic, "Exploring Bayesian surprise to prevent overfitting and to predict model performance in non-intrusive load monitoring," 2020, *arXiv:2009.07756*.
[166] C. Klemenjak, S. Makonin, and W. Elmenreich, "Towards comparability in non-intrusive load monitoring: On data and performance evaluation," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, 2020, pp. 1–5.
[167] M. DrIncecco, S. Squartini, and M. Zhong, "Transfer learning for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1419–1429, Mar. 2020.
[168] L. Massidda, M. Marrocu, and S. Manca, "Non-intrusive load disaggregation by convolutional neural network and multilabel classification," *Appl. Sci.*, vol. 10, no. 4, p. 1454, 2020.
[169] M. Zeifman, "Disaggregation of home energy display data using probabilistic approach," *IEEE Trans. Consum. Electron.*, vol. 58, no. 1, pp. 23–31, Feb. 2012.
[170] J. Kelly and W. Knottenbelt, "Neural NILM," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Build. (BuildSys)*, Seoul, South Korea, Nov. 2015, pp. 55–64.
[171] M. Aftab, C.-K. Chau, and M. Khonji, "Real-time appliance identification using smart plugs: Demo abstract," in *Proc. 8th Int. Conf. Future Energy Syst. (e-Energy)*, Hong Kong, May 2017, pp. 262–263.
[172] Y.-H. Lin and M.-S. Tsai, "Development of an improved time–frequency analysis-based nonintrusive load monitor for load demand identification," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 6, pp. 1470–1483, Jun. 2014.
[173] C. Klemenjak, A. Faustine, S. Makonin, and W. Elmenreich, "On metrics to assess the transferability of machine learning models in non-intrusive load monitoring," 2019, *arXiv:1912.06200*.

Pascal A. Schirmer received the B.Eng. degree in electrical engineering from the University of Applied Sciences, Esslingen, Germany, in 2018, and the Ph.D. degree in electrical engineering from the University of Hertfordshire, U.K., in 2021. Since 2021, he has been working with the Research and Development Department for the development of power electronics, BMW AG, Munich, Germany, where he is responsible for the lifetime evaluation of power electronic systems. Furthermore, he is a Visiting Research Fellow with the University of Hertfordshire, U.K., where he is focusing on NILM (software and hardware applications) and a Visiting Lecturer for electro-mobility with TAE, Esslingen, Germany.

Iosif Mporas received the Diploma (5-years degree) degree in electrical and computer engineering and the Ph.D. degree from the University of Patras, Greece, in 2004 and 2009, respectively. He is a Reader of Signal Processing and Machine Learning (Associate Professor) with the University of Hertfordshire, U.K. He has participated in more than Ten EU-funded Research and Development Projects as a Researcher, a Senior Researcher, and a Principal Investigator. He has authored or coauthored more than 100 articles in international journals and conferences cited more than 1385 times (H-index: 18). His research interests include applications of signal processing and machine learning. He serves as a reviewer for grant applications and international journals and a Programme Committee Member in international conferences, while he was the General Chair of the joint SPECOM/ICR 2017 conference and the Technical Chair of the ICESF 2020 Conference.