

Non-Intrusive Load Monitoring (NILM) using Deep Neural Networks: A Review

Mohammad Irani Azad
Faculty of ECE
Qom University of Technology
Qom, Iran
iraniazad.m@qut.ac.ir

Roozbeh Rajabi
Faculty of ECE
Qom University of Technology
Qom, Iran
rajabi@qut.ac.ir

Abouzar Estebarsari
School of the Built Environment and Architecture
London South Bank University
London, United Kingdom
estebasaa@lsbu.ac.uk

Abstract—Demand-side management now encompasses more residential loads. To efficiently apply demand response strategies, it's essential to periodically observe the contribution of various domestic appliances to total energy consumption. Non-intrusive load monitoring (NILM), also known as load disaggregation, is a method for decomposing the total energy consumption profile into individual appliance load profiles within the household. It has multiple applications in demand-side management, energy consumption monitoring, and analysis. Various methods, including machine learning and deep learning, have been used to implement and improve NILM algorithms. This paper reviews some recent NILM methods based on deep learning and introduces the most accurate methods for residential loads. It summarizes public databases for NILM evaluation and compares methods using standard performance metrics.

Index Terms—Smart Grids, NILM, Deep Learning, Energy Management.

I. INTRODUCTION

The non-intrusive load monitoring (NILM) method has gained popularity in recent years as a way to monitor appliance and electrical utility energy usage in buildings and events (on/off) using a single energy meter. If consumers had data on appliance-level energy usage, they could better understand their energy consumption behavior and take action to reduce it. The aim of this study is to present an overview of the latest algorithms currently being investigated by researchers to create a precise non-intrusive load monitoring (NILM) method for effective energy management. The article discusses the potential applications of NILM across different fields, along with future research objectives. The development of sustainable and smart cities has been made possible by advancements in artificial intelligence (AI), smart meters, the internet of things (IoT), and smart grids, as cited in [1] and [2]. Effective energy management is a crucial component of sustainable city development, which aims to utilize resources responsibly, protect the environment, and enhance society's well-being. The objective of energy management is to promote energy system self-reliance and sustainability [1].

Energy management involves monitoring and controlling electrical utilities to optimize energy use and reduce consumption. However, with the increase in energy needs,

energy conservation has become a challenge in recent years [3]. Greater energy use can lead to an energy crisis, climate change, and a negative impact on the economy [4]. It is estimated that the rise in carbon emissions will increase global temperatures by 2.5 to 10 °C this century, causing more frequent floods, droughts, a rise in sea level, and the spread of infectious illnesses [5]. Therefore, it is essential to reduce carbon emissions across all sectors, including construction, industry, and transportation, to mitigate climate change. Researchers are working on developing technology solutions for energy conservation [3]. Buildings are one of the major contributors to energy consumption [6], with energy consumption in this sector steadily increasing over time. In order to mitigate carbon emissions, optimizing energy consumption in residential and commercial buildings is crucial. This can be achieved through the construction or design of energy-efficient structures, as well as improving energy usage in existing buildings.

The paper is organized as follows. Section II introduces the mathematical definition of the NILM problem. Section III discusses deep learning-based NILM methods. Section IV provides a summary of the public NILM datasets. Section V presents a comparison study of NILM methods, and finally, Section VI concludes the paper.

II. NILM PROBLEM DEFINITION

A. Mathematical Problem Definition

The issue at hand can be described as follows: at a given time t , the total active power consumed by a system is represented by $y(t)$, while $y_i(t)$ represents the active power consumed by the i th appliance at the same time. The overall load is the sum of the energy consumed by individual appliances and an unmeasured residual load, expressed as:

$$y(t) = \sum_{i=1}^N y_i(t) + e(t), \quad (1)$$

where N denotes the number of appliances considered, and $e(t)$ represents the undetermined residual load. The aim is to estimate $F(y(t))$ by determining the values of $y_i(t)$, given only the value of $y(t)$, as:

$$y_1(t), y_2(t), \dots, y_i(t), \dots, y_N(t) = F(y(t)), \quad (2)$$

where F is an operator that produces N distinct values when applied to the total active power. These numbers represent the most accurate estimate of the power consumed by each appliance. It should be noted that $y_i(t)$ typically does not reflect the entire set of home appliances but rather a subset of them. As a result, the unknown term $e(t)$ takes into account the loads caused by unmonitored appliances. If simultaneous measurements of the aggregate consumption and load of each appliance are available, approximating the F operator can be considered a supervised learning problem. When mainly concerned with activation times and cumulative consumption, as is the case in real situations, the estimated individual appliance consumption ($\hat{y}_i(t)$) can be obtained using functions that are constant over the device's activation period:

$$\hat{y}_i(t) = p_i \hat{a}_i(t), \quad (3)$$

where p_i represents the average consumption of appliance i , and $\hat{a}_i(t)$ represents an estimate of the activation state of the particular appliance at time t . Its value is one if the device is in use and uses energy, and zero otherwise. Therefore, starting with the aggregate load, a technique is provided to derive the most accurate and feasible assessment of the activation state of the appliances:

$$\hat{a}_1(t), \hat{a}_2(t), \dots, \hat{a}_i(t), \dots, \hat{a}_N(t) = F_a(y(t)), \quad (4)$$

After learning the average nominal consumption of the considered equipment, one can use Equation 3 to estimate consumption.

B. Appliance types

Based on their operational characteristics, appliances can be classified into four types as discussed in [7]. Type I appliances have two modes of operation - on and off. These include appliances such as kettles, toasters, and light bulbs, which consume energy only when turned on. Type I appliances are predominantly resistive with few linear reactive components. Type II appliances are characterized as multi-state or finite state machines with a limited number of operational states that may be run repeatedly. Changes in these appliances' states can be observed by monitoring the power consumption's falling/rising edges over time. Stove burners, refrigerators, and washing machines are some examples of Type II appliances [7], [8]. Figure 1 demonstrates the distinct appliance operation conditions.

Category III appliances, also known as Continuously Variable Devices (CVDs), exhibit a non-repetitive power usage pattern, which poses a challenge for energy consumption disaggregation. Examples of Type III appliances include power drills and dimmer lights [8]. Type IV appliances are those that run continuously for extended periods of time, typically lasting several days or weeks. Examples of Type IV equipment include wireless telephone devices and cable TV receivers [8].

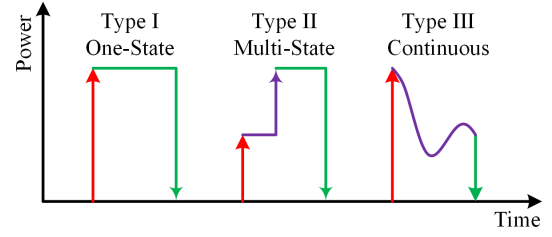


Fig. 1: A pictorial representation illustrates the different types of appliances categorized based on their operating states.

Therefore, the Non-Intrusive Load Monitoring (NILM) system is required to have the ability to differentiate between various types of appliance events, which may happen concurrently or independently and at varying time intervals.

III. DEEP LEARNING BASED NILM METHODS

NILM techniques can be broadly categorized into two groups: supervised and unsupervised methods [9]. In supervised NILM, individual appliance power usage is used to train the models. On the other hand, unsupervised methods can only utilize aggregate power usage data. Examples of unsupervised NILM techniques include Hidden Markov models (HMM) [10], [11], factorial HMM (FHMM) [12], [13], and techniques based on event detection and clustering [14], [15]. These techniques have been thoroughly examined in previous studies [9], [16]. With the advent of deep neural networks (DNNs), many neural network-based supervised NILM techniques have been developed [17], [18]. Convolutional neural networks (CNN) have also recently made significant advances [19], [20]. Graph signal processing [21], HMM [12], [13], [15], [22], [23], and DNNs [24], [25] are commonly used in suggested NILM approaches. As the cost of employing appliance data for training has grown dramatically, researchers have focused on developing unsupervised approaches and incorporating appliance models. Despite the significant progress made in NILM research in recent years, challenges remain in terms of application, identification accuracy, training time, and online deployment techniques in smart metering frameworks.

A. Event-Based Non-Intrusive Detection

The event-based NILM method is based on the concept of detecting and categorizing events within a combined electrical signal. Figure 2 shows the block diagram of the this approach. A robust event detector should be developed to cope with noisy fluctuations and identify events with decay and growth patterns, which is a bottleneck and inherent difficulty in existing event detectors [26]. One approach includes the steps of event detection, extraction, clustering, and matching in the event-based block [27]. It should be noted that the accuracy of previous event-based frameworks is dependent on the power features that are introduced. Since some appliances may have identical active power curves but radically distinct reactive

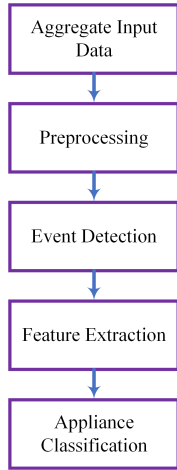


Fig. 2: Block diagram scheme of event-based NILM.

power trends, increasing the number of features can enhance the accuracy of the appliance model, particularly for non-linear loads. One of the advantages of incorporating reactive power is that it enables discrimination between different types of devices.

B. NILM Disaggregation by CNN

The proposed method [28] employs a convolutional neural network (CNN) that takes the time interval of a home’s energy consumption as input and predicts the activation status of each device at every time step. The network architecture, referred to as Temporal Pooling NILM (TP-NILM), is an updated version of Zhao et al.’s PSPNet (Pyramid Scene Parsing Network) used for semantic image segmentation [29]. The TP-NILM follows the conventional approach to image segmentation, with an encoder comprising pooling and convolutional layers that enhance the feature space of the signal but decrease its temporal resolution, and a decoder module that uses these features to approximate the activation state of the devices at the original resolution. To establish a temporal context that covers extended periods without compromising the signal’s resolution, the TP-NILM incorporates a Temporal Pooling module that accumulates features at various resolutions, enabling accurate reconstruction of the activation state.

Figure 3 illustrates the architecture of the network used in this study. The encoder uses a rectified linear unit activation function, batch-independent normalization downstream of the activations, and a regularization dropout layer, with three convolutional filters interleaved by max-pooling layers. The encoder reduces the signal’s temporal resolution by a factor of eight and raises the number of output characteristics from a single aggregate power consumption value to 256. The TP block provides context information to the decoding block, allowing it to create extra features for decoding by aggregating encoder output with various resolutions. The encoder output is passed through four average pooling modules

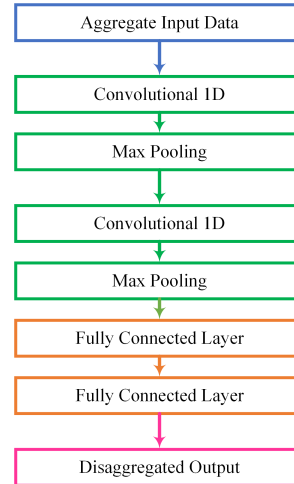


Fig. 3: Outline of network structure for NILM by CNN.

with various filter sizes, which degrade the temporal resolution while maintaining the number of features, before being convoluted with a unit filter size. This reduces the number of characteristics to one-quarter of those used in the input. The convolutional results are used as the input for a rectified linear unit activation function, which is then batch normalized. Lastly, linear up-sampling yields a temporal resolution at the TP block’s output that is equal to the encoder’s output. A dropout is also added to the output, allowing the network to be controlled. The context features generated by this block are connected to the encoder’s detail features, doubling the total number of features in the decoder’s input. To raise the signal’s temporal resolution and lower the number of features, the decoder contains a transposed convolutional layer with a stride and kernel size of 8. The ReLU is still used as the activation function, followed by an extra convolutional layer with an unified kernel size that keeps the temporal resolution and increases the number of output channels to match the number of devices being analyzed. In the output, a sigmoid function is used. This is because in the semantic segmentation of photos, each pixel is connected to a single class, however in the current application, many appliances might be in use at the same time. While working in this manner, the network decomposes all appliances at the same time. This should enable an encoder to use more broad convolutive filters that aren’t specialized for a single kind of appliance, boosting the neural network’s capacity to generalize. Gradient descent optimization may be used to find the net’s weights. The loss function is a binary crossentropy applied to each output channel that assesses the disparity between the activations predicted by the net $\hat{a}_i(t)$ and the actual ones $a_i(t)$ for each appliance under consideration and for each instant of the period under consideration.

IV. NILM PUBLIC DATASETS

To develop NILM (non-intrusive load monitoring) algorithms and assess their performance, the research

community provides various NILM datasets in the public domain [30]. Since each dataset monitors different appliances in diverse environments and buildings over a varied time period, each dataset has its own specific criteria [31], [32]. However, it has been observed that many public datasets have structural variations that require pre-processing before usage. To address this issue, the dsCleaner Python module was developed to standardize, clean, and convert time series data into a consistent file format, and also includes a resampling method for datasets. Typically, NILM datasets consist of aggregated energy data from a single meter and the actual energy consumption of each appliance, which is measured by plug-level meters and serves as the ground truth for evaluating NILM algorithms. Table I lists the most popular publicly accessible NILM datasets for research purposes [30].

In 2011, the Reference Energy Disaggregation Dataset (REDD) [32] became available as the first openly accessible dataset designed specifically to aid NILM research. Following this, the building-Level fully-labeled dataset for Electricity Disaggregation (BLUED) [33] was released in 2012, which contained data from a single household.

The Almanac of Minutely Power dataset (AMPds) [34], on the other hand, was made public in 2013 and comprised both aggregate and sub-metered power data from a single household. The Almanac of Minutely Power dataset Version 2 (AMPds2) [35] is another dataset that captures all three primary types of consumption, including electricity, water, and natural gas, over an extended period of 2 years. Furthermore, it provides 11 measurement characteristics for electricity. The data in AMPds2 has been pre-cleaned to ensure consistent and comparable accuracy results among researchers and machine learning algorithms.

The REFIT Electrical Load Measurements [36] dataset is another one that includes cleaned electrical consumption data in Watts for 20 households at both aggregate and appliance levels, timestamped and sampled at 8-second intervals. It is designed to support research into energy conservation and advanced energy services, ranging from non-intrusive appliance load monitoring, demand response measures, tailored energy and retrofit advice, appliance usage analysis, consumption and time-use statistics, and smart home/building automation. Finally, the UK Domestic Appliance-Level Electricity data set (UK-DALE) [37] was released, containing data from four households.

The available NILM datasets have varying sample rates ranging from 1 Hz to 100 kHz and cover individual appliances as well as residential complexes. While data collected from individual appliances can be valuable for modeling and training the NILM system, its performance may not be optimal when tested on the entire residential building. Conversely, relying solely on whole-household datasets may not be appropriate for training the algorithms, especially when individual appliance data is unavailable.

Furthermore, certain databases provide primary current and voltage signals, whereas others provide calculated electrical parameters such as active power, reactive power, apparent

power, and power factor. However, in order to create an effective NILM system, it is crucial to obtain unprocessed electrical signals in order to extract fundamental and harmonic characteristics.

V. EXPERIMENTAL RESULTS

A. Evaluation Metrics

To evaluate the performance of algorithms in recognizing appliance switching ON or OFF, the classification metrics presented in Eqs. 5-8 were used. The metrics are calculated using True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP represented the number of times a device was correctly recognized as ON, whereas TN represented the number of correctly identified OFF occurrences. FP highlighted instances when ON states were recorded despite the appliance not consuming power. On the other hand, FN displayed the number of OFF occurrences that were incorrectly recognized.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

$$F_1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

The recall metric in Eq. 7 measures the ratio of correctly identified positive instances (TP) to the total number of positive instances in the dataset. On the other hand, the precision metric in Eq. 6 represents the ratio of correctly identified positive instances (TP) to the total number of instances identified as positive by the algorithm. The F_1 score is a weighted mean of precision and recall, which is used to determine the accuracy of the algorithm in identifying appliance states. Higher F_1 scores indicate better algorithm performance in recognizing appliance state transitions.

The mean absolute error (MAE) and proportion of energy correctly allocated (PECA) metrics in Eqs. 9 and 10, respectively, are non-event-based metrics used to evaluate the accuracy of load disaggregation systems in calculating and assigning electricity usage. MAE measures the average absolute difference between the estimated and actual energy usage, while PECA evaluates the percentage of energy correctly allocated to individual appliances.

$$\text{MAE} = 1/T * \sum_{t=1}^T |\hat{y}_t^i - y_t^i| \quad (9)$$

$$\text{PECA} = 1 - \frac{\sum_{t=1}^T \sum_{i=1}^N |\hat{y}_t^i - y_t^i|}{2 \sum_{t=1}^T \bar{y}_t} \quad (10)$$

In the preceding equations, \hat{y}_t^i and y_t^i are the estimated and ground-truth power of the i^{th} device at time step t , respectively. Furthermore, \bar{y}_t is the total power at time t [24].

TABLE I: Summary of NILM public datasets

Dataset	Sampling Rate Time Resolution	Duration	Type	Number of Households	Number of Appliances	Country	URL
REDD	16.5 KHz	19 days	Residential	6	6	US	https://tokhub.github.io/dbecd/links/redd.html
BLUED	12 KHz	1 week	Residential	1	50	US	https://tokhub.github.io/dbecd/links/Blued.html
UK-DALE	16 KHz	2 years	Residential	5	5	UK	https://jack-kelly.com/data/
REFIT	8 seconds	2 years	Residential	20	9	UK	https://pureportal.strath.ac.uk/en/datasets/refit-electrical-load-measurements
AMPds2	1 min	2 years	Residential	1	21	Canada	http://dx.doi.org/10.7910/DVN/FIE0S4

B. Performance Evaluation and Comparison Study

In this part, the disaggregation findings for the NILM approaches using the REDD, UK-DALE, and REFIT datasets are shown. The performance indicators obtained by executing the tests on these datasets shows that F_1 produced the best results for refrigerators, air conditioners, freezers, televisions, and washing machines across the three datasets, with values greater than 0.70. Toasters and electronics, on the other hand, have lower F_1 scores of roughly 0.25 owing to misclassification caused by the non-uniform pattern of these items.

The accuracy metrics of the findings were compared for published methods: on-line NILM [27], NILM-TK [14], an FHMM implementation; Neural-NILM [7], a DNN adaption for energy estimation. The Neural-NILM used three DNN architectures: i) long short-term memory, ii) de-noising auto-encoders, and iii) rectangles. Rectangle networks, in particular, regress the start-time, end-time, and average power of appliance activation.

In experiments using UK-DALE dataset, a comparison of on-line NILM, NILM-TK and Neural-NILM methods for five appliances (fridge, washing machine, dishwasher, microwave, and kettle) is done. The microwave gives the lowest marks for all three methods. MAE and an F_1 score of roughly 195 Watts and 0.01, respectively, are reported by NILM-TK. The best MAE of 6 Watts and an F_1 score of 0.21 is shown by the neural-NILM. With an F_1 score of about 0.35, the on-line NILM method outperforms the other two. The MAE and F_1 scores reported by NILM-TK are roughly 67 watts and 0.55, respectively. The Neural-NILM has an MAE of 18 Watts and an F_1 score of 0.82. In terms of energy estimate, the Neural-NILM outperformed the suggested technique, particularly for complicated equipment like dishwashers and washing machines. Nonetheless, the time and computational resources required to train the neural network and generate the models need a large amount of appliance-level data. The on-line NILM technique, on the other hand, may generate appliance models using aggregate data without the necessity for appliance-level sub-metered data.

VI. CONCLUSION

The three-phase distribution network often feeds the final residential customers through single phase cables. It is important to make the loads balanced. Demand side management and different optimisation techniques would help on this. However, understanding the contribution of different

appliances to energy consumption is beneficial. NILM is demonstrated to be a good approach to this end. The accuracy of NILM depends on the method applied. This paper reviewed some deep learning-based methods which outperform other existing NILM algorithms. The paper compared the results of applying these advanced methods to provide a basis for future implementation. These datasets have public access and are widely used in NILM literature. Several performance criteria are formulated to analyze the performance of the methods.

REFERENCES

- [1] A. Garulli, S. Paoletti, and A. Vicino, "Models and techniques for electric load forecasting in the presence of demand response," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 3, pp. 1087–1097, 2014.
- [2] A. Estebarsari and R. Rajabi, "Single residential load forecasting using deep learning and image encoding techniques," *Electronics*, vol. 9, no. 1, p. 68, 2020.
- [3] R. Rajabi and A. Estebarsari, "Deep learning based forecasting of individual residential loads using recurrence plots," in *2019 IEEE Milan PowerTech*. IEEE, 2019, pp. 1–5.
- [4] N. Rezaei, R. Rajabi, and A. Estebarsari, "Electricity price forecasting model based on gated recurrent units," in *2022 IEEE International Conference on Environment and Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, 2022, pp. 1–5.
- [5] T. Wang, B. Shen, C. H. Springer, and J. Hou, "What prevents us from taking low-carbon actions? a comprehensive review of influencing factors affecting low-carbon behaviors," *Energy Research & Social Science*, vol. 71, p. 101844, 2021.
- [6] I. Ishak, N. S. Othman, and N. H. Harun, "Forecasting electricity consumption of malaysia's residential sector: Evidence from an exponential smoothing model," *F1000Research*, vol. 11, no. 54, p. 54, 2022.
- [7] D. Shi, R. Li, R. Shi, and F. Li, "Analysis of the relationship between load profile and weather condition," in *2014 IEEE PES General Meeting—Conference & Exposition*. IEEE, 2014, pp. 1–5.
- [8] M. U. Fahad and N. Arbab, "Factor affecting short term load forecasting," *Journal of Clean Energy Technologies*, vol. 2, no. 4, pp. 305–309, 2014.
- [9] 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)*. IEEE, 2015, pp. 1175–1180.
- [10] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
- [11] —, "An unsupervised training method for non-intrusive appliance load monitoring," *Artificial Intelligence*, vol. 217, pp. 1–19, 2014.
- [12] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial hmms with application to energy disaggregation," in *Artificial intelligence and statistics*. PMLR, 2012, pp. 1472–1482.
- [13] X. Sun, P. B. Luh, K. W. Cheung, W. Guan, L. D. Michel, S. Venkata, and M. T. Miller, "An efficient approach to short-term load forecasting at the distribution level," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2526–2537, 2015.

TABLE II: Comparison of Non-Intrusive Load Monitoring (NILM) Methods

Method	Pros	Cons	Data Requirements	Real-time Performance	Interpretability	Evaluation Performance
Template Matching	Simple, easy to implement	Requires pre-defined templates for each appliance, limited accuracy	None	Fast	High	Limited by pre-defined templates
Steady-State Analysis	Accurate for steady-state loads	Limited accuracy for transient loads, requires detailed knowledge of the power system	Detailed power system data	Fast	Medium	Limited by power system model
HMM and fractional HMM	Can handle transient loads, can identify multiple appliances simultaneously	Requires training data for each appliance, can be computationally expensive	Labeled training data	Slow	Low	Sensitive to initialization and noise
Deep Neural Networks	High accuracy, can handle complex appliance behavior, can identify multiple appliances simultaneously	Requires a large amount of training data, can be computationally expensive, limited interpretability	Labeled training data	Fast	Low	Dependent on quality and quantity of training data
Autoencoders	Can handle variable appliance behavior, can identify multiple appliances simultaneously, can learn without labeled data	Requires a large amount of training data, can be computationally expensive	Unlabeled training data	Fast	Low	Dependent on quality and quantity of training data

- [14] H. Gonçalves, A. Oceau, M. Bergés, and R. Fan, "Unsupervised disaggregation of appliances using aggregated consumption data," in *The 1st KDD workshop on data mining applications in sustainability (SustKDD)*, 2011.
- [15] J. Zheng, C. Xu, Z. Zhang, and X. Li, "Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network," in *2017 51st Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2017, pp. 1–6.
- [16] M. Zhuang, M. Shahidepour, and Z. Li, "An overview of non-intrusive load monitoring: Approaches, business applications, and challenges," in *2018 international conference on power system technology (POWERCON)*. IEEE, 2018, pp. 4291–4299.
- [17] 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)*. IEEE, 2015, pp. 63–67.
- [18] J. Kelly and W. Knottenbelt, "Neural nilm: Deep neural networks applied to energy disaggregation," in *Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments*, 2015, pp. 55–64.
- [19] C. Shin, S. Joo, J. Yim, H. Lee, T. Moon, and W. Rhee, "Subtask gated networks for non-intrusive load monitoring," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 1150–1157.
- [20] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for non-intrusive load monitoring," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.
- [21] E. Busseti, I. Osband, and S. Wong, "Deep learning for time series modeling," *Technical report, Stanford University*, pp. 1–5, 2012.
- [22] S. Makonin, F. Popowich, I. V. Bajić, B. Gill, and L. Bartram, "Exploiting hmm sparsity to perform online real-time nonintrusive load monitoring," *IEEE Transactions on smart grid*, vol. 7, no. 6, pp. 2575–2585, 2015.
- [23] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proceedings of the 2011 SIAM international conference on data mining*. SIAM, 2011, pp. 747–758.
- [24] J. Kelly and W. Knottenbelt, "Neural nilm: Deep neural networks applied to energy disaggregation," in *Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments*, 2015, pp. 55–64.
- [25] P. P. M. do Nascimento, "Applications of deep learning techniques on nilm," *Diss. Universidade Federal do Rio de Janeiro*, 2016.
- [26] S. Henriot, U. Şimşekli, B. Fuentes, and G. Richard, "A generative model for non-intrusive load monitoring in commercial buildings," *Energy and Buildings*, vol. 177, pp. 268–278, 2018.
- [27] M. A. Mengistu, A. A. Girmay, C. Camarda, A. Acquaviva, and E. Patti, "A cloud-based on-line disaggregation algorithm for home appliance loads," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3430–3439, 2019.
- [28] L. Massidda, M. Marrocu, and S. Manca, "Non-intrusive load disaggregation by convolutional neural network and multilabel classification," *Applied Sciences*, vol. 10, no. 4, 2020.
- [29] B. Zhao, L. Stankovic, and V. Stankovic, "On a training-less solution for non-intrusive appliance load monitoring using graph signal processing," *IEEE Access*, vol. 4, pp. 1784–1799, 2016.
- [30] T. Kriechbaumer, D. Jorde, and H.-A. Jacobsen, "Waveform signal entropy and compression study of whole-building energy datasets," in *Proceedings of the Tenth ACM International Conference on Future Energy Systems*, 2019, pp. 58–67.
- [31] L. Wen, K. Zhou, and S. Yang, "Load demand forecasting of residential buildings using a deep learning model," *Electric Power Systems Research*, vol. 179, p. 106073, 2020.
- [32] J. Z. Kolter and M. J. Johnson, "Redd: A public data set for energy disaggregation research," in *Workshop on data mining applications in sustainability (SIGKDD)*, San Diego, CA, vol. 25, no. Citeseer, 2011, pp. 59–62.
- [33] K. Anderson, A. Oceau, D. Benitez, D. Carlson, A. Rowe, and M. Berges, "Blued: A fully labeled public dataset for event-based non-intrusive load monitoring research," in *Proceedings of the 2nd KDD workshop on data mining applications in sustainability (SustKDD)*, vol. 7. ACM New York, 2012, pp. 1–5.
- [34] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajić, "Amps: A public dataset for load disaggregation and eco-feedback research," in *2013 IEEE Electrical Power & Energy Conference*, 2013, pp. 1–6.
- [35] S. Makonin, "Amps2: The almanac of minutely power dataset (version 2)," 2021. [Online]. Available: <https://dx.doi.org/10.21227/c2rq-bx53>
- [36] D. Murray, L. Stankovic, and V. Stankovic, "An electrical load measurements dataset of united kingdom households from a two-year longitudinal study," *Scientific data*, vol. 4, no. 1, pp. 1–12, 2017.
- [37] J. Kelly and W. Knottenbelt, "The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes," *Scientific data*, vol. 2, no. 1, pp. 1–14, 2015.