Advances in Machine-Learning Based Disaggregation of Building Heating Loads: A Review

Synne Krekling Lien[1], Behzad Najafi[2] and Jayaprakash Rajasekharan[1]

[1]Department of Electric Energy, Norwegian University of Science and Technology, Norway
[2]Department of Energy, Politecnico di Milano, Italy
Synne.k.lien@ntnu.no

Abstract. This review article investigates the methods proposed for disaggregating the space heating units' load from the aggregate electricity load of commercial and residential buildings. It explores conventional approaches together with those that employ traditional machine learning, deep supervised learning and reinforcement learning. The review also outlines corresponding data requirements and examines the suitability of a commonly utilised toolkit for disaggregating heating loads from low-frequency aggregate power measurements. It is shown that most of the proposed approaches have been applied to high-resolution measurements and that few studies have been dedicated to low-resolution aggregate loads (e.g. provided by smart meters). Furthermore, only a few methods have taken account of special considerations for heating technologies, given the corresponding governing physical phenomena. Accordingly, the recommendations for future works include adding a rigorous pre-processing step, in which features inspired by the building physics (e.g. lagged values for the ambient conditions and values that represent the correlation between heating consumption and outdoor temperature) are added to the available input feature pool. Such a pipeline may benefit from deep supervised learning or reinforcement learning methods, as these methods are shown to offer higher performance compared to traditional machine learning algorithms for load disaggregation.

Keywords: Load disaggregation, non-intrusive load monitoring, Smart Meter Analytics, Machine learning, Space Heating, Building Energy Use

1 Introduction

In 2021, the operation of buildings was responsible for 30% of final global energy consumption and 27% of total energy sector emissions (out of which 8% is related to direct emissions from buildings, while 19% refers to emissions from generation of heat and electricity consumed by buildings) [1]. The electrification of buildings has been identified as a key alternative to achieve a more sustainable energy system and mitigate the corresponding emission of gases that result in climate change [2]. The Norwegian building sector is a special case as its heating supply is highly electrified, due to historically low electricity prices; the building stock was accordingly responsible for approximately 37% of delivered energy and 55% of delivered electricity [3] in 2021. This has a significant effect on the peak electricity use and, consequently, the peak experienced by the grid coincides with the coldest hours of the year. Therefore, as the transport and industrial sector are becoming electrified, electricity peaks are expected to rise, increasing the strain on the electricity grid. A substantial share of grid investments will be made to avoid bottlenecks that are expected to occur for only a few hours each year [4]. To limit the growth in the peak load that is expected in the green transition with increased electrification, and the resulting cost to society and consumers, further knowledge is needed about buildings' electricity use behind the main meter. Electricity meters only show how much electricity is delivered to the customer, but not how the electricity is used or how different loads and appliances drive the peaks. By separating the electricity consumption specifically for heating purposes from the overall electricity consumption in buildings, one can attain a deeper understanding of both the proportion of total energy consumption and the increase in peak load attributable to heating. This approach not only enhances our comprehension of the influence of various heating appliances and types of buildings on the peak load, but also provides valuable data for optimising grid planning and facilitating more efficient demand management within buildings. Consequently, this practice can result in cost savings for building occupants and reduce the need for investments in the grid, thereby benefiting society at large. As a consequence, this paper aims to investigate previous research on disaggregation of heating loads from the total electricity load of buildings. The goal is to gain insight into the most promising methodologies and to identify any existing research gaps in the domain of heating load disaggregation.

2 Background

By 2019, Norway mandated the installation of smart electricity meters for all electricity consumers as part of advanced metering systems (AMS) [5]. These meters record customers' hourly electricity usage and transmit data to grid companies. Moreover, the meters can provide high frequency (seconds) and medium frequency (minutes) data on electricity usage, as well as information on active and reactive power, voltage and frequency through the Home Area Network gate (HAN-gate). This allows for collection of aggregate electricity consumption data. However, to gain a deeper understanding of the total peak load and how to limit it, it is necessary to disaggregate the load to specific appliances, particularly those corresponding to heating systems [6].

One way to perform load disaggregation is by using intrusive metering, which involves installing separate meters for each appliance or in each building. However, this approach can be expensive due to costs associated with the manufacturing, installation, maintenance and monitoring of the required measurement devices. It can also be inconvenient for building residents, and new meters would have to be installed for every new appliance or installation, making it impractical and challenging to scale [7]. An alternative load disaggregation method is non-intrusive load monitoring (NILM), which involves using software tools to analyse power signals and disaggregate total energy load into individual loads or appliances from a single point of measurement. The concept of NILM was first proposed by Hart in the 1980s [8]. The objective of NILM techniques is to determine the individual power consumption or on/off state of electrical loads. These methods rely solely on measuring the aggregate energy consumption of these loads. NILM has various applications, including monitoring energy use in residential and service buildings, as well as in the industrial sector. Typically, NILM techniques are classified into two categories: ‘low-frequency approaches’ that use data or features at a frequency lower than the AC base frequency in buildings, or ‘high-frequency approaches’ with higher frequencies [9], [10]. The AC base frequency is usually 50 Hz or 60 Hz for AC power systems in Europe, Asia and North America.

Traditionally, NILM approaches for building loads involve 4 steps: data acquisition, appliance and feature extraction, inference and learning, and finally, load disaggregation and classification Several techniques have been studied for NILM [12], linear regression models and unsupervised methods. Optimisation and regression techniques are computationally efficient and can yield good results with small datasets. However, in recent years there has been significant research into other machine learning (ML) methods, particularly supervised learning techniques, which have gained substantial attention. These methods include Bayesian classifiers[13], support vector machines[14] and K-nearest neighbours[15], among others. Approaches used in unsupervised training instead include blind source separation[16], and the most researched method, which is hidden Markov models(HMM) [17], [18]. In addition, Deep Neural Networks (DNNs) have seen tremendous success in the domains of vision and natural language processing in recent years. Accordingly, since 2015 there has been a rapid increase in the number of DNN-based approaches and applications for building load disaggregation [9] [19].

With the increasing number of disaggregation techniques, applications and research, the NILM toolkit (NILMTK) was developed in an effort to create reproducible NILM experiments that serves as a reference library for dataset parsers and benchmark algorithm implementations [20]. The original NILMTK library comes with implemented methods for combinatorial optimisation (CO), mean regression, factorial hidden Markov models (FHMM), and the original algorithm by Hart from 1985. The toolkit can be used to disaggregate any dataset which has been structured as an NILMTK dataset, either manually or through a simple API, and the results can be reviewed using the performance measures implemented. Furthermore, the NILMTK-Contrib repository is an extension of the NILMTK toolkit that offers additional disaggregation algorithms, such as recurrent neural networks (RNN), FHMM, sequence-to-sequence models (Seq2Seq), and more [21]. Another toolkit that is an extension of NILMTK is Torch-NILM, which offers a suite of tools for training deep neural networks in the task of energy disaggregation [22].

In the Norwegian case, it is assumed that heating loads/heating technologies are responsible for the majority of the annual electricity consumption, as well as the electric peak load in buildings [23]. A deeper knowledge of how this varies for different building categories, and how the heating loads can be disaggregated from hourly measurements from the AMS meters, is missing. Disaggregation approaches are usually applied, trained and validated on datasets with labelled data for energy use in buildings. In some cases, these datasets have separate measurements for space heating units of buildings, and NILM techniques are used to identify and disaggregate heating loads from the total electricity load in buildings. The experience from these studies can help to gain more knowledge about disaggregation of heating loads in buildings, the nature of electricity used for heating in all electric buildings in cold climates, and insights on how to limit the growth in the peak electricity demand in Norway.

2.1 Scope

The scope of this review is to look into methods used specifically for disaggregation of space heating technologies and space heating demand from the total electric load of both commercial and residential buildings. This is particularly relevant for the Norwegian case, as the peak demand load is mainly caused by electrical heating in buildings. When we talk about heating loads in this paper, we specifically mean electricity used for space heating and heating of ventilation air. When talking about heating appliances, we consider all electrical appliances that can be used for space/ventilation heating, including heat pumps (air-to-air, ground-source, air-to-water, etc.), electric space heaters, electric heating batteries, electric floor heating and electrical boilers.

This paper explores methods that utilise both traditional learning methods, as well as deep supervised learning and reinforcement learning (RL) approaches, and outlines the data requirements for load disaggregation of the space heating demand in all electric, and partially electric, buildings, with recommendations for further work. The paper also looks into and provides an overview of building energy measurement datasets used for developing disaggregation methods. Finally, the paper briefly examines the suitability of NILMTK for disaggregation of heating technologies from low frequency electricity use in all electric buildings.

In this paper, the resolution of different datasets and approaches are referred to either as frequency given in Hz or as per time unit, given as seconds/minutes/hours. These units are used interchangeably, but essentially 1 Hz is the same as 1/s, meaning that a dataset with the resolution of one Hz has measurements with 1-second resolution. In this review paper, we also consider datasets and methods which are applied to datasets with resolution in the seconds domain to be of high resolution, while datasets with measurements in the minute or hour domain are considered to be of low resolution.

2.2 Related works & contributions

Several review studies have examined NILM and disaggregation techniques that employ machine learning to disaggregate individual appliances. A selection of these are summarised in Table 1. Some studies have partly examined methods for disaggregating heating appliances’ electricity use from the building’s aggregate load. However, there is a lack of a systematic overview of techniques that specifically address disaggregation of heating loads in buildings. Such a review is essential to determine the research gaps in the disaggregation field regarding heating loads and heating appliances and can contribute to increasing the knowledge of how heating loads contribute to peak loads in all-electric buildings and the building stock.

Other recently conducted reviews and relevant articles focus on disaggregation of building energy loads and/or machine learning approaches in the disaggregation research field. Rafati et al. [24] performed a review of NILM used for fault detection and efficiency assessment of HVAC systems. This review considered different methods of NILM applied to building HVAC systems with different measurement durations and sampling frequencies and showed that, even though NILM could be successfully implemented for Fault Detection and Diagnosis (FDD) and the energy efficiency (EE) evaluation of HVAC, and enhance the performance of these techniques, there are many research opportunities to improve or develop NILM-based FDD methods to deal with real-world challenges. Huber et al. [9] reviewed NILM approaches that employ deep neural networks to disaggregate appliances from low-frequency data, i.e. data with sampling rates lower than the AC base frequency. The study looked at around 100 studies in which deep learning approaches were used for NILM. Energy use for heat pumps was disaggregated in ten of the studies examined that investigated deep learning methods on the AMPds-datasets [25], while two studies disaggregated electric heaters. The study also found that the number of deep neural network approaches to solve NILM problems has increased rapidly since 2015. Himeur et al. [26] looked into machine learning methods for anomaly detection of energy consumption in buildings using machine learning. The method briefly reviewed ML-based NILM for anomaly detection of energy consumption in buildings and concluded that even though the performance of NILM to identify abnormal consumption is not yet as accurate as using sub-metering feedback, its performance could be further improved, to allow a robust identification of faulty behaviour. Himeur et al. (2) [27] made a second review of recent trends in smart NILM frameworks (event-based, non-event-based), as well as a more technical review describing sensors and devices utilised to collect energy consumption data in residential and public buildings before applying NILM. They also reviewed real-life applications of NILM systems. Angelis et al. [28] undertook a more general literature review about commonly used methodology and applications of NILM for building energy consumption. Earlier, Ruano et al. [29] reviewed NILM methods specifically for Home Energy Management Systems (HEMS) and Ambient Assisted Living (AAL).

Table 1 Overview of other review articles on disaggregation of building energy use.

Reference	Scope	Building category and appliances	References	Disaggregation approaches
Rafati et al. 2022 [24]	Fault detection, energy efficiency assessment	HVAC systems in residential and commercial buildings	53	All kinds
Huber et al. 2021[9]	DNN for NILM applications	Mostly residential. Electrical appliances, some including HVAC and HP	190	DNN
Himeur et al. 2021(1) [26]	ML approaches for anomaly detection of energy consumption, including ML for disaggregation	Energy consumption for all appliances	7 specifically about disaggregation out of 264	ML, supervised, unsupervised, DNN
Himeur et al. 2021(2) [27]	Overview of event and non-event based NILM-methods. Applications of NILM-systems	Residential and public buildings, all appliances	200	Both ML and other
Angelis et al.2022 [28]	General overview of NILM methods and applications for building energy use	Residential and public buildings	237	Both ML and other
Ruano et al. 2019 [29]	Review of NILM methods, focusing particularly on recent proposals and their applications, particularly in the areas of Home Energy Management Systems (HEMS) and Ambient Assisted Living (AAL)	Residential, all appliances	152	NILM specifically for HEMS and AAL. Some discussion of ML methods. No special attention to heating.

To the best of the authors' knowledge, this review article is the first review to specifically look into the scientific literature on disaggregation of space heating appliances/space heating loads from the aggregate building load. In contrast to other review papers that concentrate on the disaggregation of all household appliances, including electrical heating devices among them, there is reason to believe that heating appliances could potentially benefit from more tailored methodologies. This assertion is grounded in the relation with outdoor temperature, building characteristics and heating appliance usage patterns. The primary contribution of this paper is to investigate this proposition and to assess the effectiveness of machine learning advancements in the specific context of disaggregating heating loads.

2.3 Outline of the paper

The paper is structured into different sections. Section 3, entitled "Methodology", explains the approach taken in the literature search. Additionally, it offers an overview of commonly used datasets within the NILM/disaggregation field, together with insights into the availability of separate meters for heating appliances in these datasets.

Proceeding to Section 4, "Disaggregation of Buildings' Heating Loads", the main findings from the literature review are presented. This section includes details of various disaggregation studies, categorised as traditional methods, deep supervised learning approaches and reinforcement learning methods.

Section 5 provides an evaluation of the data requirements essential for the development of effective disaggregation approaches.

Finally, in Section 6, the paper concludes by summarising the key findings and insights.

3 Methodology

The aim of this article is to conduct a literature review of proposed methods of disaggregation of building heating loads and the advances of machine learning methods within this field. This literature review looks into three main categories of literature: 1) other relevant review studies on disaggregation of building loads; 2) documented datasets for energy use in buildings used in NILM/disaggregation research; and 3) methods and results of disaggregation approaches for building heating loads.

To conduct this review, a literature search was executed on Google scholar, Elsevier library and IEEE Xplore using various combinations of key words: "disaggregation", "NILM", "HVAC", "space heating", "machine learning", "buildings", "deep learning", in January/February 2023. This resulted in a total of 1,970 articles being extracted, of which around 200 articles were screened and marked as relevant for the topic of disaggregation of energy use in buildings. An additional step in the literature search was conducted using Connectedpapers.org for articles which appeared to be specifically relevant, e.g. datasets containing measurements of heating technologies or disaggregation-methods utilising these datasets. This search describes citations within each article and points to other articles where the article in question is cited.

During the work with this article, the authors also tested NILMTK [20] and the extension NILMTK-Contrib [21] to get a better overview of the datasets used for NILM approaches, and some of the methods used for NILM. To get a better understanding of the requirements and limitations of NILM-Toolkit and commonly used datasets, NILMTK and NILMTK-Contrib were tested by following the user guide accessible through GitHub, using the IAWE [30], UK-DALE [31] and AMPds [25] datasets. The toolkit was also tested to give better insights into the NILM methodology and results. The toolkit was tested by writing a dataset converter and using the authors' proprietary dataset with hourly measurements of energy use in two Norwegian school buildings. Metadata for the datasets which are compatible with NILMTK was examined within the toolkit itself by accessing the information in the NILMTK GitHub repository[32] giving supplementary information about the datasets.

3.1 Datasets

To train, test and benchmark various disaggregation techniques, a variety of datasets are used in the literature. While some researchers gather and utilise proprietary datasets for their novel disaggregation approaches, acquiring high-resolution data on multiple buildings and appliances necessitates a significant investment of time and resources. As a result, most disaggregation methods are built, tested and benchmarked using existing datasets. Understanding these widely utilised datasets and their content is critical for gaining insight into which approaches have been utilised for disaggregating heating loads and technologies in buildings. Some of the datasets most frequently referenced in this paper include the datasets AMPDs [25], UK-Dale [31], IAWE [30] and REFIT [33] among others. AMPDs (Almanac of Minutely Power dataset) is a public dataset for load disaggregation and eco-feedback research and is a record of energy consumption of a single house in Vancouver with 21 sub-meters for an entire year (from April 1, 2012 to March 31, 2013) at one minute read intervals [31]. UK-Dale (UK recording Domestic Appliance-Level Electricity) is an open access dataset of disaggregated energy use data from 5 houses in the UK, measured over 655 days [31]. REFIT is another UK dataset he with electrical load measurements of 20 houses with nine individual appliance measurements at 8-second intervals per house, collected continuously over a period of two years [33]. The IAWE (Indian Dataset for Ambient Water and Energy) contains measurements of 73 days in 2013 of energy and water use data for a single family house [30]. Table 2 gives an overview of these datasets and other datasets for building energy use which are used in research on load disaggregation. The table indicates the location, building category and number of buildings in the datasets, as well as the measurement duration, sampling rate and available measured quantities. The availability of separate measurements for heating loads and heating technologies within the dataset, as well as NILMTK-compatibility, are also indicated in the table. A list of abbreviations is given at the end of the table.

Table 2 Overview of different datasets containing energy measurements for building loads.

Dataset	*Location*	*Buildings*	*Duration*	*Sampling rate*	*Meters*	*Appliances*	*NILMTK compatibility*
UK-DALE [31]	UK	5 SFH	2.5 years	16 kHz, 6s	P	H1: TOT, BOIL, STP, 50 EA H2: TOT, 19 EA H3: TOT, EH, 3 EA H4: TOT, GBOIL, 4 EA H5: TOT, 24 EA	Yes
REFIT [33]	UK	20 SFH	2 years	8 sec	P	H1, H9, H15: EH 19 EA (up to 9 per house)	Yes
REDD [34]	USA	3 SFH, 3 APT	3-19 days	0,33 Hz-15kHz	V, P	H2, H3: EH 10-20 EA per house	Yes
ECO [35]	Switzerland	6 SFH	8 months	1 Hz	V, I, P, Q, Φ	5-8 EA per house	Yes
SynD[36]	Austria	2 SFH	180	5 Hz	P	EH and 20 EA	Yes
ENERTALK [37]	Korea	22 SFH	29-122 days	15 Hz	P, Q	1-7 EA+TOT per building	Yes
AMPds(2)[25]	Canada	1 SFH	1 year (AMPds) 2 years (AMPds2)	1 min	V, I, f, pf, P, Q, S, E	Several EA, including HP. Electricity, water, natural gas.	Yes
HES[38]	United Kingdom	26 (year) 224 (month) RES	1 year/1 month	1/600 Hz	I, V, P, T	Only total consumption + recording of which appliances that were on/off	Yes
PLAID[39]	USA	Lab and 64 RES	5 seconds	30 kHz	V, I	TOT. 1876 appliances, mostly AE, 17 types. 15 EH (not in aggregate measurements). 27 AC.	Unknown
BLUED[40]	USA	1 SFH	8 days	12 kHz	V, I, P	TOT, 43 EA	Unknown
Tracebase [41]	Germany	-	1 day	1 Hz	P	122 EA	Unknown
BERDS[42]	USA	1 COM	1 year	20s	P, Q, S	Unknown #EA, HVAC	Unknown
iAWE[30]	India	1 SFH	73 days	1 Hz, 6s	V, I, f, P, Q, S, E, Φ	33 sensors, inc. EWH and AC	Yes
GREEND[43]	Austria/Italy	9 SFH	1 year	1 Hz	P	9 EA per house, TOT	Yes
DRED[44]	Netherlands	1 SFH	6 months	1 Hz, 1 s	P, T	11 EA, 1 CH	Yes
Dataport[45]	USA	722 RES	3.5 years	1 Hz-1min		<10 meters per house, mostly EA. EH and HWH present in very few houses.	Yes
RAE [46]	Canada	2 SFH	72 days	1 Hz	V, P, Q, S, f, E	20 EA, HP, EB	Yes
I-BLEND[47]	India	7 UNI	1 year	1 Hz, 1 min	V, P, f, I, Pf	No single appliances. Only per building and aggregate (7 buildings). Includes occupancy data.	Unknown
IRISE[48]	France	100 households	1 year	10 min, 1 hour	E	Unknown EA, 3 houses with HWH, EH	Unknown
Smart*[49]	USA	3 SFH	3 months	1 Hz	V, f, P, S	H1: 19 EA/rooms, 1 gBoil H2: 28 EA/rooms, 1 gBoil H3: 19EA/rooms, 2 Boilers (unknown fuel)	Yes
COMBED[50]	India	6 UNI (same as I-BLEND) 1 MFH (student housing)	1 Month	30s	P, Pf, f, E	No heating. Unknown # EA.	Yes
ElDeK [51]	Norway	75 RES+500 apps.	1 year	1 hour (households), 1 min (appliances)	P, E	500 appliances, inc. heaters	No
treASURE[52]	Norway	316 in 10 building categories (2 res and 8 com)	1-4 years	1 hour	P, E	Electricity, District heating (disaggregated to SH and DHW)	No
Denmark SMDH [53]	Denmark	2460 SFH, 564 MFH, 8 Service, 97 unclear	3 years	1 hour	P, E, T	District heating smart meter data	No

Abbreviations

Services/appliances		Building category		Measurements	
TOT	Aggregated electric load	RES	Residential buildings, SFH and/or MFH where this is not specified.	P	Active power
EA	electrical appliances (all electric appliances such as TV and kitchen appliances not used for SH, DHW, HVAC)	SFH	single family house (detached. semi-detached,	Q	Reactive power
BOIL	solar thermal pump	MFH	Multifamily house (apartments)	S	Apparent power
STP	Solar pump	COM	Commercial	E	Energy
EH	Electric heater	UNI	University building	f	Frequency
gBOIL	Gas boiler			Pf	Power Factor
AC	Air conditioning			Φ	Phase angle
EWH	Electric water heater			V	Voltage
CH	Central heating			I	Current
				T	Temperature

4 Disaggregation of buildings’ heating loads

This section presents the literature review on methods for disaggregation of heating loads from aggregate building electricity loads. All investigated methods are summarised in Table 3 at the end of the section. The section is divided into three sub-sections based on the main machine learning class used in the different articles – namely “Traditional methods and shallow-algorithms”, “Deep supervised learning” and “Reinforcement learning”.

4.1 Traditional methods and shallow algorithms

The first disaggregation methods were based on rule-based algorithms, statistical methods, and shallow learning algorithms, such as combinatorial optimisation, clustering and regression models. In this article, shallow learning refers to non-deep machine learning methods or traditional machine learning methods. These methods are still prevalent today due to their simplicity, interpretability and computational efficiency, making them well-suited for real-time monitoring and control applications. For instance, optimisation and linear regression models can be trained on small datasets, making them useful when data is limited or expensive to collect.

Liu et al.’s method, which is based on Affinity propagation clustering (AP) and time-segmented state probability (TSSP), was proposed as a fast working algorithm for real time disaggregation in 2021 [54]. This method was tested on the AMPds dataset and offered an average load state identification accuracy of over 96% and power decomposition accuracy of over 89%, while including all appliances. The state recognition accuracy for heat pumps was notably lower than corresponding average accuracy obtained for all appliances, although it maintained a high power composition accuracy.

Another method that is based on optimisation was proposed by Balletti et al. [55], with a novel penalty-based binary quadratic programming formulation with appliance-specific as well as an optimisation-based automatic state detection algorithm to estimate power levels of appliances and their respective transient behaviour. Their approach was trained and tested on AMPds, UKDALE and REFIT and its capability to disaggregate many appliances with high accuracy was demonstrated. However, in the training procedure, some issues were faced for heat pumps, as the same unit is employed for cooling in summer and for heating in winter. The latter situation resulted in the wrong parameters for the heat pump in summer when the model is trained on winter data. To overcome this challenge, the parameters were re-estimated using two weeks for training and one for validation immediately before the test week.

Several methods have also been proposed for low resolution (1 hour) data in this area. One statistical method developed in 2013 by Morch et al. [56] used linear regression to segment hourly electricity loads from households into weather-dependent (e.g. space heating) and weather-independent loads. This method considers the dependency of energy consumption on current and past-day temperatures. Lien et al. [57] also use linear regression, seasonality and temperature dependency to generate average profiles for domestic hot water heating (DHW) in buildings based on heat load and outdoor temperature measurements. However, this method is only tested on heating loads and not electricity loads, and it is best used for generating average load profiles, rather than disaggregating total heating loads in single buildings.

Other works have investigated unsupervised methods for disaggregating heating and cooling loads from hourly energy data. Zaeri et al. [58] used unsupervised time-series decomposition to disaggregate hourly aggregate electricity load into corresponding heating and cooling loads by decomposing the total signal into trend, seasonality and residuals before comparing with submeter data. The study showed promising results, but the results are difficult to interpret and replicate, as the type and characteristics of the office building's heating system are not provided in the article. Amayri et al. [59] developed NILM methods based on random forest to disaggregate flexible electricity use in three houses. The method aimed to classify whether the hot water heater and electric heater were on or off and were shown to perform well for two of the houses, but not the third one.

Najafi et al. [60] proposed another method for disaggregation of air-conditioning load from smart meter data, in which an extended pool of input features was extracted from both smart meter data and the corresponding weather conditions' dataset. Input features included calendar-based variables and features inspired by buildings' thermal behaviour (e.g. outdoor temperature in previous timestamps), along with statistics-based, regression-based and pattern-based features that were originally proposed by Miller et al. [61] for building characterisation. A feature selection algorithm was then employed to identify the most promising set of features and an optimisation process was utilised to determine the most promising algorithm (showing that extra trees algorithm offers highest performance). The method achieved an average R2 score of 0.905 for disaggregating cooling load and thus demonstrated the utility of the proposed set of features and the fact that high-frequency data or appliance-wise measurements are not always necessary to achieve high accuracy. However, this approach was tested on a relatively small dataset and should be tested on a larger set of buildings to further assess its generalisability.

Overall, shallow learning algorithms and rule-based algorithms can be useful for building load disaggregation when interpretability and computational efficiency are needed. However, if the data is highly complex and non-linear, and a large dataset is available, deep learning models may provide better accuracy and performance.

4.2 Deep supervised learning

In the past years, deep learning methods have been increasingly utilised for disaggregation of building energy use. Models based on deep supervised learning require more computational power compared to traditional methods, although they can provide better accuracy, scalability and performance compared to regression models in disaggregation applications, specifically while dealing with complex and non-linear relationships between the aggregate electricity consumption and appliance usage. Deep learning methods consist of a range of different models such as Convolutional Neural Networks (CNN), Residual Network (ResNet), Seq2Seq and Generative Adversarial

Network (GAN), alone or in combination with each other and/or mechanisms such as Gated Recurrent Units (GRU) and Denoising Autoencoder (dAE).

Most of the proposed deep-learning based approaches for the disaggregation of heating loads that have been reviewed in this sub-chapter are designed for disaggregation of datasets with a resolution of 1 minute or more. Considering the information that can be extracted from the measurement data with such resolution, these methods are typically designed to recognise patterns in different appliances' consumption and their states. In this context, Kaselimi et al., 2019 [62] introduced a Bayesian-optimised bidirectional Long Short-Term Memory (LSTM) method for energy disaggregation of household aggregate load. The method was evaluated using the AMPds dataset. In general, the model was shown to have a higher performance than the one achieved using other methods such as LSTM, CNN, CO and FHMM, but the proposed method performed significantly worse on disaggregating the heat pump's load compared to other appliances such as the dryer, the dishwasher and the oven. Methods based on LSTM were also examined by Xia et al[63], who proposed a composite deep LSTM for load disaggregation. The method was not tested on any heating appliances, but it was tested on an air conditioner from the Dataport dataset with 1 minute's resolution. The air conditioner was used frequently in the training period and the method performed better than all traditional methods and the DAE method for disaggregation of the air conditioner's load. Wang et al. [64] proposed an ensemble-based deep learning method for NILM, which used both LSTM and feed forward NN. The model used the real power readings from the dataset and considered sliding window data and additional attributes such as month and time of day for disaggregation of six appliances from the AMPds-dataset, including the heat pump and HVAC. The method achieved 93.9% accuracy for the heat pump disaggregation.

Davies et al. [65] proposed some CNN models for appliance classification, trained and tested on a PLAID dataset. Results showed that appliance classification is possible to some extent at low "smart meter"-like sampling frequencies, but performance increases greatly with sampling resolution. In general, their CNN architectures showed good separation of appliances on a PLAID dataset, but the models performed poorly on electric heater class, however, which was confused with the hairdryer. This is because heater onset events are generated by a single heating element turning on, corresponding to a simple step shaped transient. Such appliance classes are very difficult to separate since they contain a heating element whose onset appears as a plain step.

Li et al. [66] proposed a fusion framework using an integrated neural network for NILM with two tasks: load identification and power estimation. The foundation of load identification is event detection, achieved by using the CUSUM method. Experimental results on an AMPds dataset with 1 minute's resolution showed that the proposed model could be used for NILM on datasets using low sampling-rate power data, and the method achieved 98.5% accuracy for identification of the heat pump.

Wang et al. [67] proposed an end-to-end method to identify individual appliances from aggregated data using a combination of DAE and LSTM networks on an AMPds dataset. The method was trained on aggregated data and tested on synthesised data. The results of the model showed that it had high performance for some appliances, but low performance on reconstructing appliances with continuous states (as opposed to on/off-appliances), such as a washing machine and a heat pump.

Harell et al. [68] proposed a causal 1-D CNN, inspired by WaveNet, for NILM on low-frequency data on the AMPds dataset. The study found that when implementing current, active power, reactive power and apparent power the model showed faster convergence and higher performance for NILM, but the study does not, however, present any results specifically for the heat pump or any other appliance.

Xia et al. [69] proposed sequence-to-sequence methods for NILM based on a deep delated convolution residual network. The original power data from UK-DALE and Dataport was normalised before sliding window was used to create input for the residual network. The method can improve disaggregation efficiency and the accuracy of disaggregation of electrical appliances with low usage. The method was not tested on any heating appliances, but on an air conditioner with promising performance, but the authors argue that other methods such as KNN, DEA, CNN, seq-2-point, and their own method based on DA-ResNet [70], offered just as high performance on the disaggregation of the air conditioner.

Kaselimi et al. [71] proposed a contextually adaptive and Bayesian optimised bidirectional LSTM model for modelling different household appliances' consumption patterns in a NILM operational framework. The model showed low accuracy in detecting the HPE appliance (AMPds), mainly due to recurring signal changes caused by external (seasonal) contextual conditions. Later in the same year, Kaselimi et al. [72] investigated the suitability of a GAN-based model for NILM. GAN-based models can generate longer instances of the signal waveform, thereby enhancing NILM modelling capability. The model includes a seeder component and generates specific appliance signatures in accordance with an appliance operation, and should accurately detect events occurring (e.g. switch-on events) during a day. The method was tested on two buildings, including one with a heat pump (from AMPds), with measurements performed for one month (17/5-17/6), when heat pumps are rarely in use. The model shows promising results for NILM on most appliances, performing as well or better than traditional and other deep learning methods for all appliances tested, but the study also shows that out of all appliances, all methods performed the worst for disaggregating the heat pump compared to other appliances (not for heating). This model was improved by Kaselimi in [73] by including a deep learning classifier in the discriminator component of GAN, which gave a slight improvement in disaggregation performance of the heat pump compared to [72].

Liu et al. [74] proposed a deep learning method for NILM, which used a GRU, as well as multi-feature fusion. The method considers the coupling relationship between the electrical signals of different appliances, as well as water and gas use, meaning that correlations between working states of appliances were considered in the disaggregation. They used an AMPds-dataset, which has a significant correlation between the working states of heat pump and furnace. The method improved the F1 score and accuracy greatly compared to methods that do not consider the relationship between the electricity use of the appliances, as well as the gas and water use.

Kianpoor et al. [75] proposed a deep adaptive ensemble filter based on various signal processing tools integrated with an LSTM for NILM. Their framework searches ensemble filtering techniques, including discrete wavelet transform, low-pass filter and seasonality decomposition, to find the best filtering method for disaggregating different flexible loads (e.g. heat pumps). The discrete wavelet transform gave best results for the heat pump combined with LSTM. Their study showed that using LSTM greatly improved performance compared to traditional methods, such as linear regression, and that introducing adaptive filtering improved the results even more, although the peaks of the heat pump power consumption are still not perfectly captured.

Zou et al. [76] introduced a method based on CNN and bidirectional LSTM (BiLSTM). In this approach, periodical changes in total demand (e.g. daily, weekly and seasonal variations) are disaggregated into corresponding frequency components and are correlated with the same frequency components in meteorological variables (e.g. temperature and solar irradiance), allowing selection of the combinations of frequency components with the strongest correlations as additional explanatory variables. Their study found that heating and lighting loads were identified with greater accuracy when the correlated frequency components were used as additional information during the disaggregation.

All of the research mentioned within this field has looked at datasets with resolutions of 1 minute or higher. In 2022, Hosseini [77], however, suggested an LSTM-based method for disaggregating heating demand from the aggregated load profiles, with 15 minute resolution, belonging to houses equipped with electric space heaters and water heaters. Their proposed method aims to identify major appliances by first extracting overall heating demand from the aggregated load before extracting the remaining appliances. To extract the electric space heaters, an LSTM network is used, with a sliding window that considers the past 7 instances of aggregated load and the past 8 instances of energy for the electric space heaters. It is worth noting that ambient conditions were not considered as input features in the latter disaggregation procedure. The remaining load is disaggregated through an unsupervised clustering procedure (Density-Based Spatial Clustering of Applications with Noise).

Deep supervised learning methods can be more difficult to interpret compared to traditional models and they require larger amounts of data and computational resources to train. To overcome the latter, reinforcement learning RL may be suitable for disaggregation of heating loads with high performance and less data.

4.3 Reinforcement learning

Deep learning approaches typically require large datasets in the training procedure. Although many labelled datasets exist for developing and testing disaggregation techniques, providing a large amount of perfectly labelled training data for specific application may not always be feasible. RL and deep RL is an alternative data-driven approach which requires no labelled training data. In projects with data collection of energy use measurements in buildings, it can take a long time to acquire a full set of representative data. Algorithms developed for disaggregation of heating loads can benefit strongly from having data corresponding to more than one year, as the heating demand may vary greatly from one year to another. Considering ambient conditions, such as the outdoor temperature, could in addition improve the recognition of heating loads. With RL, one can start training the algorithm on a small dataset and continue to improve the learning algorithm as the dataset grows.

Only a few methods for disaggregation of heating loads in buildings based on RL have been proposed in the literature. Li [78] proposed an NILM recognition method based on adaptive K-nearest neighbours RL (AD-KNN-RL) and compared it to other models, such as the conventional KNN, genetic algorithm (GA) and Hidden Markov Model (HMM). The method was applied to an AMPds dataset and aimed at state recognition of 5-8 different appliances, including a heat pump. It proved that the accuracy of the state recognition of electrical appliances with simple state changes such as lamps and heat pumps is higher than for other electrical appliances, but that the accuracy of electrical identification is generally low for multi-state continuous changes. AD-KNN-RL proved to have the highest performance, while HMM performed the worst.

Zaoali et al. [79] used LSTM-based reinforcement Q-learning to disaggregate the REDD dataset. The experiment showed that the accuracy of the disaggregation was significantly improved by using this method, compared to using the deep learning approach, TFIDF-DAE, achieving an accuracy of 85%. The buildings 1, 2, 3, 4 and 6 from the REDD dataset were used for training of the model, while building number 4 was used to test the algorithm. Building numbers 2 and 3 have electric heaters, while building number 4 does not, so that the disaggregation of electric heaters from the aggregate load using this approach was not tested here.

Table 3 Overview of methods used for disaggregation of building electricity load in the literature.

	Ref	Year	method	B.cat	N buildings	Location	Dataset	Res	Relevant appliance	NILM-TK
Traditional methods and shallow algorithms	Liu et al. [54]	2022	Affinity propagation clustering (AP) and time-segmented state probability (TSSP)	SFH	1	Canada	AMPds	1 min	Heat pump	No
	Morch et al. [56]	2013	Linear regression with temperature dependency	Residential	75	Norway	ElDeK	1 h	Space heating	No
	Lien et al. [57]	2020	Linear regression to find average DHW consumption	Apartment and hotels	78	Norway	treASURE	1h	Space heating	No
	Najafi et al. [60]	2020	Feature selection, Extra trees regressor	Houses	20	TX, USA	Dataport	1h	Air conditioning	No
	Zaeri et al. . [58].	2022	Unsupervised Time-series decomposition	Office	1	Ottawa, Canada	Original	1h	Unknown heating and cooling	No
	Amayri et al. [59]	2022	RF (NILM and interactive learning)	Houses	3	France	IRISE	1h and 10 min	HWH, EH and clothes dryer	No
	Balletti et al. [55]	2022	Mixed Integer Optimization	SFH	1	Canada, UK,	AMPds, UKDALE (h1 and h2) and REFIT (h3 and H9)	1 min	Heat pump (AMPds)	No
Deep supervised learning	Kaselimi et al.[62]	2019	Bayes-Bi-LSTM	Residential	1	Canada	AMPds	1 min	HP	No
	Wang et al.[64]	2019	LSTM-RNN and LSTM-FF	Residential	1	Canada	AMPds	1 min	HP+HVAC	No
	Davies et al[65]	2019	CNN, FF	Residential	55	USA	PLAID	30 kHz	HVAC + heater	No
	Li et al.[66]	2019	CNN	SFH	1	Canada	AMPds	1 min	HP	No
	Wang et al.[67]	2019	LSTM-dAE	SFH	1	Canada	AMPds	1 min	HP	No
	Harell et al[68]	2019	CNN-wn	SFH	1	Canada	AMPds	1 min	HP	No
	Xia et al.[69]	2019	CNN-seq2seq	Residential	Unknown	UK, USA	UK-DALE, dataport	6s, 1 min	Air conditioner	No
	Xia et al. [70]	2019	CNN, DA-ResNet	Residential	Unknown	UK, USA	UK-DALE, dataport	6s, 1 min	Air conditioner	No
	Kaselimi et al. [71]	2020	CoBiLSTM	SFH	1	Canada	AMPds, REDD and REFIT	6s	HP	No
	Kaselimi et al. [72]	2020	CNN-dAE-GAN	SFH	Unknown. Results for 6 appliances presented.	Canada, UK	AMPds, REFIT	1 min	HP	Yes, for benchmarking
	Kaselimi et al. in [73]	2020	CNN-dAE- GRU-GAN	SFH	Unknown. Results for 6 appliances presented.	Canada, UK	AMPds, REFIT	1 min	HP	Yes, for benchmarking
	Xia et al.[63]	2020	CNN-LSTM	Residential	Selected appliances from 4 buildings	USA	REDD, dataport	1 min	HVAC	No
	Liu et al. al [74]	2021	GRU and multi feature fusion	Residential	1	Canada	AMPds	1 min	Heat pump and gas boiler	No
	Zou et al. [76]	2021	CNN-BiLSTM	Residential	6	Canada, UK	AMPds, UK-DALE and proprietary	1 min	Heat pump (AMPds)	Yes
	Hosseini [77]	2022	LSTM	SFH	10	Quebec	Original	15min	ESH (8-10 per house)	No
	Kianpoor et al.al [75]	2023	AEF LSTM	Residential	1	Canada	AMPds	1 min	Heat pump	No
RL	Li [78]	2020	AD-KNN-RL	SFH	1	Canada	AMPDS	1 min	1 HP and 7 EA	No
	Zaoali et al. [79]	2022	recurrent LSTM-RL-Q-learning	SFH/MFH	6	USA	REDD	0,33 Hz-15kHz	EH	No

5 Evaluation of datasets and requirements

The datasets described in Table 2 are both widely used and sometimes rarely used for the development of methods for disaggregation of building heating appliances, as shown in Section 4. The content of the datasets can be summarised as follows:

Building category: 24 datasets are investigated. Four include measurements from commercial buildings and 22 include measurements from residential buildings – mostly from single-family houses but also some multi-family houses.
Sampling rate: Most of the datasets have a frequency of 1 Hz or higher, while a few datasets have only very low resolution, of minutes or per hour.
Duration: The duration of the datasets varies between 1 day and several years, with a median value of 180 days.
Locations: The datasets are from different locations – 11 datasets are from buildings in Europe, 8 from North America, 3 from India and 1 from Korea. The datasets together represent buildings from both cold and warm climates.
Appliances: 12/24 datasets contain buildings with single measurements of heating appliances or heating loads, while the rest of the datasets have no measurements connected to space heating load.
Measurements: All contain measurements of power (active) or hourly energy use (apparent). Several datasets also contain corresponding measurements with current, voltage, phase factor, reactive power and phase angle.
NILMTK compatibility: The majority of the datasets in Table 2 are available in the NILMTK format as hdf5 files and with available converters.

Most of the datasets investigated include high-resolution data (1 second or lower). However, for real world applications, energy measurement data is usually available at a much lower resolution (15-60 minutes). Hosseini [77] shows that their suggested model performs efficiently with low-resolution data (15 minutes) in identifying most of the ESH loads (electric heaters), although the model performs inadequately in capturing the peaks and causing unwanted variations in lower demand. Najafi et al. [60], however, achieved a high R2 value for recognising AC loads through the use of feature selection. High-resolution data measurements are widely used for development of disaggregation methods, but may not be applicable to hourly datasets, which are far more available and more commonly used in real-life applications.

Space heating loads are highly dependent on outdoor temperature, season, time of day, type of day (weekday/end) and building metadata (such as building type, heating appliances and energy efficiency, etc.) [6]. Buildings with electrical heating typically exhibit significant fluctuations in their load profiles. These fluctuations stem from the varying outdoor temperatures, which can differ substantially from year to year. Consequently, datasets with extended time spans prove exceptionally valuable. Most of the datasets in Table 2 contain less than one year of data.

Given the notable difference between information that is available in measured load profiles with low and high resolution, the corresponding pipelines benefit from being treated differently. The methods proposed for disaggregating high-resolution load profiles are typically designed to recognise patterns in different appliances and their states, while this information is often lost when moving into low resolution (15 minutes to 1 hour). For the specific case of disaggregating the heating loads from low-resolution datasets, the pipeline may benefit from generating and employing features that are inspired by the thermal behaviour of buildings (e.g. the lagged values of ambient conditions such as outdoor temperature and the corresponding seasonality) to capture additional information that is not available in the load measurement's data.

Some datasets in Table 2 include measurements for heating appliances, but these are typically one single heating appliance per building. In the Norwegian setting, it is common to have more than one heating appliance per building, e.g. combinations of electric floor heating, electric panel heaters, air-to-air heat pump, electric boilers, electric water heaters, ground source heat pumps, etc. [80]. Several of these electric heating appliances and their combinations are not found in the existing datasets.

The availability of metadata varies for different datasets. The heating appliances and heating distribution system in a building greatly affect the use of electric heating appliances. An electric boiler used for both heating of domestic hot water and space heating typically has a different user pattern compared to an electric boiler solely used for one of these purposes. An air-to-air heat pump is typically used differently in a single-family house that also has access to non-electric heating options. The heating systems of the buildings in the different datasets are not always available to users, but could provide useful information for the disaggregation of the heating appliances.

While some datasets like AMPds contains hourly climate data, several of the datasets investigated include climate/ambient data or district heating data, in addition to measured electricity consumption. Space heating is highly dependent on outdoor temperature and climate conditions. This is rarely considered in traditional disaggregation approaches implemented in NILMTK. Although NILMTK can take in temperature and gas measurements, it is not utilised in the implemented methods. However, NILMTK currently does not support heating measurements from district heating.

6 Conclusion

This review paper has investigated existing approaches for disaggregation of space heating loads and appliances from buildings' total energy load that utilise traditional methods, shallow learning methods, deep supervised learning and RL methods. Previous research shows that several approaches have disaggregated single heating appliances from total electricity load. These methods are often applied to high-resolution energy measurements (50-60 Hz) from buildings, with varying durations of training datasets. Deep learning methods are shown to offer higher disaggregation performance compared to traditional and shallow learning methods such as FHMM, CO, Mean, etc. The review also shows that RL approaches for disaggregation are promising, but with a limited number of studies that can be further investigated in the future. Most of the disaggregation approaches investigated typically recognise and disaggregate single appliances, and the majority of them look at the heat pump from AMPds, or other single appliances such as heat pumps or electric space heaters from datasets with buildings with these appliances. However, few methods have been proposed in the literature for disaggregating total electricity use for space heating from the total electricity use in buildings with electric heating, or the loads of heat appliances in buildings with more than one heating appliance. At the same time, there is a demand for disaggregation algorithms tailored for cold climates that include electric heating from different heating options that work on low-frequency data. However, only a few methods have considered disaggregating the consumption of heating units based on features that are inspired from the thermal behaviour of buildings (e.g. lagged values of outdoor conditions or relationship between heating consumption and outdoor temperature). In most of the studies, power consumed by heating systems is commonly disaggregated using the same pipelines as those utilised for other appliances. Recommendations for future work include integrating deep supervised learning techniques with features inspired by building physics to develop pipelines for disaggregating heating loads from low-resolution aggregate electricity data, as this method shows significant promise, and as building energy data is mostly available as low-resolution data. Additionally, there remains a notable research gap in the disaggregation of heating loads in both commercial and public buildings, as well as in the application of reinforcement learning methods for this purpose.

7 Acknowledgment

This article has been written within the research project “Coincidence factors and peak loads of buildings in the Norwegian low carbon society” (COFACTOR). The authors gratefully acknowledge the support from the Research Council of Norway (project number 326891), research partners, industry partners and data providers.

8 References

[1] IEA, ‘https://www.iea.org/reports/buildings’, 2022.

[2] Energy Transitions Commission, ‘Making Mission Possible’, *Health Prog.*, vol. 76, no. 6, pp. 45-7, 60, 2020.

[3] SSB, ‘11561: Energibalanse. Tilgang og forbruk, etter energiprodukt 1990 - 2018’, 2020. https://www.ssb.no/statbank/table/11561 (accessed Jun. 02, 2020).

[4] L. Ødegården and S. Bhantana, *Status og prognoser for kraftsystemet 2018 rapportnr. 103-2018*. NVE, 2018. Accessed: May 31, 2023. [Online]. Available: http://publikasjoner.nve.no/rapport/2018/rapport2018_103.pdf

[5] NVE, ‘NVE.no: Smarte strømmålere (AMS)’, 2022. https://www.nve.no/reguleringsmyndigheten/kunde/strom/stromkunde/smarte-stroemmaalere-ams/

[6] K. B. Lindberg, *Doctoral thesis Impact of Zero Energy Buildings on the Power System A study of load profiles , flexibility and Impact of Zero Energy Buildings on the Power System A study of load profiles , flexibility and system*, vol. 6. 2017.

[7] K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert, ‘Is disaggregation the holy grail of energy efficiency? The case of electricity’, *Energy Policy*, vol. 52, pp. 213–234, 2013, doi: 10.1016/j.enpol.2012.08.062.

[8] G. W. Hart, ‘Nonintrusive appliance load monitoring’, *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992, doi: 10.1109/5.192069.

[9] P. Huber, A. Calatroni, A. Rumsch, and A. Paice, 'Review on Deep Neural Networks Applied to Low-Frequency NILM', *Energies*, vol. 14, no. 9, Art. no. 9, Jan. 2021, doi: 10.3390/en14092390.

[10] M. Zeifman and K. Roth, 'Nonintrusive appliance load monitoring: Review and outlook', *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 76–84, Feb. 2011, doi: 10.1109/TCE.2011.5735484.

[11] S. Makonin, 'Approaches to Non-Intrusive Load Monitoring (NILM) in the Home APPROACHES TO NON-INTRUSIVE LOAD MONITORING (NILM) by Doctor of Philosophy School of Computing Science', no. October 2012, 2014.

[12] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, 'Improving Nonintrusive Load Monitoring Efficiency via a Hybrid Programing Method', *IEEE Trans. Ind. Inform.*, vol. 12, no. 6, pp. 2148–2157, 2016, doi: 10.1109/TII.2016.2590359.

[13] L. Farinaccio and R. Zmeureanu, 'Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses', *Energy Build.*, vol. 30, no. 3, pp. 245–259, 1999, doi: 10.1016/S0378-7788(99)00007-9.

[14] M. B. Figueiredo, A. De Almeida, and B. Ribeiro, 'An experimental study on electrical signature identification of non-intrusive load monitoring (NILM) systems', *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, vol. 6594 LNCS, no. PART 2, pp. 31–40, 2011, doi: 10.1007/978-3-642-20267-4_4.

[15] K. T. Nguyen *et al.*, 'Event Detection and Disaggregation Algorithms for NIALM System', *NILM Workshop*, no. June 2014, pp. 2–5, 2014.

[16] H. Gonçalves, A. Ocneanu, and M. Bergés, 'Unsupervised disaggregation of appliances using aggregated consumption data', *Environ*, 2011.

[17] L. Wang, X. Luo, and W. Zhang, 'Unsupervised energy disaggregation with factorial hidden Markov models based on generalized backfitting algorithm', in *EEE International Conference of IEEE Region 10 (TENCON 2013)*, 2013.

[18] D. Egarter, V. P. Bhuvana, and W. Elmenreich, 'PALDi: Online load disaggregation via particle filtering', *IEEE Trans. Instrum. Meas.*, vol. 64, no. 2, pp. 467–477, 2015, doi: 10.1109/TIM.2014.2344373.

[19] Q. Wu and F. Wang, 'Concatenate convolutional neural networks for non-intrusive load monitoring across complex background', *Energies*, vol. 12, no. 8, 2019, doi: 10.3390/en12081572.

[20] J. Kelly *et al.*, 'NILMTK v0.2: a non-intrusive load monitoring toolkit for large scale datasets: demo abstract', in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, in BuildSys '14. New York, NY, USA: Association for Computing Machinery, Nov. 2014, pp. 182-183. doi: 10.1145/2674061.2675024.

[21] N. Batra *et al.*, 'Towards reproducible state-of-the-art energy disaggregation', in *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, New York NY USA: ACM, Nov. 2019, pp. 193-202. doi: 10.1145/3360322.3360844.

[22] N. Virtsionis Gkalinikis, C. Nalmpantis, and D. Vrakas, 'Torch-NILM: An Effective Deep Learning Toolkit for Non-Intrusive Load Monitoring in Pytorch', *Energies*, vol. 15, no. 7, Art. no. 7, Jan. 2022, doi: 10.3390/en15072647.

[23] K. B. Lindberg, 'Impact of Zero Energy Buildings on the Power System', p. 192.

[24] A. R. Abbasi, 'Fault detection and diagnosis in power transformers: a comprehensive review and classification of publications and methods', *Electr. Power Syst. Res.*, vol. 209, p. 107990, Aug. 2022, doi: 10.1016/j.epsr.2022.107990.

[25] S. Makonin, B. Ellert, I. V. Bajić, and F. Popowich, 'Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014', *Sci. Data*, vol. 3, no. 1, Art. no. 1, Jun. 2016, doi: 10.1038/sdata.2016.37.

[26] Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, and A. Amira, ‘Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives’, *Appl. Energy*, vol. 287, p. 116601, Apr. 2021, doi: 10.1016/j.apenergy.2021.116601.

[27] Y. Himeur, A. Alsalemi, F. Bensaali, A. Amira, and A. Al-Kababji, ‘Recent trends of smart nonintrusive load monitoring in buildings: A review, open challenges, and future directions’, *Int. J. Intell. Syst.*, vol. 37, no. 10, pp. 7124–7179, 2022, doi: 10.1002/int.22876.

[28] G.-F. Angelis, C. Timplalexis, S. Krinidis, D. Ioannidis, and D. Tzovaras, ‘NILM applications: Literature review of learning approaches, recent developments and challenges’, *Energy Build.*, vol. 261, p. 111951, Apr. 2022, doi: 10.1016/j.enbuild.2022.111951.

[29] A. Ruano, A. Hernandez, J. Ureña, M. Ruano, and J. Garcia, ‘NILM Techniques for Intelligent Home Energy Management and Ambient Assisted Living: A Review’, *Energies*, vol. 12, no. 11, Art. no. 11, Jan. 2019, doi: 10.3390/en12112203.

[30] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava, ‘It’s Different: Insights into home energy consumption in India’, in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, in BuildSys’13. New York, NY, USA: Association for Computing Machinery, Nov. 2013, pp. 1–8. doi: 10.1145/2528282.2528293.

[31] J. Kelly and W. Knottenbelt, ‘The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes’, *Sci. Data*, vol. 2, no. 1, Art. no. 1, Mar. 2015, doi: 10.1038/sdata.2015.7.

[32] ‘NILMTK: Non-Intrusive Load Monitoring Toolkit’. nilmtk, Feb. 27, 2023. Accessed: Feb. 28, 2023. [Online]. Available: https://github.com/nilmtk/nilmtk

[33] D. Murray, L. Stankovic, and V. Stankovic, ‘An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study’, *Sci. Data*, vol. 4, no. 1, Art. no. 1, Jan. 2017, doi: 10.1038/sdata.2016.122.

[34] J. Z. Kolter and M. J. Johnson, ‘REDD: A Public Dataset for Energy Disaggregation Research’.

[35] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, ‘The ECO dataset and the performance of non-intrusive load monitoring algorithms’, in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, Memphis Tennessee: ACM, Nov. 2014, pp. 80–89. doi: 10.1145/2674061.2674064.

[36] 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, Art. no. 1, Apr. 2020, doi: 10.1038/s41597-020-0434-6.

[37] C. Shin, E. Lee, J. Han, J. Yim, W. Rhee, and H. Lee, ‘The ENERTALK dataset, 15 Hz electricity consumption data from 22 houses in Korea’, *Sci. Data*, vol. 6, no. 1, p. 193, Oct. 2019, doi: 10.1038/s41597-019-0212-5.

[38] ‘Household Electricity Survey - Data Briefing’.

[39] R. Medico *et al.*, ‘A voltage and current measurement dataset for plug load appliance identification in households’, *Sci. Data*, vol. 7, no. 1, Art. no. 1, Feb. 2020, doi: 10.1038/s41597-020-0389-7.

[40] K. D. Anderson, Ocneanu, Adrian Filip, D. Benitez, D. Carlson, A. Rowe, and M. Berges, ‘BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research’, Aug. 2023.

[41] A. Reinhardt *et al.*, *On the accuracy of appliance identification based on distributed load metering data*. 2012, p. 9.

[42] M. Maasoumy, B. M. Sanandaji, K. Poolla, and A. S. Vincentelli, 'BERDS-BERkeley EneRgy Disaggregation Dataset'. https://tokhub.github.io/dbecd/links/berds.html (accessed Jan. 24, 2023).

[43] A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro, and A. M. Tonello, 'GREEND: An energy consumption dataset of households in Italy and Austria', in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Nov. 2014, pp. 511–516. doi: 10.1109/SmartGridComm.2014.7007698.

[44] A. S. N. Uttama Nambi, A. Reyes Lua, and V. R. Prasad, 'LocED: Location-aware Energy Disaggregation Framework', in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, Seoul South Korea: ACM, Nov. 2015, pp. 45–54. doi: 10.1145/2821650.2821659.

[45] O. Parson *et al.*, 'Dataport and NILMTK: A building dataset designed for non-intrusive load monitoring', in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL, USA: IEEE, Dec. 2015, pp. 210–214. doi: 10.1109/GlobalSIP.2015.7418187.

[46] S. Makonin, Z. J. Wang, and C. Tumpach, 'RAE: The Rainforest Automation Energy Dataset for Smart Grid Meter Data Analysis', *Data*, vol. 3, no. 1, p. 8, Feb. 2018, doi: 10.3390/data3010008.

[47] H. Rashid, P. Singh, and A. Singh, 'I-BLEND, a campus-scale commercial and residential buildings electrical energy dataset', *Sci. Data*, vol. 6, no. 1, Art. no. 1, Feb. 2019, doi: 10.1038/sdata.2019.15.

[48] K. Basu, V. Debusschere, and S. Bacha, 'Residential appliance identification and future usage prediction from smart meter', in *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, Nov. 2013, pp. 4994–4999. doi: 10.1109/IECON.2013.6699944.

[49] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, 'Smart*: An Open Dataset and Tools for Enabling Research in Sustainable Homes', *Proc SustKDD*, Jan. 2012.

[50] N. Batra, O. Parson, M. Berges, A. Singh, and A. Rogers, *A comparison of non-intrusive load monitoring methods for commercial and residential buildings*. 2014.

[51] H. Sæle, E. Rosenberg, and N. Feilberg, 'State-of-the-art Projects for estimating the electricityu end-use demand', 2010. Accessed: Mar. 20, 2023. [Online]. Available: https://www.sintef.no/globalassets/project/eldek/publisering/tr-a6999-state-of-the-art-projects-for-estimating-the-electricity-end-use-demand.pdf

[52] K. Heimar Andersen, S. Krekling Lien, K. Byskov Lindberg, H. Taxt Walnum, and I. Sartori, 'Further development and validation of the "PROFet" energy demand load profiles estimator', presented at the 2021 Building Simulation Conference, Sep. 2021. doi: 10.26868/25222708.2021.30159.

[53] M. Schaffer, T. Tvedebrink, and A. Marszal-Pomianowska, 'Three years of hourly data from 3021 smart heat meters installed in Danish residential buildings', *Sci. Data*, vol. 9, no. 1, Art. no. 1, Jul. 2022, doi: 10.1038/s41597-022-01502-3.

[54] L. Liu *et al.*, 'Non-Intrusive Load Monitoring Method Considering the Time-Segmented State Probability', *IEEE Access*, vol. 10, pp. 39627–39637, 2022, doi: 10.1109/ACCESS.2022.3167132.

[55] M. Balletti, V. Piccialli, and A. M. Sudoso, 'Mixed-Integer Nonlinear Programming for State-Based Non-Intrusive Load Monitoring', *IEEE Trans. Smart Grid*, vol. 13, no. 4, pp. 3301–3314, Jul. 2022, doi: 10.1109/TSG.2022.3152147.

[56] A. Morch, H. Sæle, N. Feilberg, and K. B. Lindberg, 'Method for development and segmentation of load profiles for different final customers and appliances', *ECEEE 2013 Summer Stud. Proc.*, 2013, Accessed: Mar. 20, 2023. [Online]. Available: https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2013/7-

monitoring-and-evaluation/method-for-development-and-segmentation-of-load-profiles-for-different-final-customers-and-appliances/
[57] S. K. Lien, D. Ivanko, and I. Sartori, *Domestic hot water decomposition from measured total heat load in Norwegian buildings*. SINTEF Academic Press, 2020. Accessed: Jan. 17, 2023. [Online]. Available: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2684373
[58] N. Zaeri, H. B. Gunay, and A. Ashouri, 'Unsupervised energy disaggregation using time series decomposition for commercial buildings', in *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, Boston Massachusetts: ACM, Nov. 2022, pp. 373–377. doi: 10.1145/3563357.3566155.
[59] M. Amayri, C. S. Silva, H. Pombeiro, and S. Ploix, 'Flexibility characterization of residential electricity consumption: A machine learning approach', *Sustain. Energy Grids Netw.*, vol. 32, p. 100801, 2022.
[60] B. Najafi, L. Di Narzo, F. Rinaldi, and R. Arghandeh, 'Machine learning based disaggregation of air-conditioning loads using smart meter data', *IET Gener. Transm. Distrib.*, vol. 14, no. 21, pp. 4755–4762, 2020, doi: 10.1049/iet-gtd.2020.0698.
[61] C. Miller and F. Meggers, 'Mining electrical meter data to predict principal building use, performance class, and operations strategy for hundreds of non-residential buildings', *Energy Build.*, vol. 156, pp. 360–373, Dec. 2017, doi: 10.1016/j.enbuild.2017.09.056.
[62] M. Kaselimi, N. Doulamis, A. Doulamis, A. Voulodimos, and E. Protopapadakis, 'Bayesian-optimized Bidirectional LSTM Regression Model for Non-intrusive Load Monitoring', in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom: IEEE, May 2019, pp. 2747–2751. doi: 10.1109/ICASSP.2019.8683110.
[63] M. Xia, W. Liu, K. Wang, W. Song, C. Chen, and Y. Li, 'Non-intrusive load disaggregation based on composite deep long short-term memory network', *Expert Syst. Appl.*, vol. 160, p. 113669, Dec. 2020, doi: 10.1016/j.eswa.2020.113669.
[64] J. Wang, S. El Kababji, C. Graham, and P. Srikantha, 'Ensemble-Based Deep Learning Model for Non-Intrusive Load Monitoring', in *2019 IEEE Electrical Power and Energy Conference (EPEC)*, Oct. 2019, pp. 1–6. doi: 10.1109/EPEC47565.2019.9074816.
[65] P. Davies, J. Dennis, J. Hansom, W. Martin, A. Stankevicius, and L. Ward, 'Deep Neural Networks for Appliance Transient Classification', in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 8320–8324. doi: 10.1109/ICASSP.2019.8682658.
[66] C. Li, R. Zheng, M. Liu, and S. Zhang, 'A Fusion Framework using Integrated Neural Network Model for Non-Intrusive Load Monitoring', in *2019 Chinese Control Conference (CCC)*, Guangzhou, China: IEEE, Jul. 2019, pp. 7385–7390. doi: 10.23919/ChiCC.2019.8865721.
[67] T. S. Wang, T. Y. Ji, and M. S. Li, 'A New Approach for Supervised Power Disaggregation by Using a Denoising Autoencoder and Recurrent LSTM Network', in *2019 IEEE 12th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*, Toulouse, France: IEEE, Aug. 2019, pp. 507–512. doi: 10.1109/DEMPED.2019.8864870.
[68] A. Harell, S. Makonin, and I. V. Bajic, 'Wavenilm: A Causal Neural Network for Power Disaggregation from the Complex Power Signal', in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom: IEEE, May 2019, pp. 8335–8339. doi: 10.1109/ICASSP.2019.8682543.

[69] M. Xia, W. Liu, K. Wang, X. Zhang, and Y. Xu, 'Non-intrusive load disaggregation based on deep dilated residual network', *Electr. Power Syst. Res.*, vol. 170, pp. 277–285, May 2019, doi: 10.1016/j.epsr.2019.01.034.

[70] M. Xia, W. Liu, Y. Xu, K. Wang, and X. Zhang, 'Dilated residual attention network for load disaggregation', *Neural Comput. Appl.*, vol. 31, no. 12, pp. 8931–8953, Dec. 2019, doi: 10.1007/s00521-019-04414-3.

[71] M. Kaselimi, N. Doulamis, A. Voulodimos, E. Protopapadakis, and A. Doulamis, 'Context Aware Energy Disaggregation Using Adaptive Bidirectional LSTM Models', *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3054–3067, Jul. 2020, doi: 10.1109/TSG.2020.2974347.

[72] M. Kaselimi, A. Voulodimos, E. Protopapadakis, N. Doulamis, and A. Doulamis, 'EnerGAN: A GENERATIVE ADVERSARIAL NETWORK FOR ENERGY DISAGGREGATION', in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain: IEEE, May 2020, pp. 1578–1582. doi: 10.1109/ICASSP40776.2020.9054342.

[73] M. Kaselimi, N. Doulamis, A. Voulodimos, A. Doulamis, and E. Protopapadakis, 'EnerGAN++: A Generative Adversarial Gated Recurrent Network for Robust Energy Disaggregation', *IEEE Open J. Signal Process.*, vol. 2, pp. 1–16, 2021, doi: 10.1109/OJSP.2020.3045829.

[74] H. Liu, C. Liu, L. Tian, H. Zhao, and J. Liu, 'Non-intrusive Load Disaggregation Based on Deep Learning and Multi-feature Fusion', in *2021 3rd International Conference on Smart Power & Internet Energy Systems (SPIES)*, Shanghai, China: IEEE, Sep. 2021, pp. 210–215. doi: 10.1109/SPIES52282.2021.9633819.

[75] N. Kianpoor, B. Hoff, and T. Østrem, 'Deep Adaptive Ensemble Filter for Non-Intrusive Residential Load Monitoring', *Sensors*, vol. 23, no. 4, Art. no. 4, Jan. 2023, doi: 10.3390/s23041992.

[76] M. Zou, S. Zhu, J. Gu, L. M. Korunovic, and S. Z. Djokic, 'Heating and Lighting Load Disaggregation Using Frequency Components and Convolutional Bidirectional Long Short-Term Memory Method', *Energies*, vol. 14, no. 16, Art. no. 16, Jan. 2021, doi: 10.3390/en14164831.

[77] S. S. Hosseini, B. Delcroix, N. Henao, K. Agbossou, and S. Kelouwani, 'A Case Study on Obstacles to Feasible NILM Solutions for Energy Disaggregation in Quebec Residences', *Power*, vol. 2, p. 4, 2022.

[78] H. Li, 'A Non-intrusive home load identification method based on adaptive reinforcement learning algorithm', *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 853, no. 1, p. 012030, May 2020, doi: 10.1088/1757-899X/853/1/012030.

[79] K. Zaouali, M. L. Ammari, and R. Bouallegue, 'LSTM-Based Reinforcement Q Learning Model for Non Intrusive Load Monitoring', in *Advanced Information Networking and Applications*, L. Barolli, F. Hussain, and T. Enokido, Eds., in Lecture Notes in Networks and Systems. Cham: Springer International Publishing, 2022, pp. 1–13. doi: 10.1007/978-3-030-99619-2_1.

[80] NVE and SSB, *Oppvarming i boliger. Kartlegging av oppvarmingsutstyr of effektiviseringstiltak i husholdningene. NVE Rapport 85-2014.* 2014. Accessed: Mar. 23, 2023. [Online]. Available: https://publikasjoner.nve.no/rapport/2014/rapport2014_85.pdf