Energy & Buildings 305 (2024) 113890

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

A review of current methods and challenges of advanced deep learning-based non-intrusive load monitoring (NILM) in residential context

Hasan Rafiq [a], Prajowal Manandhar [a], Edwin Rodriguez-Ubinas [a,*], Omer Ahmed Qureshi [a], Themis Palpanas [b]

[a] Dubai Electricity and Water Authority, DEWA R&D Centre, Dubai, United Arab Emirates
[b] LIPADE - Université Paris Cité, Paris 75006, France

ARTICLE INFO

Keywords:
Non-intrusive load monitoring
NILM
Data-driven algorithms
Real-time deployment
Deep learning
Load disaggregation

ABSTRACT

The rising demand for energy conservation in residential buildings has increased interest in load monitoring techniques by exploiting energy consumption data. In recent years, hundreds of research articles have been published that have mainly focused on data-driven non-intrusive load monitoring (NILM) approaches. Due to the high volume of research articles published in this domain, it has become necessary to provide a review of the up-to-date research in NILM and highlight the current challenges associated with its application. This paper reviews the state-of-the-art of NILM by following a structured assessment process to consider relevant and most recent documents in the literature. It presents the pros and cons of data-driven NILM methods, available datasets, and performance evaluation mechanisms. Even though research in NILM solutions has matured in recent years thanks to the use of deep learning models, there are still gaps in their effective deployment related to data requirements, real-time performance, and interpretability. Therefore, the paper also addresses the NILM development and implementation challenges and includes promising improvement measures that can be utilized to solve them.

1. Introduction

Greenhouse gas emissions are the prevalent threat to the world's sustainable energy future. Since the mid-21st century, incremental change in CO_2 concentration has risen. The increase in CO_2 concentration by 2.13 parts per million (ppm) in 2022, the 11th consecutive year with concentrations increase higher than 2 ppm, has put the global CO_2 concentration level at 417.06 ppm [1]. The natural greenhouse effect has been supercharged by this increase in CO_2 levels, resulting in global warming and an increase in the average global surface temperature [2]. Considering carbon emissions as a global challenge, developed countries have set targets to reduce emissions and conserve energy [3]. Building energy use, representing 30 % of global energy use and 27 % of total emissions from the energy sector [4], increased to 135 EJ (1 EJ = 1018 J) in 2021 from the previous level of 115 EJ in 2010. To get on track with Net Zero Scenario by 2030, the specific energy consumption in buildings must be 35 % less than the 2021 levels [5].

These numbers show that energy conservation in buildings, especially residential ones, can significantly reduce overall energy demand and associated environmental impact. Therefore, significant research efforts are directed at determining efficient energy utilization in residential buildings. Energy-use-intensity (EUI) in residential buildings can be reduced through efficient design by adopting climate-appropriate passive [6] and active strategies [7]. However, smart electrification can also contribute to reducing energy waste in residential buildings by monitoring the energy use of individual systems or appliances and providing real-time actionable feedback to consumers. The International Electrotechnical Commission (IEC) has stated that "smart electrification, the intelligent and economical use of electricity as a major energy source, is the best way to solve energy problems" [8,9]. Vassileva et al. [10] reported that 80 % of consumers showed interest in appliance-based electricity consumption segregation. This segregation helps to identify appliances with the highest share of consumption, and the consumers can take corrective measures to optimize their energy use. Real-time awareness of consumption trends and information on energy-intensive appliances can help to reduce electricity consumption by up to 12 % in residential and commercial buildings [11,12]. Conventional appliance-based electrical load monitoring requires multiple energy meters installed on different electrical appliances. Although this method helped achieve the desired energy segregation, it has various

* Corresponding author.
E-mail address: edwin.ubinas@dewa.gov.ae (E. Rodriguez-Ubinas).

https://doi.org/10.1016/j.enbuild.2024.113890
Received 8 September 2023; Received in revised form 19 December 2023; Accepted 3 January 2024
Available online 7 January 2024

disadvantages, such as high cost, time-consuming installation, and complexity [13].

To overcome these challenges, Hart [14] introduced the problem of breaking down the aggregated power signal into several appliance-level signals, using a non-intrusive method. Non-intrusive means distributed direct sensing is not required; single-point sensing is employed [15]. Therefore, Non-Intrusive Load Monitoring (NILM), or simply energy disaggregation, is an imperative part of computational sustainability research where researchers develop algorithms to disaggregate the energy consumption of individual appliances from an aggregated energy signal retrieved from the main energy meter. Unlike traditional techniques, it uses an advanced energy analysis technique that does not need to install sensors or monitoring devices for each appliance. Fortunately, due to large-scale deployments of smart meters in buildings, clean, high-resolution, and high-accuracy aggregated data is available.

The availability of high-resolution aggregated data supports the ease of implementation and usability of the NILM in residential buildings. Applying advanced energy analysis techniques on the aggregated power usage data (mainly voltage and currents) acquired at the main electrical panel of the building or obtained from smart meters can enable consumers to monitor appliance-based energy usage and operational status information. This consumption breakdown is helpful for utility companies as NILM helps better understand consumers' energy usage patterns and load constitution. Such information is essential for utility companies to manage the load, peak demand management, and ensure grid stability by educating their customers about more efficient power consumption based on individual consumption information [16]. It is also helpful to increase the accuracy of power usage prediction to assist system planning [17]. Consumers also benefited by gaining insights into their electricity usage and identifying areas for energy savings through disaggregated data [18]. Additionally, the disaggregated consumption information helps to predict the appliance's health. Further, customers can get actionable feedback about their appliance's health in terms of (1) increased consumption alert: if a particular appliance was consuming high energy in the past few days compared to last week or last month, (2) irregularities in consumption patterns alert: if sharp peaks or lower magnitude activations were found in the patterns, (3) fault detection alert: the disaggregated information can aid in detecting over-voltage and over-current faults along with smart meter's aggregate readings, and (4) periodic maintenance alert: a type of feedback that allows customers to get their appliances checked and maintained on regular basis. Some people might negatively perceive NILM due to privacy concerns and possible unwanted feedback. However, the multiple positive effects of NILM outweigh the negative ones.

Energy disaggregation is imperative for energy conservation in the building sector. Considering this objective, researchers worldwide have proposed novel ways to disaggregate household appliance's energy. Many new articles are published every year in the NILM domain, and there is always a need to do a detailed and up-to-date study that will present a comprehensive review of the current state-of-the-art of NILM. In past years, various articles have reviewed these works by highlighting different aspects of energy disaggregation's state-of-the-art. Some impactful review articles that discussed NILM dataset analysis, feature extraction, learning methods, and evaluation methods can be found in [19–22]. However, the applicability of those learning methods is not discussed in any of the review works. Due to the large-scale application of deep learning for energy disaggregation, it is vital to review current NILM research from its applicability perspective and highlight the challenges associated with these methods so that researchers can focus on solving those challenges that would help NILM solutions to be successfully applied in the real world. Therefore, the main objectives of the present study are:

1) To comprehensively review the evolution and current NILM methods applied to residential buildings/households, categorizing them with respect to their maturity and advancements. The review also discusses their pros and cons, data requirements, interpretability, and performance.
2) To analyze in detail the challenges of applying deep-learning-based NILM methods in the real world and provide valuable insights into the hurdles these methodologies face.
3) To discuss potential improvement measures by following a structured and solution-oriented approach, i.e., discuss solutions considering three essential parts of the NILM framework: data preprocessing and feature selection, model implementation, and evaluation.

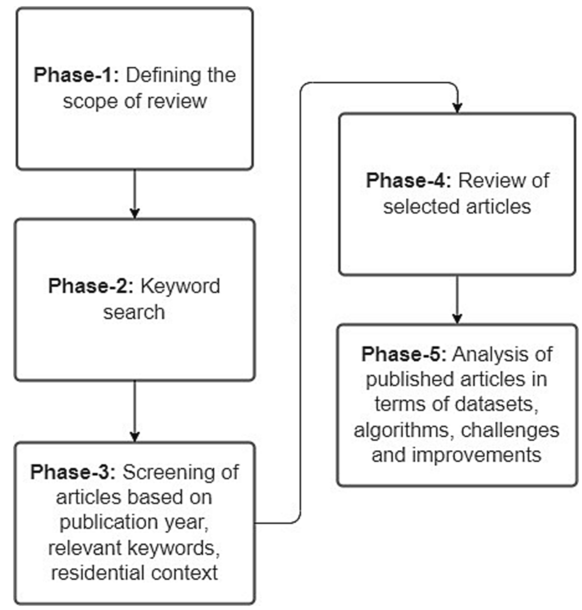

Fig. 1. The study screening and review process phases.

The paper is structured in the following manner. Section 2 provides a step-by-step process for reviewing NILM literature. Section 3 presents a brief background of NILM technology from the algorithm's framework and real-time implementation point-of-view. Section 4 provides a review of data-driven NILM approaches. Section 5 discusses publicly available datasets and metrics necessary for accurate evaluation and real-time deployment. Sections 6 and 7 explain some challenges in data-driven NILM technology, followed by improvement measures. Section 8 concludes this review paper.

2. Method

This review paper follows a structured assessment process to select relevant literature and provide up-to-date reviews of non-intrusive load monitoring articles. The review process has five phases, as shown in Fig. 1.

2.1. Phase 1: Scope of review

Non-intrusive load monitoring (NILM) applications can utilized in residential, commercial, and industrial sectors. However, residential or household energy disaggregation has received more attention from researchers. Most research work and commercial NILM products target this domain; therefore, this review's scope is limited to residential energy disaggregation systems and methods.

Residential energy disaggregation provides individual consumption information on household appliances and systems. NILM provides actionable feedback to consumers and engages them in energy-saving programs. Whereas commercial and industrial applications would include disaggregation of their loads to reduce overall energy waste.

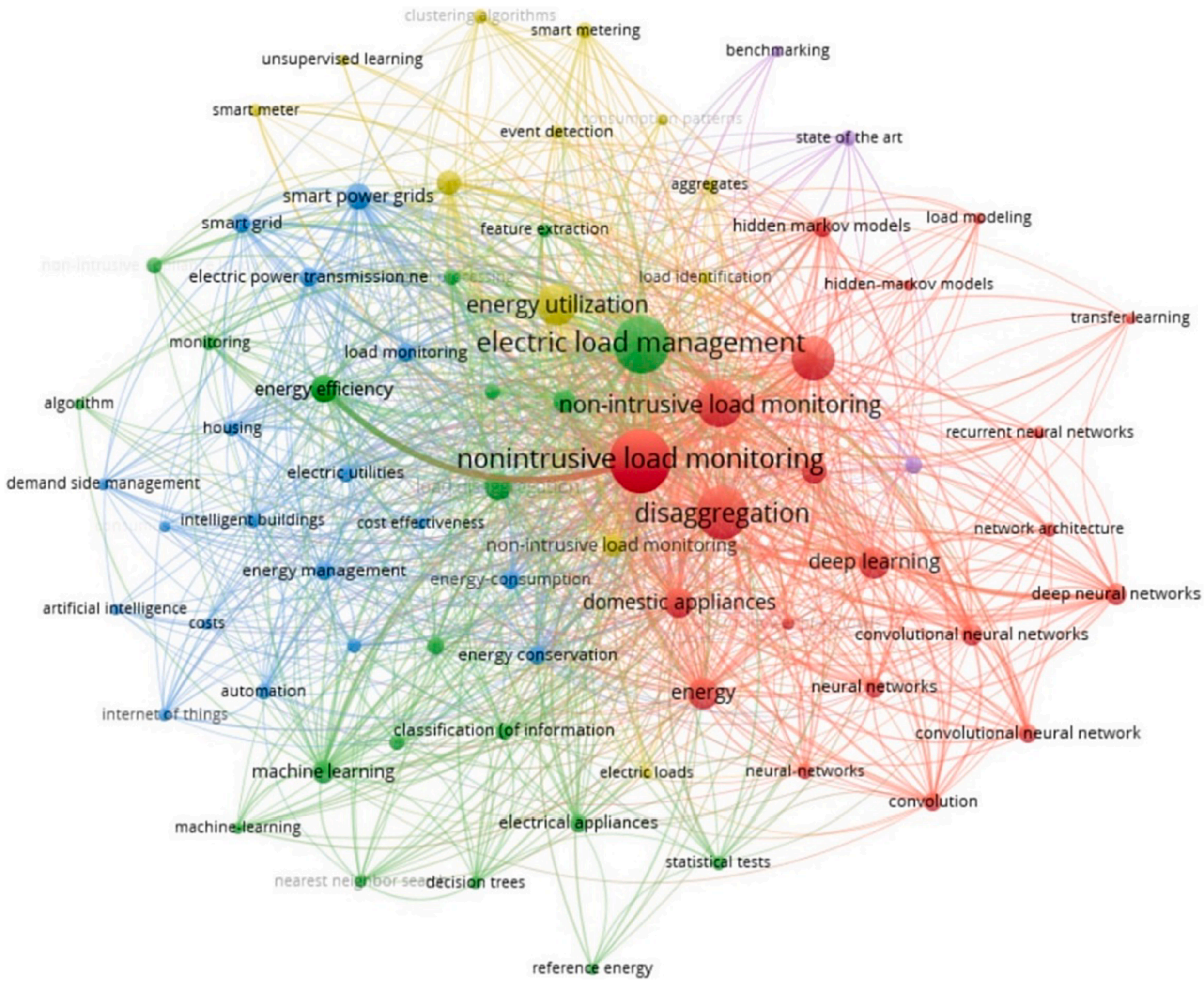

Fig. 2. A network of keyword co-occurrences associated with NILM. This figure shows a clear trend of using deep learning and machine learning for NILM problems.

2.2. Phase 2: Keyword search

A keyword-based search of research articles, including review articles, was done using Scopus. The keywords used for the search were non-intrusive load monitoring, non-intrusive appliance load monitoring, energy disaggregation, and load disaggregation. Fig. 2 shows a co-occurrence network of keywords associated with non-intrusive load monitoring. The co-occurrence network shows the total link strength of all keywords found in the literature (specifically machine learning and deep learning) during a keyword-based search with NILM or energy disaggregation. This strong relation indicates that deep learning has often been applied to solve energy disaggregation problems.

2.3. Phase 3: Screening of articles

The articles were screened using three-step criteria: (1) the publication year of the articles, (2) whether NILM was applied to residential load, commercial load, or industrial load, and (3) whether data-driven approaches (mainly machine learning or deep learning) were proposed in the articles or not. Once published articles were screened based on their application on residential buildings only, then selected articles were analyzed based on:

1) Type of data-driven algorithm (e.g., traditional time series, shallow learning algorithms, advanced machine learning, and deep learning algorithms),
2) Type of dataset used (e.g., public, private, and their name),
3) Sampling frequency of the data (e.g., low frequency or high frequency),
4) Type of input features (e.g., electrical measurements, calendar information, and seasons),
5) Types of evaluation metrics used (e.g., MAE, F1-score, accuracy, or other domain-specific metrics).

2.4. Phase 4: Review of selected articles

As mentioned, the articles were reviewed based on data type, resolution, algorithm type, and how those algorithms were evaluated. Furthermore, main contributions and future work suggestions were noted down for each reviewed article. All the relevant information was extracted and compiled as tables for our analysis.

2.5. Phase 5: Analysis of published articles

Based on the available information, the analysis is presented in terms of:

1) NILM background, including NILM problem formulation and NILM algorithm's framework.
2) Learning algorithm types used in NILM (including their pros and cons).

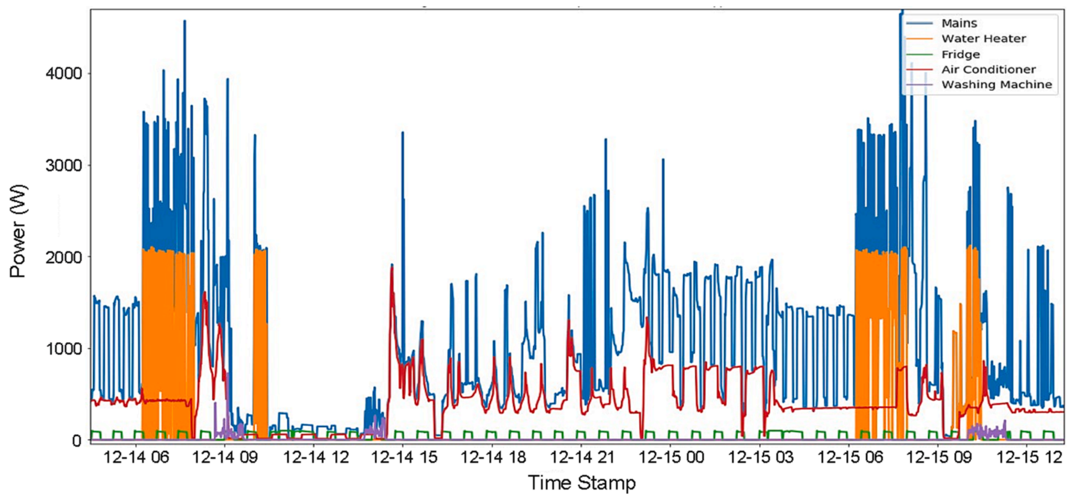

Fig. 3. Total power consumption (aggregate mains) plot is shown in blue colour. The target is to estimate the individualized consumption of the appliances at a given time index. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

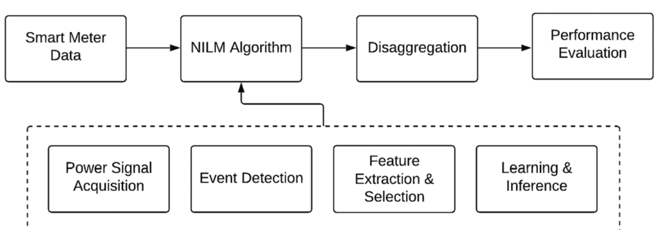

Fig. 4. NILM algorithm workflow ().
reproduced from [23]

3) Datasets and performance evaluation metrics.
4) Challenges in NILM research.
5) Potential improvements.

3. Background

3.1. NILM problem formulation

For any household, if the power consumption per appliance is known, the well-posed forward problem is to estimate the aggregated power consumption. On the contrary, energy disaggregation, an inverse problem, aims to determine the unknown power consumption of each appliance using the known information about the aggregate measurements. Refer to Fig. 3, which shows the aggregate and disaggregated consumption of a group of household appliances.

Assuming the aggregate signal $X(t)$ at any discrete time stamp t is equal to the sum of all the appliances' energy usage rate $Y_i(t)$ plus an unidentifiable electrical signal as additive noise $e(t)$. Thus:

$$X(t) = \sum_{i=0}^{K} Y_i(t) + e(t) \tag{1}$$

where, I refers to the i^{th} appliance from the total number of appliances K.

The aggregated consumption signal $X(t)$ is given at each time stamp t. The energy consumption rate of the individual appliance $Y_i(t)$ could be available a priori (leading to supervised learning), or in the absence of individual metering, this data could be unavailable (leading to unsupervised learning). In the NILM framework, the best estimates $\widehat{Y}_i(t)$ of the individual appliance power are determined using the known aggregated value $X(t)$ and applying advanced energy analysis techniques. The key challenge to the energy disaggregation problem is the accurate identification of the live data from the smart meter in real-time and the generalization of applicability on other unseen datasets.

3.2. NILM Algorithm's framework

A typical framework of the NILM algorithm, as described by Desai et al. [23], consists of the following steps: power data acquisition, event detection, feature extraction, and learning and inference, as shown in Fig. 4 below.

3.2.1. Power data acquisition

The first step for any NILM algorithm is Power data acquisition. This process involves receiving aggregated energy consumption measurements at an acceptable rate to identify distinctive load patterns. Considering how data is collected, three characteristics affect the performance of the NILM system.

1) The type of power measure: real, reactive, or both.
2) The power meter resolution: the smallest change in power that can be detected.
3) The data sampling frequency: low- or high-frequency.

Based on the selected NILM algorithm, the required characteristics of power data collection is determined. Different algorithms require

Table 1
Different disaggregation capabilities at different sampling frequencies [12].

Data Sampling Frequency	8hr – 15 min	1 min – 1 sec	10 Hz – 2 kHz	10 kHz – 15 kHz	>1MHz
Data Type used by the Algorithms	Visually observable patterns; duration and time of use if consistent.	Steady-state steps	Low order harmonics	Medium-order harmonics to identify the type of electrical circuitry in an appliance.	Very high-order harmonics to identify both transients and background noise by appliances.
Identified Appliances	General categories of base load and variable load.	<10 appliance types: Fridges, ACs, Heaters, Washing machines, Pool pumps.	Not known	20 – 40 appliance types, such as toasters and computers.	40 – 100 specific appliances (e.g., differentiates 2 lights; requires separate power consumption data stream).

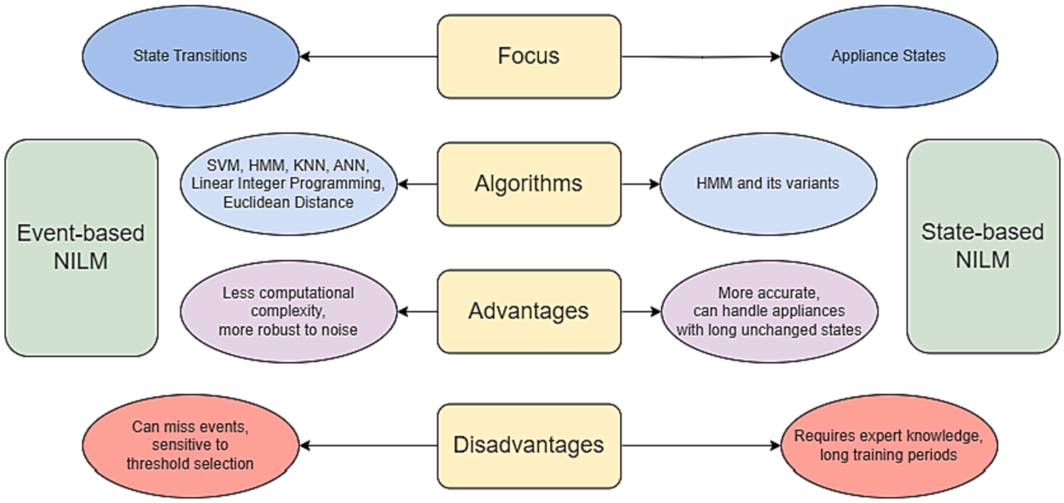

Fig. 5. Summary of the features for event-detection.

different measurements and electrical characteristics [24]. A sampling rate of below 1 Hz is considered low-frequency sampling. However, if the rate is between 1 and 2000 Hz, it is classified as high-frequency sampling [12]. Smart meters belong to low-frequency samples since they communicate readings every 15 min. Smart meter data collection is a cost-efficient approach for acquiring aggregate power data, rolled out in almost all major countries. Power meters required for high-frequency sampling are often custom-built and expensive due to sophisticated hardware [24]. Armel et al. graphically characterise the disaggregated appliances depending upon the sample frequency range, as shown in Table 1 below.

3.2.2. Event detection

NILM algorithms are generally classified into event-based and state-based. An event or edge refers to any change or discontinuity in the energy rate. In the NILM algorithm workflow, the event detection module senses the instances when state transition actions have occurred in the aggregated power signal. The state transition actions can be characterised as changes in power levels (like ON/OFF), appliance speed alterations, and operational mode changes.

The event-based approaches focus on these state transition actions that are generated by appliances and use a change detection algorithm to identify the start and end of an event [25,26]. It is a complex process due to the presence of multiple like and unlike appliances and variations in the state transition actions exhibited by these appliances. The task of the change detection algorithm is to detect these changes in the aggregated power signal. If a significant change (i.e., a ramp or step change) occurs in the primary power signal, the detection module will extract this event and compare it with the characteristic amplitude and signature of appliances that are typically found in buildings. Hence, the event is associated with a specific appliance's state transition actions. This type of event detection approach is also called the "Expert Heuristics Models." There is another category of event detection algorithms called "Matched Filter Models," that match unknown signals to templates in a features library. These methods rely on establishing a comprehensive features library, making them dependent on extensive device motion information and updating when new appliances are added.

A review of event detection algorithms used in the NILM literature is presented in [27]. Different classification methods such as Neural Networks, Fuzzy Logic, Support Vector Machine (SVM), Naive Bayes, k-Nearest Neighbors (kNN), decision trees, Hidden Markov Model (HMM), and many other hybrid approaches have been used [28,29]. Zhang et al. [30] proposed a novel NILM event detection algorithm based on different frequency scales. The proposed event detection approach was composed of two parts: high sensitivity event pre-detection sub-algorithm called voting improved isolated forest (VIIF), and an event verification sub-algorithm called time shift down-sampling matching (TSDM). The input data was high-frequency sampled, which was processed to low-frequency sampled data before applying VIIF and TSDM sub-algorithms. Rehman et al. [31] proposed two event detection algorithms by considering mean absolute deviation and variance of electrical signals as input features. Lu and Li [32] proposed hybrid event detection, which consisted of a main algorithm and two sub-algorithms. The main algorithm detected suspicious events, and sub-algorithms verified the suspected events.

State-based NILM approaches do not rely on event detectors; instead, they represent each appliance operation using a state machine with distinct state transitions based on appliance usage patterns [33]. These approaches are based on the fact that when an appliance turns ON/OFF or changes running states, it creates different edge measurements with a probability distribution that matches that appliance. State-based NILMs are usually based on HMM and its variants [34–37]. State-based approaches are limited by the need for expert knowledge to set a prior value for each appliance state via long training periods. Besides, they have high computational complexity [24,33] and do not have a way to handle the fact that states may stay unchanged for long time intervals [38]. A summary of event-based and state-based NILM approaches is presented in Fig. 5.

3.2.3. Feature extraction and selection

Unique features, or signatures, of each specific appliance are essential for any NILM algorithm. These features help distinguish between power signals coming from various appliances as a single aggregated signal. A meaningful appliances-based feature vector is a subset of all these relevant features. These features are filtered based on their statistical importance using multiple algorithms, such as the Relief algorithm [39]. This feature vector, associated with internal components of appliances, provides information about each appliance's intrinsic state transition actions. Depending on the sampling rate, different features can be extracted from the aggregated time-series power data.

The concern about the data availability coming from low-frequency measurements has been overcome in many locations due to the widespread deployment of commercial low-sampling rate smart meters. These meters can sample granularity of 1–10 s. Other concerns regarding low-frequency measurements are data storage, data management, and user privacy. It's important to note that low-frequency NILM differs fundamentally from its high-frequency counterpart. While high-frequency NILM effectively leverages signal transients and harmonic signal analysis to differentiate between multiple appliances and even two appliances of the same type (e.g., two air conditioners)

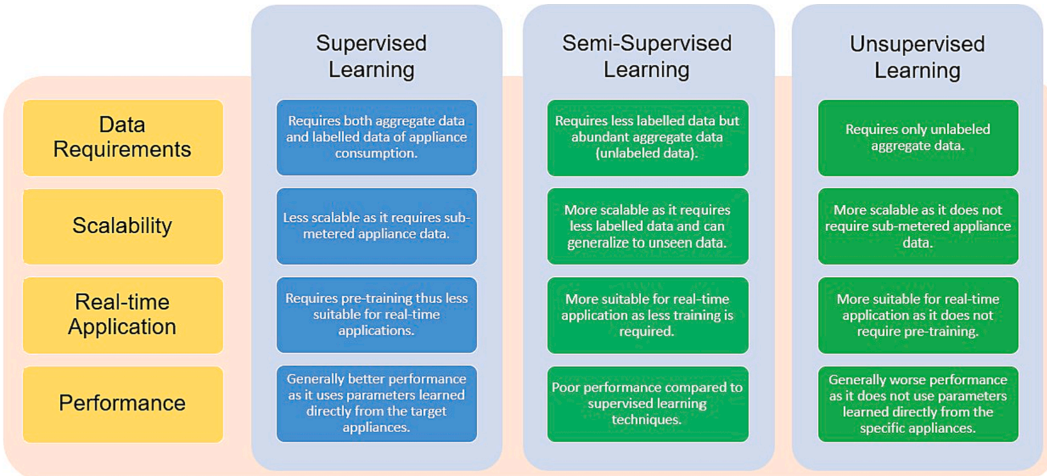

Fig. 6. Summary of supervised and unsupervised NILM algorithms (general).

[40], low-frequency NILM relies solely on steady-state power measurements. This necessitates entirely different methodological approaches to tackle the inverse problem, involving effective and robust machine learning or statistical data analysis, as emphasized in [33]. Some of the commonly used low-frequency features for load identification are the active-reactive power plane (P-Q plane) [25], macroscopic transients, active power [41], and current- and voltage-based features [42]. Among these features, most algorithms utilise active power measurement as it is available in most open-access datasets. Various statistical measures can be extracted from the active power data, including minimum, maximum, mean, and median values, percentiles, standard deviation, skewness, and kurtosis.

High-frequency measurements: High-frequency steady state and transient features are used in load identification [43], such as the spectral envelope, wavelets, shape features, raw waveforms, and voltage-current (V-I) trajectory. High-frequency-based NILM methods found in the literature are based either on spectrogram analysis [24,44] or on voltage-current trajectories [45]. Rather than relying solely on time-domain analysis, the power signal can also be characterised using two-dimensional (2D) representation [46]. Furthermore, high-frequency current data has been converted to spectrograms by the short-time Fourier transform (STFT) and set as the model input [44]. V-I-trajectory-enabled transfer learning method has also been proposed for the NILM algorithm training [47]. At first, the V-I trajectories are transformed into coloured images. Then, a pre-trained convolutional neural network (CNN) is used to classify the V-I trajectory images. Chang et al. [48] proposed a method for nonintrusive aging load identification (NIALM) in residential buildings using the Hellinger distance (HD) algorithm and a particle swarm optimization (PSO)-based back-propagation artificial neural network (BP-ANN). The HD algorithm was used to extract real power (P), reactive power (Q), total voltage harmonic distortion (V_{THD}) and current harmonic distortions (I_{THD}), and the various effective odd-order harmonic components of the current (I_{nth}) from 5-second sampled power data. Then the PSO-based BP-ANN was used to classify the extracted features. Chang et al. [49] proposed a method for non-intrusive fault identification (NIF) of power distribution systems (PDS) in intelligent buildings. The proposed method used the wavelet transform (WT) to extract wavelet transform coefficients (WTCs) from current waveforms of the complete fault transient signals. After that, a back-propagation artificial neural network (BP-ANN) was used to identify the fault events. Nour et al. [50] proposed data augmentation strategies for high-frequency NILM datasets. The proposed data augmentation approach involved mixing individual high-frequency sampled appliance signals, their placement at any instant, and manipulating appliance signal length by cutting, expanding, or leaving it. The synthetic data generation on high-frequency sampled data allows the extraction of high-frequency features for NILM model development.

Sometimes, additional information other than the power signal is also available. The additional information includes rated power, appliance types, historical data, and the number of appliances. NILM algorithms that utilize such additional information exhibit superior accuracy and reliable performance [42,51]. However, this information is not always available in most open-access datasets.

3.2.4. *Learning & inference*

Learning and inference is the final step of the NILM algorithm. This step uses a mathematical model to disaggregate the aggregate power signal into appliance-level signals. Once the appliance level signal is extracted, it is further analysed to classify appliance-specific states and estimate their corresponding power consumption. The learning algorithms (training) are used to develop the model parameters, and the inference algorithms (testing) are employed to infer appliance states from observed aggregate power data and estimate their corresponding power consumption.

The learning algorithm in NILM can be supervised, unsupervised, and semi-supervised based on the source of the appliances' model parameters. Supervised NILM techniques require a training phase in which both the aggregate data and the individual appliance consumption energy signatures are used. In this case, sub-metered appliance data or labelled observations must be collected from the target building. Collecting these data is expensive, time-consuming, and limits the scalability of NILM systems. Several earlier works have focused on supervised learning techniques such as Support Vector Machine (SVM) [52,53], Nearest Neighbor (kNN) [28], and some forms of Hidden Markov Models (HMM) [54]. Unlike supervised NILM, unsupervised NILM techniques do not require pretraining and are thus more suitable for real-time NILM applications. Unsupervised NILM approaches do not require individual appliance data, and the model parameters are captured only using the aggregated load without user intervention [55].

Semi-supervised NILM techniques use limited labelled data and plentiful unlabelled data to improve appliance identification and consumption estimation performance. Barsim et al. [56] work on event-based energy disaggregation is one of the early studies to use a semi-supervised technique, where they used a self-training approach to train a classifier model based on externally labelled data and its own high-confidence predictions. Fatouh et al. [57] proposed semi-supervised and active learning-based appliance classification approach. The proposed approach leverages both semi-supervised and active learning to process most certain and uncertain instances and provides high-confidence data. However, the author's approach relied on accurate labelling of data provided by the users, which is not always true. Iwayemi and Zhou [58] proposed a semi-supervised learning appliance annotation scheme for appliance signatures. Their approach used a mixture of labelled and unlabelled data to prepare appliance detection classifiers, focusing on stopping criteria in semi-supervised learning. Zhang et al. [59] proposed a temporal convolutional network-conditional random field (TCN-CRF) model that used optimal interval filtering and unanimous voting to identify high-precision segments in the sequence-to-sequence algorithm as semi-supervised learning targets. Similarly, Hur et al. [60] proposed a novel semi-supervised multi-label classification approach for NILM that combined robust knowledge transfer and network stabilization to perform generalization across domains. The proposed approach successively improved appliance detection on unlabelled target data.

Generally, the performance of the supervised NILM methods is superior to the semi-supervised and unsupervised methods. The supervised methods use parameters learned directly from the specific appliances. These parameters are then utilised to identify these appliances from the aggregated energy rate signal. However, from a real-time applicability point of view, semi-supervised and unsupervised techniques are preferable. A summary of learning-based NILM algorithms is presented in Fig. 6.

4. Learning methods for NILM

4.1. Early research on NILM

The idea of NILM was first proposed by R. D. Little [61] in 1991 for the disaggregation of electrical loads (specifically HVAC) in commercial buildings. However, G. W. Hart [14] proposed clustering comparable occurrences based on appliance features to disaggregate residential electrical loads [14]. Hart's study utilized the combinatorial optimization (CO) method, commonly accepted as the principal technique for resolving disaggregation issues. The initial methodology exhibited a significant drawback wherein the combinatorial optimization process executed power disaggregation for each instant individually without acknowledging the temporal evolution of the load. As previously noted, prevailing methodologies for addressing the Non-Intrusive Load Monitoring (NILM) predicament, typically, rely on unsupervised event detection within the aggregate signal. Various classification methods have been extensively employed, encompassing support vector machines (SVMs) [52], neural networks, decision trees (DTs) [62], and hybrid classification techniques [33,63]. In contrast to traditional methodologies, dynamic time warping (DTW) presents an alternative approach to analyzing daily profiles by comparing and merging windows while detecting distinctive load signatures [64]. The primary focus of contention within these approaches pertains to the arduousness of categorizing multi-state appliances, as outlined in references [63,65]. Multi-state appliances require training a long-range model to facilitate their detection. The concept of Graph Signal Processing (GSP), as described in the literature [33], is a highly effective approach for capturing spatiotemporal correlation among data samples by embedding the underlying signal structure into a graph. A recent proposal has been made for event classification through a modified cross-entropy method rooted in CO. This proposal presents NILM as a cross-entropy problem [66]. Hidden Markov models (HMMs) and their extended versions have been suggested as a means of investigating the plausible

Table 2
Summary of traditional machine learning approaches for NILM.

Ref.	Year	Dataset	Features	NILM Algorithm	Evaluation Metrics
[79]	2011	REDD	Active Power	Finite Hidden Markov Models (FHMM)	Total Energy Correctly Assigned
[37]	2012	REDD	Active Power	Hidden Markov Models (HMM)	Normalized Energy Error, RMSE
[80]	2014	HES	Active Power	Additive FHMM, Additive Factorial Approximate Maximum a Posteriori (AFAMAP)	Normalized Disaggregation Error, Signal Aggregate Error
[74]	2014	REDD REFIT	Active Power (edges detection)	Decision Trees, Dynamic Time Warping, HMM	Precision, Recall, F1-score
[28]	2014	REDD	Active Power, Min/Max value, Event duration	HMM, Support Vector Machine (SVM), k-Means-SVM	Precision, Recall, F1-score
[81]	2015	Simulated data	Active Power, Reactive Power, Apparent Power	Decision Trees	Accuracy
[82]	2015	Simulated data	Dynamic Time Warping (DTW)	AdaBoost	Accuracy
[83]	2015	Simulated data	Energy of Wavelet Coefficients	Decision Trees	Accuracy
[84]	2016	AMPds	Active Power, Reactive Power	AFAMAP and Forward Differential AFAMAP	Precision, Recall, F1-score
[33]	2016	REDD REFIT	Active Power	Graph Signal Processing	Precision, Recall, F1-score
[85]	2016	REDD AMPds	Active Power and Apparent Power	Aided Integer Linear Programming	Accuracy
[86]	2016	REDD	Active Power and Apparent Power	Dynamic Time Warping	Random k-label sets, Multi-label kNN,
[87]	2017	BLUED	Current	Decision Trees, SVM, k-Nearest Neighbors, Random Forests	Accuracy, Precision, Recall, F1-score
[88]	2018	Dataport	Seasonal Features (incl. weather information)	Adagrad (Transfer Learning)	Percentage of Energy Correctly Assigned
[89]	2019	REDD AMPds	Active Power and Reactive Power	Multi-objective evolutionary optimization	Accuracy, Precision, Recall, F1-score, Total Power Correctly Assigned
[90]	2019	REDD Dataport REFIT	Active Power and Discrete Fourier Transform (DFT)	Similar Time Window (STW), FHMM, Powerlet-based Energy Disaggregation, Multilabel, Classification, Sparse Coding (SC)	Accuracy
[91]	2019	AMPds	Active Power	Optimized SVR (OSVR), FHMM, Dynamic Time Warping	F1-score, Percentage of Energy Correctly Assigned, Root Mean Square Error
[92]	2020	PLAID Private dataset	Fast Fourier Transform-based (Magnitude and phase current harmonics)	Bagging Decision Tree	Accuracy, Precision, Recall, F1-score
[93]	2020	Georges Hebrail UCI UK-DALE	Active Power	Fuzzy Clustering	Precision, Recall, F1-score

amalgamations among diverse state sequences of appliances [35,67,68]. Theoretically, it was established that Hidden Markov Models (HMMs) operate on state-based principles, necessitating the inclusion of discrete states within the appliance signatures under investigation [69]. As the number of appliances rises, the number of feasible permutations of state sequences increases exponentially, resulting in a commensurate increase in complexity inherent to the problem.

Moreover, the augmentation of time complexity results in a concomitant deterioration of the classification efficacy of the model, as posited by reference [65]. Makonin et al. [70] presented a super-state Hidden Markov Model (HMM) along with a sparse Viterbi algorithm to mitigate computational complexity. One notable constraint of HMM-based methodologies is their susceptibility to malfunction when confronted with unfamiliar appliances. Rahimpour et al. [71] proposed a matrix factorization method for the linear decomposition of the aggregated signal. The approach utilizes the appliances' signatures as the basis of the model, which enhances the estimation of the energy consumption per appliance. The resultant technique presents an efficient solution for estimating energy consumption.

4.2. NILM research based on Machine learning and deep learning

The field of Non-Intrusive Load Monitoring (NILM) has garnered renewed academic interest in the last ten years, primarily attributed to the growth in the availability of datasets originating from smart electric meters implemented in household settings [72,73]. This increase in datasets has consequently led to a shift in proposed solutions for NILM, wherein supervised learning methods have become more prevalent. The advent of deep learning has ushered in a novel class of techniques that leverage deep neural network architectures to tackle the intractable task energy disaggregation. Since 2015, the application of deep learning techniques has primarily been focused on low-frequency NILM methods [55]. A widely adopted methodology is to consider the collective signal as a corrupted data sequence resulting from the presence of extraneous noise within the operation of a particular electrical device. The present perspective posits that denoising autoencoders (DAEs) constitute valuable methodologies particularly employed to reconstruct a signal derived from its noisy rendition. The current architecture was first introduced by Kelly and Knottenbelt [72], with subsequent contributions made by other researchers who proposed alternative designs and configurations of DAEs [67]. Leveraging the temporal characteristics and interrelationships exhibited by the power signal, a distinct category of deep learning models, known as recurrent neural networks (RNNs), has demonstrated notable efficacy within NILM. In this context, NILM is viewed through the lens of a supervised learning problem that incorporates time series analysis. Recurrent neural networks (RNNs) and their modifications, namely long short-term memory (LSTM) networks and gated recurrent units (GRUs), have predominantly been employed owing to their prevalence and efficacy in processing one-dimensional (1D) temporal data. Several pertinent studies have been conducted previously, as reported in references [55,74] and [65]. A context-aware Long Short-Term Memory (LSTM) model capable of adapting to external environmental conditions was introduced in reference [75]. Convolutional neural networks (CNNs) have historically been designed to process two-dimensional imagery data [76]. However, it is possible to leverage one-dimensional CNNs to effectively capture the temporal features inherent in sequential time series data. Few researchers [51] have endeavored to augment Convolutional Neural Network (CNN) architectures with a recurrent component, exemplified by CNN-LSTM and recurrent convolutional networks. A study by [42] introduced a causal 1D convolutional neural network for Non-Intrusive Load Monitoring (NILM) applications. Bonfigli et al. [45] introduced the concept of data sequences to provide historical power load values to the conventional structure. There have been proposals [77] advocating for a CNN architecture operating on a sequence-to-point basis, emphasizing the significance of incorporating a sliding windows approach for processing time series data with prolonged durations. An alternative proposal has been put forth in the form of sequence-to-sequence architectures in reference [78]. A summary of traditional machine-learning approaches is shown in Table 2.

4.3. NILM research based on advanced deep learning

Numerous advanced machine-learning techniques have been employed late in the NILM research domain. These methodologies not only exhibit a comparable level of precision as the conventional NILM techniques but also offer prospective resolutions to address the outstanding predicaments in NILM, thereby contributing to the pursuit of a trustworthy NILM approach characterized by accuracy, resilience, dependability, transparency, and impartiality. The employment of Generative Adversarial Networks (GANs) in Non-Intrusive Load Monitoring (NILM) has been investigated in recent studies. A GAN-based framework was utilized initially to address NILM, as documented in [94]. Kaselimi et al. [95] investigated the GAN application aimed at facilitating sequence-to-sequence learning. Similarly, Pan et al. [96] also offer their proposed approach to the same problem. The researchers successfully implemented the utilization of conditional GANs for sequence-to-sub-sequence learning.

Chen et al. [97] utilized a context-aware convolutional network for energy disaggregation, which has undergone adversarial training. Numerous findings rely on the ability of the adversarial training procedure to withstand noise. Furthermore, as an advancement to Kaselimi et al. [95] work, the inclusion of two Gated Recurrent Units into the D network was proposed and implemented by Kaselimi et al. [98]. To conduct or carry out discrimination. A Conditional Generative Adversarial Network (cGAN) has been developed for energy disaggregation. Finally, a novel method entitled GAN-NILM [99] is presented herein. The network comprises an autoencoder, denoted as G, responsible for generating the appliance-specific signal. The authors concatenated the main signal with the output of G as a method of data analysis. The technique in question is purported to guarantee stability in training. The discriminator's objective is to distinguish whether or not the disaggregated signal is authentic. This study assessed the efficacy of the methodology by conducting tests on three datasets that are readily accessible to the public.

The use of Gaussian Processes (GPs) for NILM has been studied in recent years. Desai et al. [100] explored the application of sparse GPs for energy disaggregation. The authors modelled three variants of GPs on the REDD dataset where statistical features based on a sparse GP model performed best in detecting the ON states of the target appliance. However, the top two models failed to detect off-states and required further improvements. Moomal et al. [101] used Gaussian Mixture Models (GMM) to identify the states of different appliances in an unsupervised learning manner. The authors used the Bayesian information criterion (BIC) to correctly identify specific events from the total observed events. The proposed approach performed best in disaggregating energy consumption of two-state appliances compared to the mean-shift clustering method.

Recently, there have been attempts to use transformer models as a substitute architecture for neural machine translation tasks. This trend has persisted for approximately three years [102]. A transformer-based architecture that employs self-attention for energy disaggregation has recently been implemented by [103] to manage sequential power signal data. Most extant deep learning models for energy disaggregation use a singular target learning method, whereby a neural network is solely trained for each appliance. In contradiction to a univariable learning paradigm, the research of [104] advances the notion of Unet-NILM for identifying various household appliances and estimating their energy consumption by employing a multi-label learning framework and multi-target quantile regression. The Unet-NILM is a convolutional neural network designed using the U-Net architecture first introduced for image segmentation. The notion of Explainable AI (XAI) strives to

Table 3
Summary of stand-alone deep learning-based approaches for NILM.

Ref.	Year	Dataset	Features	NILM Algorithm	Evaluation Metrics
[109]	2019	PLAID	Active Power, Reactive Power, Apparent Power	FF & CNN	Precision, Recall, F1-score, Transient Harmonics
[110]	2019	REDD	Active Power	RNN	Precision, Recall, F1-score, Accuracy
	2019	AMPds	Active Power	RNN	Accuracy
[111]	2019	AMPds2	Active Power	Bayes- Bidirectional-LSTM	Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Normalized RMSE
[75]	2020	AMPds, REFIT, REDD	Active Power	Context-aware Bidirectional-LSTM	Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Normalized RMSE
[104]	2020	UK-DALE	Active Power	Unet	Estimated Accuracy (EAC), Normalized Disaggregation Error (NDE), MAE, F1-score
[112]	2020	REDD, ENERTALK	Active Power, Reactive Power, Apparent Power, Voltage, Current, Power Factor, and combined features of the above.	CNN	Normalized Error in Assigned Power (NEAP), NDE, F1-score
[113]	2020	SynD, UK-DALE, REFIT	Active Power	2D-CNN	MAE, NDE
[114]	2020	REDD, UK-DALE, REFIT	Active Power	Pruned CNN	MAE, MSE
[115]	2020	REDD, UK-DALE	Active Power	VGG-16	Accuracy, Precision
[116]	2020	BLUED	Current	CNN	Precision, Recall, F1-score, False Positive Rate,
[117]	2020	UK-DALE, ECO	Active Power, Apparent Power, Reactive Power, Current, Power Factor	Bidirectional-LSTM	Signal Aggregate Error (SAE), Estimation Accuracy, F1-score
[118]	2020	UK-DALE	Active Power	AE	Precision, Recall, F1-score, MAE, SAE
[119]	2020	REDD	Active Power	CNN & Clustering	Precision, recall, F1-score
[120]	2020	UK-DALE, REFIT, Dataport, DRED	Active Power	CNN	MAE, F1-score
[121]	2020	UK-DALE, REDD, DRED	Active Power	CNN	NDE, F1-score
[122]	2020	UK-DALE	Active Power	ResNet	MAE, F1-score
[108]	2021	REFIT	Active Power	CNN	MAE, SAE, F1-score
[123]	2021	REDD, UK-DALE	Active Power	ResNet	MAE, SAE, Precision, Accuracy, F1-score
[124]	2021	REDD, UK-DALE, REFIT	Active Power	LSTM	Precision, Accuracy, F1-score, MAE, SAE
[125]	2021	REDD	Active Power	CNN	MSE, RMSE
[126]	2021	REDD	Active Power	CNN	Precision, Recall, Accuracy, F1-score
[127]	2021	HIPE, IMDELD, REDD	Active Power	CNN, DAE, RNN, GRU	MDE, SAE
[128]	2021	REDD, ECO, UK-DALE	Active Power	CNN	MAE, F1-score, Overlapping Energy, Extra Energy, Missed Energy
[129]	2021	PLAID	Voltage, Current	CNN	Macro-Accuracy, Macro-F1-score
[130]	2022	PLAID	Voltage, Current	Bayesian Neural Network	F1-score
[131]	2022	Low Voltage Network Simulation (LVNS)	Active Power, On-Off Events	Dictionary learning (Sparse coding)	Precision, Recall, Accuracy, F1-score
[107]	2023	REFIT	Active Power	WaveNet-based DNN	F1-score
[132]	2023	REDD, REFIT, UK-DALE	Active Power	Sequence-to-Point with Adaptive Domain Transfer Learning	Overall Disaggregation Proportion Error (ODPE)
[133]	2023	PLAID, WHITED, HRAD	2D Load Signatures	Residual CNN	Precision, Recall, Accuracy, F1-score

establish a perceptibly more lucid and trustworthy AI by devising techniques that render the operational mechanics and predictive outputs of machine learning systems comprehensible to human cognition without incurring any detrimental impacts on system effectiveness [105]. The present study aims to elucidate the Explainable Non-Intrusive Load Monitoring (NILM) networks introduced by [106], which strive to comprehend the internal mechanisms of machine learning models employed in NILM. Understanding the importance of deep learning-based approaches and realizing their challenges, Todic et al. [107] proposed an active learning framework to learn and update the deep learning-based NILM model from a small amount of data and transfer its disaggregation capabilities in a new environment. The authors explored the BatchBALD method to choose a diverse batch of samples for training the Wave-net NILM-based deep learning model [108]. A detailed summary of stand-alone deep learning and advanced approaches is presented in Tables 3 and 4, respectively.

A comparison of optimization methods, traditional machine learning methods, and advanced deep learning methods in terms of their advantages, disadvantages, and data requirements is presented in Table 5 below. A comparison of these methods in terms of real-time performance (inference time) and interpretability is shown in Fig. 7. Despite the high number of publications utilizing advanced deep learning, these methods are very hard to interpret and have a high inference time. Because of these reasons, advanced machine learning methods are not preferable for real-time employment. Instead, traditional machine learning and deep learning methods must be explored from a real-time deployment point of view.

As Fig. 8 shows, deep neural networks (DNN) have gained more attention recently. They are the preferable choice among researchers for energy disaggregation. Among various types of deep neural networks, convolutional neural networks (CNN) are the most widely used learning method because of their light architecture and fast and accurate prediction. Another essential aspect of CNN's popularity is its multiple variants, such as sequence-to-sequence, sequence-to-point, dilated

Table 4
Summary of advanced deep learning-based approaches for NILM.

Ref.	Year	Dataset	Features	NILM Algorithm	Evaluation Metrics
[95]	2020	AMPds2, REFIT	Active Power	GAN	MAE, RMSE, NRMSE
[98]	2020	AMPds2, REFIT	Active Power	Hybrid GAN-CNN-GRU	MAE, RMSE, NRMSE
[103]	2020	REDD, UK-DALE	Active Power	Transformer	Mean Residual Error (MRE), MAE, F1-score, Accuracy
[99]	2020	REFIT, REDD, UK-DALE	Active Power	GAN	F1-score
[134]	2020	REDD	Active Power	Multi-Head Attention Network	Accuracy, Signal Aggregate Error (SAE), F1-score, MAE
[135]	2020	Hospital Electric Power Demand Data	Active Power, Voltage, Current, High-Frequency Harmonics, and Total Harmonic Distortion (THD)	FCN-dAE	RMSE, NEAP
[136]	2020	REDD, UK-DALE	Active Power	Attention-based CNN-GRU	Precision, recall, F1-score, Relative Error in Total Energy (RETE), MAE
[96]	2020	UK-DALE, REFIT	Active Power	GAN	F1-score, MAE, Accurasy
[137]	2021	REFIT	Active Power	Hybrid Multitask CNN-GRU	Precision, Recall, F1-score, MAE
[138]	2021	REDD, UK-DALE	Active Power	Attention	F1-score, MAE, SAE
[139]	2021	REDD	Active Power	Hybrid LSTM-BP	Precision, Recall, F1-score
[140]	2022	UK-DALE	Active Power	Multi-head CNN	Precision, Recall, Accuracy, F1-score, MAE, SAE
[141]	2022	REDD, UK-DALE	Active Power	Attention-based LSTM, Attention-based Sequence-to-Point Network	Precision, Recall, Accuracy, F1-score, MAE, RETE
[142]	2022	REDD, AMPDs	Active Power	Hybrid LE-CRNN	Accuracy
[143]	2023	PLAID, REDD	Voltage, Current	Hybrid SE-ResNet	Precision, Recall, Accuracy, F1-score
[144]	2023	REDD, UK-DALE	Active Power	Conditional Diffusion Model (DiffNILM)	Precision, Recall, Accuracy, F1-score
[145]	2023	REDD, UK-DALE	Active Power	Domain Adversarial Neural Network (DANN)	F1-score

convolutions, residual networks, and hybrid architectures. Hidden Markov Models (HMM) are the second most widely used methods in recent years because of their inherent capability to model uncertainty and achieve accuracy comparable to DNNs. RNNs and LSTM combined are the third most commonly used methods in NILM. These networks have a proven ability to deal with sequential data, which is why these networks and their variants have been applied in NILM research for many years.

5. Datasets and performance evaluation metrics

5.1. Datasets

The current surveys on NILM suggest that it is crucial to choose the appropriate dataset. Huber et al. [21] briefly described the most important features of the publicly available datasets. Also, in [146], authors compared different datasets that are open source. Table 6 provides information about the most commonly used datasets for NILM, including details about their features and attributes. It is important to note that the selection of a dataset must be closely linked to the chosen NILM method. This is because a dataset has limitations, which can stop specific machine-learning models from being used. For instance, according to Table 6, the AMPds [70] dataset records data every 1 min. This means that the dataset is unsuitable for applications that require a lot of data processing but can only be used for applications that do not require a significant amount of data processing. On the other hand, the BLOND [147] and EMBED [148] datasets are useful for models that work quickly with lots of data. Another characteristic is the total time the data covers in each dataset. The biggest ones last for a few years, like REFIT [149], DEDDIAG [150], and IDEAL [151]. The smaller ones, like the REDD dataset, only have a few days of data. Moreover, most datasets only include a small number of houses.

The most data-driven methods for NILM need labeled data (individual power consumption data of an appliance) to train them. However, getting this data is a difficult task that slows the development of reliable and widely applicable home energy monitoring systems. The exceptions are the HES dataset [152], the PLAID dataset [153], and the IDEAL dataset [151]. Nowadays, machine learning has advanced, and generative adversarial networks (GANs) are used to create appliance power signatures [154,155]. Similarly, SynD [156] is a type of synthetic dataset that can generate synthetic aggregate power and individual power consumption profiles of specific appliances at any date, time range, and resolution. The online access links for these datasets are provided in Appendix A1.

5.2. NILM metrics for performance evaluation

Until now, there has not been a clear agreement or standard on sharing accuracy results in NILM. However, a few research studies give a general understanding of how to measure performance in NILM. The noteworthy ones are identified as [36,146,157–159]. NILM system needs to report measurements as overall disaggregation scores for the entire household and individual scores for each appliance. The household-level measurements demonstrate how accurate the model is in breaking down the total power signal into different signals. This evaluation can be helpful when many houses are combined into one NILM model. In this situation, it is essential to check how well the model works for individual houses and whether it can be used for other houses. This can be done by analyzing the model's performance in each house separately.

Additionally, it is crucial to provide information about how each appliance performs (measured at the appliance level) to understand the pros and cons of various NILM algorithms. With this more detailed information, it is possible to develop a NILM system that chooses different methods based on the load separation task's situation (such as specific past events). Because there are many ways to separate loads, there is a need to evaluate how well they classify and predict so that an accurate comparison can be made. Therefore, when it comes to NILM, two types of metrics are used. The first type measures how well the model can detect when an appliance turns on and off. These metrics are called classification metrics. The second type assesses how accurately the model can break down the total power signal into individual appliance power signals [146]. These metrics are called regression metrics.

Table 5
Comparison of learning methods applied to NILM.

Methods	Examples	Pros	Cons	Data Requirements
Optimization Methods	CO, GSP	Simple and easy to implement.	Low accuracy due to loss of temporal continuity in the presence of a large number of appliances.	Can work with any data. No specific requirements.
Probabilistic Methods	HMM and its variants, GMMs	Can achieve similar accuracy as DNNs but with less data;	Computational complexity increases, and accuracy decreases as number of appliances increases.	Can work with both labelled training data only.
Traditional Machine Learning	SVM, ANN, k-NN, RF, XGBoost, Naïve-Bayes	Lower computational complexity and easy to implement.	Hand crafted feature extraction; Requires expert knowledge; limited accuracy.	Can work with both labelled and unlabelled data; require high input features.
Deep Learning	RNN, LSTM, Bi-LSTM, CNN, GRU, AE	Automatic feature extraction from input data; highly accurate compared to traditional machine learning and optimization methods; Can disaggregate a large number of appliances.	Accuracy is dependent on data quality and quantity; difficult hyperparameter tuning process; generalization capability is the biggest challenge.	Requires large, clean, and labelled training data.
Advanced Deep Learning	Attention-based methods, Transformers, GAN, DANN, Hybrid DNNs,	Automatic feature extraction from input data; highly accurate compared to all other methods; Can disaggregate a large number of appliances.	Computationally expensive; hard to implement.	Requires large, clean, and labelled training data.

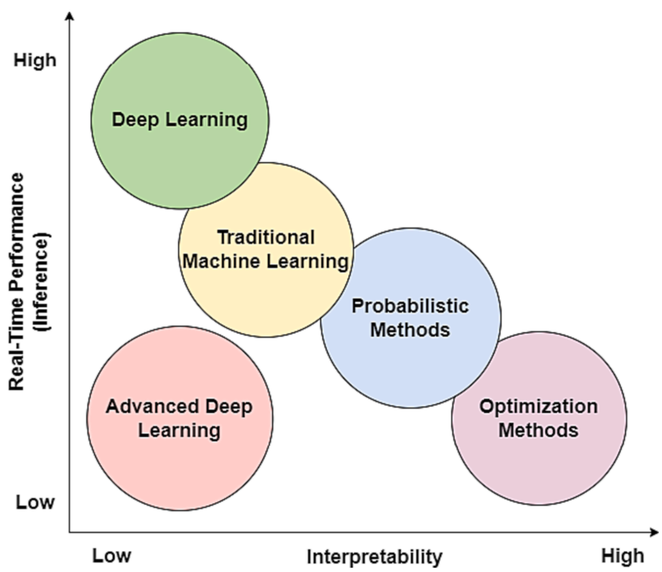

Fig. 7. Comparison of data-driven NILM methods on real-time performance and interpretability.

5.2.1. *Classification metrics*

The classification-based metrics are used to assess how well event-detection algorithms perform. This approach's performance evaluation focuses on checking how well the method can accurately estimate whether an appliance is on or off. In this case, accuracy, precision, recall, and the F1-score are the most common ways to measure performance [157,158].

$$Accuracy = \frac{Numberofcorretlyidentifiedevents}{Totalnumberofevents} \times 100 \tag{2}$$

$$precision = \frac{TP}{TP + FP} \tag{3}$$

$$recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{5}$$

In particular, Pereira et al. [158] analyzed experimentally the behavior of eighteen different performance metrics applied to event-detection algorithms. Their study found that the behavior of performance metrics used to evaluate NILM algorithms is similar to the behavior exhibited when applied to other domains. However, they also found a few differences from the NILM problem's unbalanced nature. They also suggested that performance evaluation should be done on scenarios closer to real-world situations.

5.2.2. *Regression metrics*

The performance evaluation of regression-based approaches is done

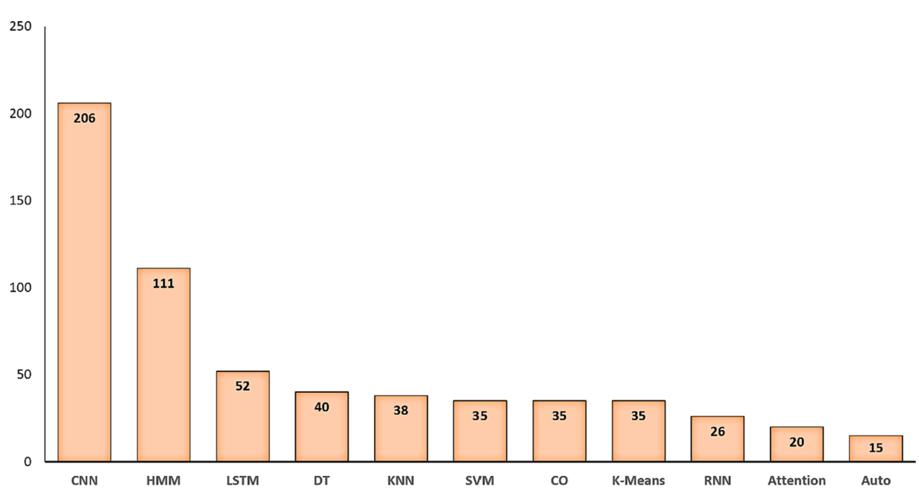

Fig. 8. Number of publications from 2018 to 2023 using deep learning and machine learning methods.

Table 6
Summary of publicly available datasets.

Dataset Name	Year	Country	No. of Houses	Duration (months)	Measurements*	Aggregate Data Sampling Frequency
REDD	2011	US	6	5	V, I	1 Hz, 15 kHz
BLUED	2011	US	1	8 days	V, I	12 kHz
AMPds	2013	CA	1	12	V, I, P, Q, S, P.F.	1 min
BERDS	2013	US	1	12	P, Q, S	20 sec
iAWE	2013	IN	1	2.5	V, I, A, Q, S	1 sec
DRED	2014	NL	1	6	P	1 Hz
ECO	2014	CH	6	8	P, Q	1 Hz
GREEND	2014	IT/AT	9	12	P	1 sec
PLAID	2014	US	60	6	V, I	30 kHz
REFIT	2015	UK	20	24	P	8 sec
UK-DALE	2015	UK	5	Up 30	V, I	6 sec, 16 kHz
COOLL	2016	FR	1	1	V, I	100 kHz
BLOND	2018	DE	1	213 days	V, I	50 kHz
EMBED	2019	US	3	14–27 days	P, Q	1 Hz
ENERTALK	2019	KR	22	29–122 days	P	15 Hz
SynD	2019	AT	1	6	P	5 Hz
DEDDIAG	2021	DE	15	<40	P	1 Hz
IDEAL	2021	UK	255	<24	P	1 sec

*Active: P, Reactive: Q; Apparent: S; Voltage: V; Current: I; Power Factor: P.F.

to see how well a method works by comparing the actual appliance signal with the predicted estimates. Three commonly used metrics are mean absolute error (MAE), signal aggregate error (SAE), and estimation accuracy (EA). The study in [111] proposed a Bayesian-optimized bidirectional LSTM network to predict the energy usage of a dishwasher in the AMPds dataset. Their proposed solution achieved minimum MAE in disaggregating dishwashers. The lowest MAE error for washing machines was achieved in [51]. Murray et al. [73] proposed a convolution neural network that performs really well in the fridge category of the REFIT dataset. Regarding the microwave and kettle appliances, Pan et al. [96] proposed conditional GANs to predict the power usage of these appliances in the UK-DALE dataset that achieved minimum MAE. Therefore, it is evident that disaggregation performance highly depends on the type of appliance.

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_t - \widehat{y}_t| \tag{6}$$

$$SAE = \frac{|E_p - E_g|}{E_g} \tag{7}$$

$$EA^k = 1 - \frac{\sum_{t=1}^{T}\left|\widehat{y}_y^k - y_t^k\right|}{2 \times \sum_{t=1}^{T} y_t^k} \tag{8}$$

Rafiq et al. [128] proposed energies-of-interest (EOI) based evaluation metrics to get more insights on regression-based algorithm performance. Their proposed evaluation metrics used ground-truth energy and predicted energy of an appliance to calculate overlapping, missing, and extra energy. Knowing these parameters helped to understand how accurately the algorithm estimated the energy consumption of a target appliance.

5.2.3. Comparison of NILM metrics for performance of different studies

The results achieved through each of the NILM Metrics for Performance Evaluation explained in the previous sections vary due to several reasons, including:

- The type of appliance being disaggregated
- Number of test data samples
- Threshold values (used to calculate Precision, Recall, and F1 score)
- Test scenarios (e.g., seen data, unseen data, and unseen house)
- Whether the test data is noised or denoised (noised data refers to the amount of un-metered load in the test data).

Therefore, to compare evaluation scores from different studies, it must be ensured that all NILM methods are evaluated on the same number of test data samples, the same test scenarios, the same percentage of noise in the test data, and the same threshold value. However, as explained before, there is no agreement or unified criteria for the NILM Metrics for Performance test conditions. That makes comparing the values found in different studies difficult or impossible.

To overcome this situation, the authors suggest the development of a standard to evaluate the NILM methods, which includes standardized testing scenarios, test data, and specific evaluation metrics. That will permit fair comparisons and identify which metrics performed best according to the established test conditions.

6. Challenges in NILM

In Section 1, it was discussed that the real-world application of a NILM system is not only cost-effective but also beneficial in reducing building energy consumption. Most research on NILM focuses on improving NILM algorithms for event detection, appliance classification, and power consumption estimation. Although this is an essential line of research, to take advantage of NILM technology, those algorithms must be robust and applicable in the real world. According to the already established technology of smart meters, Zeifman [160] proposed six requirements for NILM systems to be applicable in the real world:

1) Feature selection: Most smart meters and the lower-cost solution demand a sampling rate of 1 Hz. This requirement refers to the data acquisition and feature extraction parts of the NILM framework because the frequency of data collection affects the features that can be extracted.
2) Accuracy: There are indications that an acceptable user experience would require a minimum accuracy of 80–90 %.
3) No training: For a seamless user experience, the user should put minimum effort into configuring and training the system. Then, the system should be able to detect new, unseen, and discard unused devices.
4) Near real-time capabilities: The system should work online and give immediate and accurate feedback regarding the current energy status of the house.
5) Scalability: The system should be robust and efficient even for more complex environments (e.g., more than 15 devices).

Table A1
Summary of publicly available datasets with access links.

Dataset Name	Country	Link to the Dataset
REDD	US	https://redd.csail.mit.edu/
BLUED	US	https://portoalegre.andrew.cmu.edu:88/BLUED/
AMPds	CA	https://dataverse.harvard.edu/dataset.xhtml?persistentId = https://doi.org/10.7910/DVN/FIE0S4
BERDS	US	https://people.eecs.berkeley.edu/~maasoumy/
iAWE	IN	https://iawe.github.io/
DRED	NL	https://www.st.ewi.tudelft.nl/~akshay/dred/
ECO	CH	https://www.vs.inf.ethz.ch/res/show.html?what = eco-data
GREEND	IT/AT	https://sourceforge.net/projects/greend/
PLAID	US	https://figshare.com/articles/dataset/PLAID_-_A_Voltage_and_Current_Measurement_Dataset_for_Plug_Load_Appliance_Identification_in_Households/10084619
REFIT	UK	https://pure.strath.ac.uk/portal/en/datasets/refit-electrical-load-measurements(31da3ece-f902-4e95-a093-e0a9536983c4).html
UK-DALE	UK	https://www.doc.ic.ac.uk/~dk3810/data/
COOLL	FR	https://coolldataset.github.io/
EMBED	US	https://embed-dataset.org/
ENERTALK	KR	https://springernature.figshare.com/collections/The_ENERTALK_Dataset_15_Hz_Electricity_Consumption_Data_from_22_Houses_in_Korea/4502780
SynD	AT	https://springernature.figshare.com/collections/SynD_A_Synthetic_Energy_Dataset_for_Non-Intrusive_Load_Monitoring_in_Households/4716179
DEDDIAG	DE	https://figshare.com/articles/dataset/DEDDIAG_a_domestic_electricity_demand_dataset_of_individual_appliances_in_Germany/13615073
IDEAL	UK	https://datashare.is.ed.ac.uk/handle/10283/3647

6) Various appliance types. The system should recognize the four types of electrical devices: a) on/off, b) finite-state, c) variable power/multi-state, and d) permanent consumer.

Another equally important requirement of a NILM system is the ability to generalize to unseen houses. In the real world, it is improbable that ground truth appliance data will be available for each house. Therefore, algorithms developed on one dataset should be able to estimate the consumption of appliances on any unseen data.

Various approaches have been proposed to solve the NILM problem, as presented in Section 4. Some of the most successful ones exploit deep learning neural network structures for modelling an energy disaggregation problem. Nevertheless, NILM is a challenging task. The authors hereby provide an indicative list of the NILM challenges based on our current understanding of the field. Some of them are well-studied, whereas others are immature, and there is ongoing research on these topics:

1) **Detection and power estimation of the same type of appliances:** Most households have multiple appliances of the same kind. For example, it is common for large households to have more than one refrigerator or air conditioning unit. However, detecting and estimating the power consumption of multiple appliances of the same type when they are ON simultaneously is challenging for current NILM algorithms. Few works have tried to address this issue by developing individual models for each type of appliance. Doing so requires prior knowledge of different appliance types and their count in a house. However, when multiple appliances of the same type are present in aggregate data, the disaggregation model mispredicts appliance activations. Similarly, NILM algorithms also suffer from disaggregating appliances with similar consumption patterns or activations. For instance, deep learning models trained individually for each target appliance tend to mispredict appliances (detect an appliance when it is OFF) because of overlapping characteristics of similar appliances. Gopinath et al. [161] have tried to address this challenge using deep neural networks trained on high-frequency sampled data (10 kHz, 25 kHz, 50 kHz). However, accurately detecting similar or multiple appliances of the same type on low-frequency sampled data is still an open problem and requires active attention.
2) **Detection and power estimation of appliances on Max operation:** It has been noticed in literature that learning methods struggle to correctly estimate the consumption of target appliances even if these appliances are being used in regular operation. There are situations where certain appliances, such as air conditioners and irons, are run in maximum operation. During this state, appliances draw the highest power and consume maximum energy. Some appliances' patterns change when they are used at their maximum capacity. If that is the case, learning methods may fail to recognize them. As a result, mispredictions can be made by the learning method. This fact is an open problem, and the authors have not found a single paper in the literature discussing this challenge.
3) **Generalization ability of the algorithms**: Many advanced techniques have not worked well when evaluated on houses whose data has not been part of the training [73]. So, it is hard to make reliable algorithms with high transferability. Another thing to consider is to create strong models that handle datasets with high noise and appliances that behave strangely. Peak aggregate energy readings make NILM techniques work worse. Furthermore, a common issue in NILM is that the appliances being targeted have irregular patterns or display abnormal actions. In addition to these problems, the models do not work well because they do not have enough information to learn from [162]. To increase the transferability of NILM algorithms or to make them generalized, they must be trained on a large amount of data. To solve this challenge, researchers have introduced algorithms that generate synthetic data [128,156]. However, as explained before, few studies used automatic labelling of appliance activations to increase the amount of labelled data [159].
4) **Incorporating the user's feedback and seasonality in algorithms:** Consumers' behaviors and varying seasons significantly impact energy usage, making it harder to monitor energy usage. Many different things can affect the operation of household appliances. Various factors like the environment around them and the socio-economic situation of the people using them can make a difference. Feedback from users is essential to improve the models' accuracy in NILM algorithms. Modern NILM methods need to be continuously updated and enhanced according to the suggestions and feedback provided by users.
5) **Explainability of NILM models (the reasoning behind the model estimations):** Although the models recently proposed in the literature claim to be more accurate than the previous ones, their workings are not easily understood or replicable. Understanding and believing in the results of the learning-based algorithms can help improve designs and show important features of the data used to make decisions. It can also give a better idea of how accurate the model is and build trust in the feedback given to the user.
6) **Fairness in NILM:** Different factors, such as money, social status, and the environment, influence how people use energy. Because of this, different data patterns (like social and geographic groups) should be represented and studied in a NILM framework. Therefore,

the NILM AI framework learns how to accurately predict results based on real data from the tested appliance. It also ensures fairness by using a set of predetermined measures and having enough different training data. Aside from using accurate data and models, it is also essential to evaluate the performance of NILM fairly and adequately to establish benchmarks.

7. Measures for improvement

After identifying and highlighting relevant research works and current challenges in NILM, as the final step, the study includes the potential measures for improvement of issues related to:

1) The data pre-processing and feature selection.
2) The NILM model implementation in real-time.
3) The recommended model evaluation practices are discussed.

7.1. Data pre-processing and feature selection

The data collected from smart meters' energy consumption could be used to make a successful NILM algorithm. Assuming a supervised NILM system, the users collect extra information about how much power individual appliances consume in addition to the aggregate power measurements. Extracting valuable features from the data and using them to train models can make the models work better. For this purpose, the optimal feature extraction framework (discussed in Section 3.2.3) can also be utilized. The most common additional input variables reported in the literature are:

1) The spatial–temporal information (e.g., household's location and time period).

2) The events (e.g., holiday information, weekday-weekend activity information, appliances availability, roof-top PV panels information).

3) The socio-economic characteristics of a household (e.g., the total number of occupants, number of bedrooms, house area, and air conditioning type).

Adding more variables/features to the model makes it work better. However, in most cases, only a few of these features are used. Selecting the best input features for a machine learning-based NILM model is challenging as it affects the performance of the model and its computational efficiency, as discussed in [43]. Some studies recommend eliminating unnecessary features as the first step to determining the most useful ones to solve these problems. However, choosing the best features is limited by the data availability, and it depends on how the person responsible plans to identify the appliances [39]. Regarding data collection and sampling frequency, a rough division of 1-second is typically good enough to distinguish between low-frequency and high-frequency data features. Although high-frequency datasets are better at differentiating between various types of appliances, low-frequency datasets are becoming increasingly popular in deep learning-based NILM algorithms [21].

Another crucial step to making a successful NILM algorithm is to address the problem of missing data, which can happen when there are metering or data transmission issues. Moreover, unusual data points (large spikes) that occur because of sudden changes, increases, and irregularities in the load are also seen in the data. The K-means clustering method might be able to fix these problems. Density-based clustering (such as the OPTICS) can handle outliers. Still, it may not work well on big datasets because of issues with numbers. Secondly, filling up the data using the appropriate methods is very important and requires understanding the data and processes. Interpolation methods (linear or non-linear) are commonly used to fill in small gaps in the aggregate data. Depending on the number of missing data points, previous week/month/year-same day-same time data can also be used to fill up missing data points.

7.2. NILM model implementation

The effectiveness of NILM methods in real-world scenarios depends on the dataset used for training and the specific type of appliance it was developed for. The NILM algorithm proposed in [96] works well for microwaves and kettles. However, when it comes to dishwashers and washing machines, there are different ways to make them work even better. Usually, even though a specific method may work well for one appliance, it may not work well for other appliances. Therefore, there is no perfect method for dealing with all types of appliances. No matter the machine-learning method used, researchers working on NILM should remember that the most critical things in this field come from the real-life challenges faced. Machine learning and deep learning are constantly developing more algorithms for advanced problems. However, these algorithms are becoming more and more complicated. It is not easy to create advanced NILM algorithms that are efficient, practical, and meet the requirements for trustworthiness and explainability. The new transformer models used for NILM need a lot of computer power. When continuous learning is added to these models, it makes the computational problems even more difficult. Therefore, it is necessary to develop simple yet accurate models that can easily be deployed in real-time. Similarly, machine learning or deep learning-based NILM models must be continually re-trained to identify new or unknown appliances accurately.

7.3. NILM model evaluation

To understand the evaluation results correctly, it is essential to choose the right performance evaluation metrics from a wide range available in the literature (refer to section 5.2). It is important to remember that the accuracy metric results should be normalized. Normalizing the results helps readers understand how each appliance compares to others and how accurate they are overall [146]. Also, NILM researchers must show classification and regression measurements in their studies. Still, there is no need to provide exact scores for appliance states or their power estimation. It is also recommended to evaluate the models on different testing scenarios. These scenarios could be evaluating NILM models on:

1) Unseen data from the same house of the same dataset.
2) Unseen data from another house of the same dataset.
3) Unseen data from another dataset (cross-domain).
4) Noised data (presence of unknown loads in the test data).
5) De-noised data (presence of target loads in the test data).

Once results are analyzed based on these testing scenarios, it helps to understand NILM model capabilities, such as generalization and detecting similar or unknown appliances.

Furthermore, performance evaluation metrics have shown similar behavior when applied to the NILM problem or other problems (Medical Diagnosis or Image Recognition); however, few differences have been reported in the literature because of the unbalanced nature of the NILM problem. Therefore, domain-specific evaluation metrics, or those that have been introduced specifically to address energy disaggregation, are highly recommended to apply while evaluating NILM models.

8. Limitations of the review

One limitation of the review is its exclusive focus on NILM methods designed for residential buildings and households. The decision to limit the review work to the residential sector only was based on many factors, such as the benefits of real-time load monitoring, energy-saving targets, and the applicability of these approaches. However, the absence of coverage on commercial and industrial building applications limits the paper's applicability across diverse contexts. Furthermore, the review focuses on NILM algorithm frameworks and learning methods.

Therefore, it does not include specific discussions on data augmentation, appliance activation extraction, data post-processing, and NILM algorithm evaluation.

9. Conclusion

This paper comprehensively reviews the most recent NILM methods applied to the residential sector. One of its significant contributions is the discussion of the challenges concerning NILM applicability, which has not been addressed in prior research. Although advanced deep learning-based NILM methods have demonstrated promising results in terms of disaggregation accuracy and generalization under laboratory conditions, their effectiveness in real-time scenarios (i.e., actual buildings) is questionable due to their computational inefficiency and implementation difficulty. Also, the evaluation scenarios described in published articles are biased, as these methods have not been thoroughly assessed on unseen data and houses, which raises concerns about their generalizability.

This review also analised the challenges associated with the accuracy of deep learning methods in detecting multiple same-type appliances and appliances under maximum operation. Additionally, it discussed generic challenges, such as incorporating user feedback, data seasonality, and explainability of NILM methods.

Potential improvement measures were also discussed, including how to tackle the challenges related to the applicability of NILM. For example, supervised learning-based NILM methods require balanced training data, which synthetic data generation algorithms can facilitate. Furthermore, a comprehensive evaluation process must be conducted to test the effectiveness and limitations of the proposed solution across a range of scenarios, including accuracy, generalization, and scalability.

Lastly, for a successful and effective deployment of NILM in the residential sector, future research work should focus on developing computationally efficient and trustworthy algorithms.

CRediT authorship contribution statement

Hasan Rafiq: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Prajowal Manandhar:** Writing – original draft, Methodology, Investigation. **Edwin Rodriguez-Ubinas:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation. **Omer Ahmed Qureshi:** Writing – original draft, Investigation. **Themis Palpanas:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendix

References

[1] P. Friedlingstein, et al., Global carbon budget 2022, Nov. 2022. doi: 10.5194/essd-14-4811-2022.
[2] R. Lindsey, Climate change: atmospheric carbon dioxide.
[3] A. Savaresi, The Paris agreement: a new beginning? J. Energy Nat. Resour. Law 34 (1) (2016) 16–26, https://doi.org/10.1080/02646811.2016.1133983.
[4] IEA, CO2 Emissions in 2022, https://www.iea.org/reports/co2-emissions-in-2022.
[5] IEA, Buildings, https://www.iea.org/reports/buildings.
[6] E. Trepci, E. Rodriguez-Ubinas, Z. AlShikh, and S. Sgouridis, "Passive Design Strategies for Zero Energy Houses in Desert Environments: The Case of Solar Decathlon Middle East 2021," Aug. 2022, pp. 139–151. doi: 10.2495/SDP220121.
[7] O.A. Qureshi, E. Trepci, A.A. Shahzad, P. Manandhar, E. Rodriguez-Ubinas, Quantitative assessment of the HVAC system of zero-energy houses of the Solar Decathlon Middle East 2021, Energy Rep. 9 (Sep. 2023) 1050–1060, https://doi.org/10.1016/j.egyr.2023.05.042.
[8] I. E. C, Coping with the Energy Challenge The IEC's role from 2010 to 2030, no. September. Geneva, Switzerland: International Electrotechnical Commission, 2010.
[9] D. Larcher, J.M. Tarascon, Towards greener and more sustainable batteries for electrical energy storage, Nat. Chem. (2015), https://doi.org/10.1038/nchem.2085.
[10] I. Vassileva, E. Dahlquist, F. Wallin, J. Campillo, Energy consumption feedback devices' impact evaluation on domestic energy use, Appl. Energy 106 (2013) 314–320, https://doi.org/10.1016/j.apenergy.2013.01.059.
[11] J. Laitner, K. Erhardt-Martinez, "Examining the Scale of the Behaviour Energy Efficiency Continuum", in People-Centred Initiatives for Increasing Energy Savings, American Council for Energy Efficient Economy, Washington D.C., 2010, pp. 20–31.
[12] K. Carrie Armel, A. Gupta, G. Shrimali, A. Albert, Is disaggregation the holy grail of energy efficiency? The case of electricity, Energy Policy 52 (2013) 213–234, https://doi.org/10.1016/j.enpol.2012.08.062.
[13] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, S. Santini, The ECO data set and the performance of non-intrusive load monitoring algorithms, in: BuildSys 2014 - Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, 2014, https://doi.org/10.1145/2674061.2674064.
[14] G.W. Hart, Nonintrusive appliance load monitoring, Proc. IEEE 80 (12) (1992) 1870–1891, https://doi.org/10.1109/5.192069.
[15] J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. Reynolds, S. Patel, Disaggregated end-use energy sensing for the smart grid, IEEE Pervasive Comput. 10 (2011) 28–39, https://doi.org/10.1109/MPRV.2010.74.
[16] C. Dinesh, S. Makonin, and I. V. Bajic, "Residential Power Forecasting Using Load Identification and Graph Spectral Clustering," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 66, no. 11, 2019, doi: 10.1109/TCSII.2019.2891704.
[17] P. Manandhar, H. Rafiq, E. Rodriguez-Ubinas, Current status, challenges, and prospects of data-driven urban energy modeling: A review of machine learning methods, Energy Rep. 9 (Dec. 2023) 2757–2776, https://doi.org/10.1016/J.EGYR.2023.01.094.
[18] H. Rafiq, P. Manandhar, E. Rodriguez-Ubinas, J.D. Barbosa, O.A. Qureshi, Analysis of residential electricity consumption patterns utilizing smart-meter data: Dubai as a case study, Energy Build. 291 (Jul. 2023) 113103, https://doi.org/10.1016/j.enbuild.2023.113103.
[19] L. Pereira and N. Nunes, "Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—A review," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 8, no. 6. 2018. doi: 10.1002/widm.1265.
[20] R. Bonfigli and S. Squartini, Machine Learning Approaches to Non-Intrusive Load Monitoring. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-30782-0.
[21] P. Huber, A. Calatroni, A. Rumsch, and A. Paice, "Review on deep neural networks applied to low-frequency nilm," Energies, vol. 14, no. 9. 2021. doi: 10.3390/en14092390.
[22] G.-F. Angelis, C. Timplalexis, S. Krinidis, D. Ioannidis, D. Tzovaras, NILM applications: Literature review of learning approaches, recent developments and challenges, Energy Build. 261 (Apr. 2022) 111951, https://doi.org/10.1016/j.enbuild.2022.111951.
[23] S. Desai, R. Alhadad, A. Mahmood, N. Chilamkurti, and S. Rho, "Multi-state energy classifier to evaluate the performance of the NILM algorithm," Sensors (Switzerland), vol. 19, no. 23, 2019, doi: 10.3390/s19235236.
[24] A. Zoha, A. Gluhak, M.A. Imran, S. Rajasegarar, Non-intrusive Load Monitoring approaches for disaggregated energy sensing: a survey, Sensors 12 (2012) 16838–16866, https://doi.org/10.3390/s121216838.
[25] K.S. Barsim, R. Streubel, B. Yang, An approach for unsupervised non-intrusive load monitoring of residential appliances, Nilm 2014 (2014), https://doi.org/10.4236/ojee.2013.21002.
[26] Y.F. Wong, Y. Ahmet Sekercioglu, T. Drummond, V.S. Wong, Recent approaches to non-intrusive load monitoring techniques in residential settings, in: IEEE Symposium on Computational Intelligence Applications in Smart Grid, 2013, https://doi.org/10.1109/CIASG.2013.6611501.
[27] K.D. Anderson, M.E. Berges, A. Ocneanu, D. Benitez, J.M.F. Moura, Event detection for Non Intrusive load monitoring, IECON Proceedings (industrial Electronics Conference) (2012), https://doi.org/10.1109/IECON.2012.6389367.
[28] H. Altrabalsi, L. Stankovic, J. Liao, and V. Stankovic, "A low-complexity energy disaggregation method: Performance and robustness," in IEEE Symposium on Computational Intelligence Applications in Smart Grid, CIASG, Orlando, FL, USA, 2014, pp. 1–8. doi: 10.1109/CIASG.2014.7011569.
[29] B. Zhao, L. Stankovic, V. Stankovic, On a Training-Less Solution for Non-Intrusive Appliance Load Monitoring Using Graph Signal Processing, IEEE Access 4 (2016) 1784–1799, https://doi.org/10.1109/ACCESS.2016.2557460.
[30] F. Zhang, L. Qu, W. Dong, H. Zou, Q. Guo, Y. Kong, A Novel NILM Event Detection Algorithm Based on Different Frequency Scales, IEEE Trans. Instrum. Meas. 71 (2022), https://doi.org/10.1109/TIM.2022.3181897.

[31] A.U. Rehman, T.T. Lie, B. Valles, S.R. Tito, Event-Detection Algorithms for Low Sampling Nonintrusive Load Monitoring Systems Based on Low Complexity Statistical Features, IEEE Trans. Instrum. Meas. 69 (3) (Mar. 2020) 751–759, https://doi.org/10.1109/TIM.2019.2904351.

[32] M. Lu, Z. Li, A Hybrid Event Detection Approach for Non-Intrusive Load Monitoring, IEEE Trans. Smart Grid 11 (1) (Jan. 2020) 528–540, https://doi.org/10.1109/TSG.2019.2924862.

[33] K. He, L. Stankovic, J. Liao, V. Stankovic, Non-intrusive load disaggregation using graph signal processing, IEEE Trans. Smart Grid (2018), https://doi.org/10.1109/TSG.2016.2598872.

[34] H. Kim, M. Marwah, M.F. Arlitt, G. Lyon, J. Han, Unsupervised Disaggregation of Low Frequency Power Measurements, in: Proceedings of the SIAM Conference on Data Mining, 2011, https://doi.org/10.1137/1.9781611972832.

[35] Z. Kolter, T. Jaakkola, J.Z. Kolter, Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation, J. Mach. Learn. Res. 22 (2012) 1472–1482.

[36] S. Makonin, F. Popowich, Nonintrusive Load Monitoring (NILM) Performance Evaluation A unified approach for accuracy reporting, Energy Effic. 8 (4) (2015) 809–814, https://doi.org/10.1007/s12053-014-9306-2.

[37] O. Parson, S. Ghosh, M. Weal, A. Rogers, Non-intrusive load monitoring using prior models of general appliance types, in: 26th AAAI Conference on Artificial Intelligence, 2012, https://doi.org/10.1186/s12902-015-0014-5.

[38] L. Mauch, K.S. Barsim, B. Yang, How well can HMM model load signals. Proceeding of the 3rd International Workshop on Non-Intrusive Load Monitoring (NILM 2016), 2016.

[39] S. Bao, L. Zhang, W. Li, D. Sun, B. Zhang, and X. Han, "Feature Selection Method for Non-intrusive Load Monitoring with Balanced Redundancy and Relevancy," in 2020 IEEE/IAS Industrial and Commercial Power System Asia, I and CPS Asia 2020, 2020. doi: 10.1109/ICPSAsia48933.2020.9208609.

[40] W. Luan, Z. Wei, B. Liu, Y. Yu, Non-intrusive power waveform modeling and identification of air conditioning load, Appl. Energy 324 (Oct. 2022) 119755, https://doi.org/10.1016/j.apenergy.2022.119755.

[41] C. Dinesh, P. Perera, R.I. Godaliyadda, M.P.B. Ekanayake, J. Ekanayake, Non-intrusive load monitoring based on low frequency active power measurements, AIMS Energy 4 (3) (2016) 414–443.

[42] A. Harell, S. Makonin, and I. V Bajic, "Wavenilm: A Causal Neural Network for Power Disaggregation from the Complex Power Signal," in: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, Brighton, UK: IEEE, 2019, pp. 8335–8339. doi: 10.1109/ICASSP.2019.8682543.

[43] N. Sadeghianpourhamami, J. Ruyssinck, D. Deschrijver, T. Dhaene, C. Develder, Comprehensive feature selection for appliance classification in NILM, Energy Build. 151 (2017), https://doi.org/10.1016/j.enbuild.2017.06.042.

[44] Q. Wu, F. Wang, Concatenate convolutional neural networks for non-intrusive load monitoring across complex background, Energies (Basel) 12 (8) (2019) 1572.

[45] P. A. Schirmer and I. Mporas, "Double Fourier Integral Analysis Based Convolutional Neural Network Regression for High-Frequency Energy Disaggregation," IEEE Trans Emerg Top Comput Intell, vol. 6, no. 3, 2022, doi: 10.1109/TETCI.2021.3086226.

[46] Y. Himeur, A. Alsalemi, F. Bensaali, and A. Amira, "An intelligent nonintrusive load monitoring scheme based on 2D phase encoding of power signals," International Journal of Intelligent Systems, vol. 36, no. 1, 2021, doi: 10.1002/int.22292.

[47] 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, 2018, doi: 10.1109/TSG.2018.2888581.

[48] H.H. Chang, M.C. Lee, W.J. Lee, C.L. Chien, N. Chen, Feature Extraction-Based Hellinger Distance Algorithm for Nonintrusive Aging Load Identification in Residential Buildings, IEEE Trans. Ind. Appl. 52 (3) (May 2016) 2031–2039, https://doi.org/10.1109/TIA.2016.2533487.

[49] H.H. Chang, Non-intrusive fault identification of power distribution systems in intelligent buildings based on power-spectrum-based wavelet transform, Energy Build. 127 (Sep. 2016) 930–941, https://doi.org/10.1016/j.enbuild.2016.06.050.

[50] M. Nour, J.C. Le Bunetel, P. Ravier, Y. Raingeaud, Data Augmentation Strategies for High-Frequency NILM Datasets, IEEE Trans. Instrum. Meas. 72 (2023), https://doi.org/10.1109/TIM.2023.3261907.

[51] M. Kaselimi, E. Protopapadakis, A. Voulodimos, N. Doulamis, A. Doulamis, Multi-channel recurrent convolutional neural networks for energy disaggregation, IEEE Access 7 (2019), https://doi.org/10.1109/ACCESS.2019.2923742.

[52] H. Altrabalsi, V. Stankovic, J. Liao, L. Stankovic, Low-complexity energy disaggregation using appliance load modelling, AIMS Energy 4 (1) (2016) 1–21, https://doi.org/10.3934/energy.2016.1.1.

[53] S. Barker, Model-Driven Analytics of Energy Meter Data in Smart Homes, 2014.

[54] Y. Li, Z. Peng, J. Huang, Z. Zhang, J.H. Son, Energy Disaggregation via Hierarchical Factorial HMM. Proceeding of the 2014 NILM Workshop, 2014.

[55] R. Bonfigli, S. Squartini, M. Fagiani, and F. Piazza, "Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview," in: 2015 IEEE 15th International Conference on Environment and Electrical Engineering, EEEIC 2015 - Conference Proceedings, 2015, pp. 1175–1180. doi: 10.1109/EEEIC.2015.7165334.

[56] K. S. Barsim and B. Yang, "Toward a semi-supervised non-intrusive load monitoring system for event-based energy disaggregation," in 2015 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2015, 2016. doi: 10.1109/GlobalSIP.2015.7418156.

[57] A. M. Fatouh, O. A. Nasr, and M. M. Eissa, "New Semi-Supervised and Active Learning Combination Technique for Non-Intrusive Load Monitoring," in 2018 IEEE International Conference on Smart Energy Grid Engineering (SEGE), IEEE, Aug. 2018, pp. 181–185. doi: 10.1109/SEGE.2018.8499498.

[58] A. Iwayemi, C. Zhou, SARAA: Semi-Supervised Learning for Automated Residential Appliance Annotation, IEEE Trans. Smart Grid (2015) 1–8, https://doi.org/10.1109/TSG.2015.2498642.

[59] C. Zhang, et al., Achieving Sustained Improvement in Identification Accuracy with a Semi-Supervised Learning Approach for NILM, Systems (2023), https://doi.org/10.36227/techrxiv.22548661.v1.

[60] C.H. Hur, H.E. Lee, Y.J. Kim, S.G. Kang, Semi-Supervised Domain Adaptation for Multi-Label Classification on Nonintrusive Load Monitoring, Sensors 22 (15) (2022) Aug, https://doi.org/10.3390/s22155838.

[61] R.D. Little, Electrical power disaggregation in commercial buildings with applications to a non-intrusive load monitor, Massachusetts Institute of Technology (1991).

[62] G. Elafoudi, L. Stankovic, V. Stankovic, Power disaggregation of domestic smart meter readings using dynamic time warping, in: ISCCSP 2014–2014 6th International Symposium on Communications, Control and Signal Processing, 2014, https://doi.org/10.1109/ISCCSP.2014.6877810.

[63] J. Kim, T.-T.-H. Le, H. Kim, Nonintrusive Load Monitoring Based on Advanced Deep Learning and Novel Signature, Comput. Intell. Neurosci. 2017 (2017) 1–22, https://doi.org/10.1155/2017/4216281.

[64] R. Machlev, Y. Levron, and Y. Beck, "Modified Cross-Entropy Method for Classification of Events in NILM Systems A New Hybrid State Estimation Considering," IEEE Trans Smart Grid, vol. 10, no. 5, 2021, doi: 10.1109/TSG.2018.2871620.

[65] R. Bonfigli, A. Felicetti, E. Principi, M. Fagiani, S. Squartini, F. Piazza, Denoising autoencoders for Non-Intrusive Load Monitoring: Improvements and comparative evaluation, Energy Build. 158 (2018) 1461–1474, https://doi.org/10.1016/j.enbuild.2017.11.054.

[66] W. Kong, Z. Y. Dong, D. J. Hill, J. Ma, J. H. Zhao, and F. J. Luo, "A Hierarchical Hidden Markov Model Framework for Home Appliance Modeling," IEEE Trans Smart Grid, vol. 9, no. 4, 2018, doi: 10.1109/TSG.2016.2626389.

[67] D. Bajovic, K. He, L. Stankovic, D. Vukobratovic, and V. Stankovic, "Optimal detection and error exponents for hidden semi-markov models," IEEE Journal on Selected Topics in Signal Processing, vol. 12, no. 5, 2018, doi: 10.1109/JSTSP.2018.2851506.

[68] L. Mauch and B. Yang, "A new approach for supervised power disaggregation by using a deep recurrent LSTM network," in 2015 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2015, Orlando, FL, USA, 2015, pp. 63–67. doi: 10.1109/GlobalSIP.2015.7418157.

[69] S. Makonin, F. Popowich, I. V. Bajic, B. Gill, and L. Bartram, "Exploiting HMM Sparsity to Perform Online Real-Time Nonintrusive Load Monitoring," IEEE Trans Smart Grid, vol. 7, no. 6, 2016, doi: 10.1109/TSG.2015.2494592.

[70] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V Bajić, "AMPds: A public dataset for load disaggregation and eco-feedback research," in 2013 IEEE Electrical Power and Energy Conference, EPEC *2013*, 2013. doi: 10.1109/EPEC.2013.6802949.

[71] A. Rahimpour, H. Qi, D. Fugate, and T. Kuruganti, "Non-Intrusive Energy Disaggregation Using Non-Negative Matrix Factorization with Sum-to-k Constraint," IEEE Transactions on Power Systems, vol. 32, no. 6, 2017, doi: 10.1109/TPWRS.2017.2660246.

[72] J. Kelly, W. Knottenbelt, Neural NILM: Deep Neural Networks Applied to Energy Disaggregation, in: ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, 2015, pp. 55–64, https://doi.org/10.1145/2821650.2821672.

[73] D. Murray, L. Stankovic, V. Stankovic, S. Lulic, S. Sladojevic, Transferability of Neural Network Approaches for Low-rate Energy Disaggregation, ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings (2019), https://doi.org/10.1109/ICASSP.2019.8682486.

[74] J. Liao, G. Elafoudi, L. Stankovic, V. Stankovic, Non-intrusive appliance load monitoring using low-resolution smart meter data, in: 2014 IEEE International Conference on Smart Grid Communications, SmartGridComm 2014, 2015, https://doi.org/10.1109/SmartGridComm.2014.7007702.

[75] 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, 2020, doi: 10.1109/TSG.2020.2974347.

[76] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, Deep Learning for Computer Vision: A Brief Review, Comput. Intell. Neurosci. (2018. 2018.), https://doi.org/10.1155/2018/7068349.

[77] C. Zhang, M. Zhong, Z. Wang, N. Goddard, C. Sutton, Sequence-to-point learning with neural networks for nonintrusive load monitoring, in: Thirty-Second AAAI Conference on Artificial Intelligence, LA, USA, New Orleans, 2018, p. 8.

[78] K. Chen, Q. Wang, Z. He, K. Chen, J. Hu, J. He, Convolutional sequence to sequence non-intrusive load monitoring, J. Eng. 17 (2018) 2018, https://doi.org/10.1049/joe.2018.8352.

[79] J. Z. Kolter and M. J. Johnson, "REDD: A Public Data Set for Energy Disaggregation Research," in Proceedings of the ACM Workshop on Data Mining Applications in Sustainability (SustKDD), San Diego, California, 2011, pp. 1–6.

[80] M. Zhong, N. Goddard, C. Sutton, Signal aggregate constraints in additive factorial HMMs, with application to energy disaggregation, in: Advances in Neural Information Processing Systems, 2014, pp. 3590–3598.

[81] M. Nguyen, S. Alshareef, A. Gilani, W.G. Morsi, A novel feature extraction and classification algorithm based on power components using single-point

monitoring for NILM, Canadian Conference on Electrical and Computer Engineering (2015), https://doi.org/10.1109/CCECE.2015.7129156.
[82] S. Alshareef and W. G. Morsi, "Application of wavelet-based ensemble tree classifier for non-intrusive load monitoring," in 2015 IEEE Electrical Power and Energy Conference: Smarter Resilient Power Systems, EPEC 2015, 2016. doi: 10.1109/EPEC.2015.7379983.
[83] J.M. Gillis, S.M. Alshareef, W.G. Morsi, Nonintrusive Load Monitoring Using Wavelet Design and Machine Learning, IEEE Trans. Smart Grid (2016), https://doi.org/10.1109/TSG.2015.2428706.
[84] R. Bonfigli, M. Severini, S. Squartini, M. Fagiani, and F. Piazza, "Improving the performance of the AFAMAP algorithm for Non-Intrusive Load Monitoring," in 2016 IEEE Congress on Evolutionary Computation, CEC 2016, 2016. doi: 10.1109/CEC.2016.7743809.
[85] M. Z. A. Bhotto, S. Makonin, I. V. Bajić, Load disaggregation based on aided linear integer programming, IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 64, no. 7, 2017, doi: 10.1109/TCSII.2016.2603479.
[86] S. M. Tabatabaei, S. Dick, W. Xu, Toward Non-Intrusive Load Monitoring via Multi-Label Classification, IEEE Trans Smart Grid, vol. 8, no. 1, 2017, doi: 10.1109/TSG.2016.2584581.
[87] A. K. Jain, S. S. Ahmed, P. Sundaramoorthy, R. Thiruvengadam, and V. Vijayaraghavan, "Current peak based device classification in NILM on a low-cost embedded platform using extra-trees," in 2017 IEEE MIT Undergraduate Research Technology Conference, URTC 2017, 2018. doi: 10.1109/URTC.2017.8284200.
[88] N. Batra, Y. Jia, H. Wang, K. Whitehouse, Transferring decomposed tensors for scalable energy breakdown across regions, in: 32nd AAAI Conference on Artificial Intelligence 2018, 2018, https://doi.org/10.1609/aaai.v32i1.11343.
[89] R. Machlev, J. Belikov, Y. Beck, Y. Levron, MO-NILM: A multi-objective evolutionary algorithm for NILM classification, Energy Build 199 (Sep. 2019) 134–144, https://doi.org/10.1016/J.ENBUILD.2019.06.046.
[90] X. Shi, H. Ming, S. Shakkottai, L. Xie, J. Yao, Nonintrusive load monitoring in residential households with low-resolution data, Appl. Energy 252 (Oct. 2019) 113283, https://doi.org/10.1016/J.APENERGY.2019.05.086.
[91] Q. Yuan, H. Wang, B. Wu, Y. Song, and H. Wang, A fusion load disaggregation method based on clustering algorithm and support vector regression optimization for low sampling data, Future Internet, vol. 11, no. 2, 2019, doi: 10.3390/fi11020051.
[92] T.T.H. Le, H. Kang, H. Kim, Household appliance classification using lower odd-numbered harmonics and the bagging decision tree, IEEE Access 8 (2020), https://doi.org/10.1109/ACCESS.2020.2981969.
[93] C. Puente, R. Palacios, Y. González-Arechavala, and E. F. Sánchez-Úbeda, "Non-intrusive load monitoring (NILM) for energy disaggregation using soft computing techniques," Energies (Basel), vol. 13, no. 12, 2020, doi: 10.3390/en13123117.
[94] K. Bao, K. Ibrahimov, M. Wagner, H. Schmeck, Enhancing neural non-intrusive load monitoring with generative adversarial networks, Energy Informat. 1 (S1) (2018) Oct, https://doi.org/10.1186/s42162-018-0038-y.
[95] M. Kaselimi, A. Voulodimos, E. Protopapadakis, N. Doulamis, A. Doulamis, EnerGAN: A GENERATIVE ADVERSARIAL NETWORK for ENERGY DISAGGREGATION, ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings (2020), https://doi.org/10.1109/ICASSP40776.2020.9054342.
[96] Y. Pan, K. Liu, Z. Shen, X. Cai, Z. Jia, Sequence-To-Subsequence Learning with Conditional Gan for Power Disaggregation, ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings (2020), https://doi.org/10.1109/ICASSP40776.2020.9053947.
[97] K. Chen, Y. Zhang, Q. Wang, J. Hu, H. Fan, and J. He, "Scale- And Context-Aware Convolutional Non-Intrusive Load Monitoring," IEEE Transactions on Power Systems, vol. 35, no. 3, 2020, doi: 10.1109/TPWRS.2019.2953225.
[98] M. Kaselimi, N. Doulamis, A. Voulodimos, A. Doulamis, E. Protopapadakis, EnerGAN++: A Generative Adversarial Gated Recurrent Network for Robust Energy Disaggregation, IEEE Open J. Signal Process. 2 (2021), https://doi.org/10.1109/OJSP.2020.3045829.
[99] A. M. A. Ahmed, Y. Zhang, and F. Eliassen, "Generative Adversarial Networks and Transfer Learning for Non-Intrusive Load Monitoring in Smart Grids," in 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), IEEE, Nov. 2020, pp. 1–7. doi: 10.1109/SmartGridComm47815.2020.9302933.
[100] A. K. Desai, G. Vashishtha, Z. B. Patel, and N. Batra, "Challenges in Gaussian Processes for Non Intrusive Load Monitoring," in NeurIPS Workshop onGaussian Processes, Spatiotemporal Modeling, and Decision-making Systems, Dec. 2022.
[101] M. Qureshi, C. Ghiaus, N. Ahmad, A blind event-based learning algorithm for non-intrusive load disaggregation, Int. J. Electr. Power Energy Syst. 129 (Jul. 2021) 106834, https://doi.org/10.1016/j.ijepes.2021.106834.
[102] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, M. Zhou, MINILM: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. Advances in Neural Information Processing Systems, 2020.
[103] Z. Yue, C.R. Witzig, D. Jorde, H.A. Jacobsen, BERT4NILM: A Bidirectional Transformer Model for Non-Intrusive Load Monitoring, in: NILM 2020 - Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring, 2020, https://doi.org/10.1145/3427771.3429390.
[104] A. Faustine, L. Pereira, H. Bousbiat, S. Kulkarni, UNet-NILM: A Deep Neural Network for Multi-tasks Appliances State Detection and Power Estimation in NILM, in: NILM 2020 - Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring, 2020, https://doi.org/10.1145/3427771.3427859.
[105] M.R. Islam, M.U. Ahmed, S. Barua, S. Begum, A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks, Appl. Sci. 12 (3) (Jan. 2022) 1353, https://doi.org/10.3390/app12031353.
[106] D. Murray, L. Stankovic, V. Stankovic, Explainable NILM Networks, in: NILM 2020 - Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring, 2020, https://doi.org/10.1145/3427771.3427855.
[107] T. Todic, V. Stankovic, L. Stankovic, An active learning framework for the low-frequency Non-Intrusive Load Monitoring problem, Appl. Energy 341 (Jul. 2023) 121078, https://doi.org/10.1016/J.APENERGY.2023.121078.
[108] J. Jiang, Q. Kong, M.D. Plumbley, N. Gilbert, M. Hoogendoorn, D.M. Roijers, Deep Learning-Based Energy Disaggregation and On/Off Detection of Household Appliances, ACM Trans. Knowl. Discov. Data 15 (3) (Jun. 2021) 1–21, https://doi.org/10.1145/3441300.
[109] P. Davies, J. Dennis, J. Hansom, W. Martin, A. Stankevicius, L. Ward, Deep Neural Networks for Appliance Transient Classification, ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings (2019), https://doi.org/10.1109/ICASSP.2019.8682658.
[110] N. V. Linh and P. Arboleya, "Deep learning application to non-intrusive load monitoring," in 2019 IEEE Milan PowerTech, PowerTech 2019, 2019. doi: 10.1109/PTC.2019.8810435.
[111] M. Kaselimi, N. Doulamis, A. Doulamis, A. Voulodimos, E. Protopapadakis, "Bayesian-optimized Bidirectional LSTM Regression Model for Non-intrusive Load Monitoring", in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings (2019), https://doi.org/10.1109/ICASSP.2019.8683110.
[112] M. Ayub and E. S. M. El-Alfy, "Multi-Target Energy Disaggregation using Convolutional Neural Networks," International Journal of Advanced Computer Science and Applications, vol. 11, no. 10, 2020, doi: 10.14569/IJACSA.2020.0111085.
[113] H. Bousbiat, C. Klemenjak, and W. Elmenreich, "Exploring Time Series Imaging for Load Disaggregation," in BuildSys 2020 - Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, 2020. doi: 10.1145/3408308.3427975.
[114] J. Barber, H. Cuayáhuitl, M. Zhong, W. Luan, Lightweight Non-Intrusive Load Monitoring Employing Pruned Sequence-to-Point Learning, in: NILM 2020 - Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring, 2020, https://doi.org/10.1145/3427771.3427845.
[115] 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 Transactions on Consumer Electronics, vol. 66, no. 2, 2020, doi: 10.1109/TCE.2020.2977964.
[116] F. Ciancetta, G. Bucci, E. Fiorucci, S. Mari, A. Fioravanti, A New Convolutional Neural Network-Based System for NILM Applications, IEEE Trans Instrum Meas 70 (2021), https://doi.org/10.1109/TIM.2020.3035193.
[117] H. Rafiq, X. Shi, H. Zhang, H. Li, M.K. Ochani, A Deep Recurrent Neural Network for Non-Intrusive Load Monitoring Based on Multi-Feature Input Space and Post-Processing, Energies (Basel) 13 (9) (2020) 2195, https://doi.org/10.3390/en13092195.
[118] L. Massidda, M. Marrocu, and S. Manca, "Non-intrusive load disaggregation via a fully convolutional neural network: Improving the accuracy on unseen household," in Proceedings - 2020 2nd IEEE International Conference on Industrial Electronics for Sustainable Energy Systems, IESES 2020, 2020. doi: 10.1109/IESES45645.2020.9210661.
[119] Y. Zhang, B. Yin, Y. Cong, Z. Du, Multi-state Household Appliance Identification Based on Convolutional Neural Networks and Clustering, Energies (Basel) (2020), https://doi.org/10.3390/en13040792.
[120] R. Kukunuri *et al.*, "EdgeNILM: Towards NILM on Edge devices," in BuildSys 2020 - Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, 2020. doi: 10.1145/3408308.3427977.
[121] A. Reinhardt, M. Bouchur, On the Impact of the Sequence Length on Sequence-to-Sequence and Sequence-to-Point Learning for NILM, in: NILM 2020 - Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring, 2020, https://doi.org/10.1145/3427771.3427857.
[122] G. Zhou, Z. Li, M. Fu, Y. Feng, X. Wang, C. Huang, Sequence-to-Sequence Load Disaggregation Using Multiscale Residual Neural Network, IEEE Trans Instrum Meas 70 (2021), https://doi.org/10.1109/TIM.2020.3034989.
[123] Z. Jia, L. Yang, Z. Zhang, H. Liu, F. Kong, Sequence to point learning based on bidirectional dilated residual network for non-intrusive load monitoring, Int. J. Electr. Power Energy Syst. 129 (2021), https://doi.org/10.1016/j.ijepes.2021.106837.
[124] J. Song, H. Wang, M. Du, L. Peng, S. Zhang, and G. Xu, "Non-intrusive load identification method based on improved long short term memory network," Energies (Basel), vol. 14, no. 3, 2021, doi: 10.3390/en14030684.
[125] A. Moradzadeh, K. Pourhossein, B. Mohammadi-Ivatloo, T. Khalili, and A. Bidram, "Radial deformation emplacement in power transformers using long short-term memory networks," in 2021 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2021, 2021. doi: 10.1109/ISGT49243.2021.9372252.
[126] C. Athanasiadis, D. Doukas, T. Papadopoulos, and A. Chrysopoulos, "A scalable real-time non-intrusive load monitoring system for the estimation of household appliance power consumption," Energies (Basel), vol. 14, no. 3, 2021, doi: 10.3390/en14030767.
[127] F. Kalinke, P. Bielski, S. Singh, E. Fouché, K. Böhm, An Evaluation of NILM Approaches on Industrial Energy-Consumption Data, in: e-Energy 2021 - Proceedings of the 2021 12th ACM International Conference on Future Energy Systems, 2021, https://doi.org/10.1145/3447555.3464863.
[128] 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, 2021, doi: 10.1109/TSG.2021.3082622.
[129] D. Jia, Y. Li, Z. Du, J. Xu, B. Yin, Non-intrusive load identification using reconstructed voltage-current images, IEEE Access 9 (2021), https://doi.org/10.1109/ACCESS.2021.3082432.
[130] L. Werthen-Brabants, T. Dhaene, D. Deschrijver, Uncertainty quantification for appliance recognition in non-intrusive load monitoring using Bayesian deep learning, Energy Build. 270 (Sep. 2022) 112282, https://doi.org/10.1016/J.ENBUILD.2022.112282.
[131] Y. Liu, C. Liu, W. Wang, S. Gao, X. Huang, A robust non-intrusive load disaggregation method with roof-top photovoltaics, Electr. Pow. Syst. Res. 208 (2022), https://doi.org/10.1016/j.epsr.2022.107887.
[132] D. Li, et al., Transfer learning for multi-objective non-intrusive load monitoring in smart building, Appl. Energy 329 (2023), https://doi.org/10.1016/j.apenergy.2022.120223.
[133] L. Qu, Y. Kong, M. Li, W. Dong, F. Zhang, H. Zou, A residual convolutional neural network with multi-block for appliance recognition in non-intrusive load identification, Energy Build. 281 (2023), https://doi.org/10.1016/j.enbuild.2022.112749.
[134] N. Lin, B. Zhou, G. Yang, and S. Ma, "Multi-head Attention Networks for Nonintrusive Load Monitoring," in 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), IEEE, Aug. 2020, pp. 1–5. doi: 10.1109/ICSPCC50002.2020.9259533.
[135] D. Garcia-Perez, D. Perez-Lopez, I. Diaz-Blanco, A. Gonzalez-Muniz, M. Dominguez-Gonzalez, and A. A. Cuadrado Vega, "Fully-Convolutional Denoising Auto-Encoders for NILM in Large Non-Residential Buildings," IEEE Trans Smart Grid, vol. 12, no. 3, 2021, doi: 10.1109/TSG.2020.3047712.
[136] N. Virtsionis Gkalinikis, C. Nalmpantis, D. Vrakas, Attention in Recurrent Neural Networks for Energy Disaggregation, in: Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2020, https://doi.org/10.1007/978-3-030-61527-7_36.
[137] 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, 2021, doi: 10.1109/TSG.2020.3027491.
[138] V. Piccialli and A. M. Sudoso, "Improving non-intrusive load disaggregation through an attention-based deep neural network," Energies (Basel), vol. 14, no. 4, 2021, doi: 10.3390/en14040847.
[139] L. Huang, S. Chen, Z. Ling, Y. Cui, Q. Wang, Non-invasive load identification based on LSTM-BP neural network, Energy Rep. 7 (2021), https://doi.org/10.1016/j.egyr.2021.01.040.
[140] H. Grover, L. Panwar, A. Verma, B.K. Panigrahi, T.S. Bhatti, A multi-head Convolutional Neural Network based non-intrusive load monitoring algorithm under dynamic grid voltage conditions, Sustain. Energy Grids Netw. 32 (2022), https://doi.org/10.1016/j.segan.2022.100938.
[141] Q. Liu, et al., Improving wireless indoor non-intrusive load disaggregation using attention-based deep learning networks, Phys. Commun. 51 (2022), https://doi.org/10.1016/j.phycom.2021.101584.
[142] A. Moradzadeh, S. Zakeri, W.A. Oraibi, B. Mohammadi-Ivatloo, Z. Abdul-Malek, R. Ghorbani, Non-intrusive load monitoring of residential loads via laplacian eigenmaps and hybrid deep learning procedures, Sustainability 14 (22) (Nov. 2022) 14898, https://doi.org/10.3390/su142214898.
[143] T. Chen, H. Qin, X. Li, W. Wan, W. Yan, A non-intrusive load monitoring method based on feature fusion and SE-ResNet, Electronics (Basel) 12 (8) (Apr. 2023) 1909, https://doi.org/10.3390/electronics12081909.
[144] R. Sun, K. Dong, J. Zhao, DiffNILM: a novel framework for non-intrusive load monitoring based on the conditional diffusion model, Sensors 23 (7) (Mar. 2023) 3540, https://doi.org/10.3390/s23073540.
[145] D. Pujić, N. Tomašević, M. Batić, A semi-supervised approach for improving generalization in non-intrusive load monitoring, Sensors 23 (3) (Jan. 2023) 1444, https://doi.org/10.3390/s23031444.
[146] C. Klemenjak, S. Makonin, and W. Elmenreich, "Towards Comparability in Non-Intrusive Load Monitoring: On Data and Performance Evaluation," in 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), IEEE, Feb. 2020, pp. 1–5. doi: 10.1109/ISGT45199.2020.9087706.
[147] T. Kriechbaumer, H.A. Jacobsen, BLOND, a building-level office environment dataset of typical electrical appliances, Sci. Data 5 (2018), https://doi.org/10.1038/sdata.2018.48.
[148] F. Jazizadeh, M. Afzalan, B. Becerik-Gerber, L. Soibelman, EMBED: A dataset for energy monitoring through building electricity disaggregation, in: e-Energy 2018 - Proceedings of the 9th ACM International Conference on Future Energy Systems, 2018, https://doi.org/10.1145/3208903.3208939.
[149] D. Murray, L. Stankovic, V. Stankovic, An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study, Sci. Data 4 (2017), https://doi.org/10.1038/sdata.2016.122.
[150] M. Wenninger, A. Maier, and J. Schmidt, DEDDIAG, a domestic electricity demand dataset of individual appliances in Germany, Sci. Data, vol. 8, no. 1, 2021, doi: 10.1038/s41597-021-00963-2.
[151] M. Pullinger *et al.*, "The IDEAL household energy dataset, electricity, gas, contextual sensor data and survey data for 255 UK homes," Sci Data, vol. 8, no. 1, 2021, doi: 10.1038/s41597-021-00921-y.
[152] J.-P. Zimmermann, et al., Household Electricity Survey: A study of domestic electrical product usage, Intertek (2012).
[153] J. Gao, S. Giri, E.C. Kara, M. Bergés, PLAID, A public dataset of high-resolution electrical appliance measurements for load identification research, in: BuildSys 2014 - Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, 2014, https://doi.org/10.1145/2674061.2675032.
[154] A. Harell, R. Jones, S. Makonin, and I. V. Bajic, "TraceGAN: Synthesizing Appliance Power Signatures Using Generative Adversarial Networks," IEEE Trans Smart Grid, vol. 12, no. 5, 2021, doi: 10.1109/TSG.2021.3078695.
[155] D.P.B. Renaux, et al., A Dataset for Non-Intrusive Load Monitoring: Design and Implementation, Energies (Basel) 13 (20) (Oct. 2020) 5371, https://doi.org/10.3390/en13205371.
[156] 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, 2020, doi: 10.1038/s41597-020-0434-6.
[157] L. Pereira, "NILMPEds: A performance evaluation dataset for event detection algorithms in non-intrusive load monitoring," Data (Basel), vol. 4, no. 3, 2019, doi: 10.3390/data4030127.
[158] L. Pereira, N. Nunes, A comparison of performance metrics for event classification in Non-Intrusive Load Monitoring, in: 2017 IEEE International Conference on Smart Grid Communications, SmartGridComm 2017, 2018, https://doi.org/10.1109/SmartGridComm.2017.8340682.
[159] P. Laviron, X. Dai, B. Huquet, T. Palpanas, in: Electricity Demand Activation Extraction: from Known to Unknown Signatures, Using Similarity Search, ACM, Jun, New York, NY, USA, 2021, pp. 148–159, https://doi.org/10.1145/3447555.3464865.
[160] M. Zeifman, K. Roth, Nonintrusive appliance load monitoring: Review and outlook, IEEE Trans. Consum. Electron. 57 (1) (2011) 76–84, https://doi.org/10.1109/TCE.2011.5735484.
[161] R. Gopinath, M. Kumar, K.J. Lokesh, K. Srinivas, Performance Analysis of Similar Appliances Identification using NILM Technique under Different Data Sampling Rates, in: NILM 2020 - Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring, 2020, https://doi.org/10.1145/3427771.3427858.
[162] M. D'Incecco, S. Squartini, M. Zhong, Transfer Learning for Non-Intrusive Load Monitoring, IEEE Trans. Smart Grid 11 (2) (2020) 1419–1429, https://doi.org/10.1109/tsg.2019.2938068.