Smart-Building Applications

Deep Learning-Based, Real-Time Load Monitoring

Çimen, Halil; Garcia, Emilio Jose Palacios; Kolbæk, Morten; Cetinkaya, Nurettin; Vasquez, Juan C.; Guerrero, Josep M.

Smart-Building Applications:

Deep Learning-Based, Real-Time Load Monitoring

Halil Çimen, *(non-member)*,
Emilio José Palacios-García, *Member, IEEE*, *Member of IEEE-IES*
Morten Kolbæk, *Member, IEEE*,
Nurettin Çetinkaya, *(non-member)*,
Juan C. Vasquez, *Senior Member, IEEE*, *Member of IEEE-IES*
Josep M. Guerrero, *Fellow, IEEE*, *Member of IEEE-IES*

1. INTRODUCTION

Google's Director of Research, Peter Norvig said that "We don't have better algorithms than anyone else, we just have more data". This inspiring statement shows that having more data is directly related to better decision making and having the foresight about the future. With the development of the Internet of Things (IoT) technology, it is now much easier to gather data. Technological tools such as social media websites, smartphones, and security cameras can be considered as "data generators". When the focus is shifted to the energy field, these generators are "Smart Meters".

Smart meter technology incorporates many intelligent functions and offers great opportunities for utility operators, prosumers, and consumers. Although smart meters are referred to as 'smart', they might not be intelligent enough depending on the final purpose. Meter data generally provide more benefits for the utility side than for the consumer side. However, with the smart meter data, customers can be offered great opportunities, where they may be able to make more conscious decisions. Previous studies have reported that if instantaneous energy consumption data are given to the consumers as feedback, approximately 20% of energy savings can be achieved per household [1]. To achieve this target, more detailed data on the electricity consumed by each appliance are needed. Smart meters cannot meet this

need since they can only read the total electricity consumption. To overcome this issue, Appliance Load Monitoring (ALM) is frequently applied. ALM is used to monitor individual appliances in households by using sensors. Non-Intrusive Load Monitoring (NILM) is an ALM technique that analyzes the total household electricity consumption measured by the main meter and obtains appliance-level information by using various signal processing or pattern recognition techniques. Assuming that, there are at least 20 appliances in each household, it is clear that robust algorithms are needed to solve this problem.

Nowadays, academia and industry show great interest in learning-based data analysis methods [2, 3]. Deep Learning (DL) is the most prominent and explosively growing artificial intelligence technique. Particularly, it has been gaining popularity in many different areas, such as image classification, speech recognition, and health management, due to its superior performance over other traditional methods [4, 5]. Considering that there are millions of smart meters installed, and these meters produce data every minute, it can easily be seen that DL is one of the most suitable methods to solve the NILM problem.

This article introduces the NILM method, which can contribute to energy management and savings in residential, industrial, and naval uses. Up-to-date data-driven NILM solutions and advantages of DL-based analysis are explained in detail. Also, a multi-label DL approach, which can save training time and reduce the need for model storage, is presented and tested in real-time. Considering that the studies in the literature are carried out offline, the online analysis capacity of recent DL models has been tested in a laboratory environment. In this way, the accuracy difference between offline and online implementations has been revealed.

2. NON-INTRUSIVE LOAD MONITORING

Load monitoring is an important part of energy management in households, industry, and naval vessels [6]. There are two types of load monitoring methods: Intrusive Load Monitoring

(ILM) and NILM. ILM is an advanced, systematic, and high accuracy load monitoring technique, which is often applied for smart homes. One sensor, which can be a potentially smart plug, per appliance is used to remotely monitor and control the appliances. However, the main disadvantages are the need for a comprehensive installation, communication infrastructure, maintenance, and updating. All these features make the ILM a high-cost system, besides the data privacy breach. Users can be conservative in sharing the data, especially by installing sensors in the household. To eliminate these drawbacks, NILM is proposed as a cost-effective alternative solution [7]. In the NILM technique, also referred to as energy disaggregation, rather than using an individual sensor for each appliance, the energy consumed by the entire household, referred to as aggregated data, is monitored by using only one sensor, which can potentially be the main smart meter. Since no extra sensors are placed inside the household, it is called Non-Intrusive. Aggregated data is analyzed by various signal processing or pattern recognition methods to obtain the appliance-level disaggregated data. An example of data disaggregation is shown in Fig. 1.
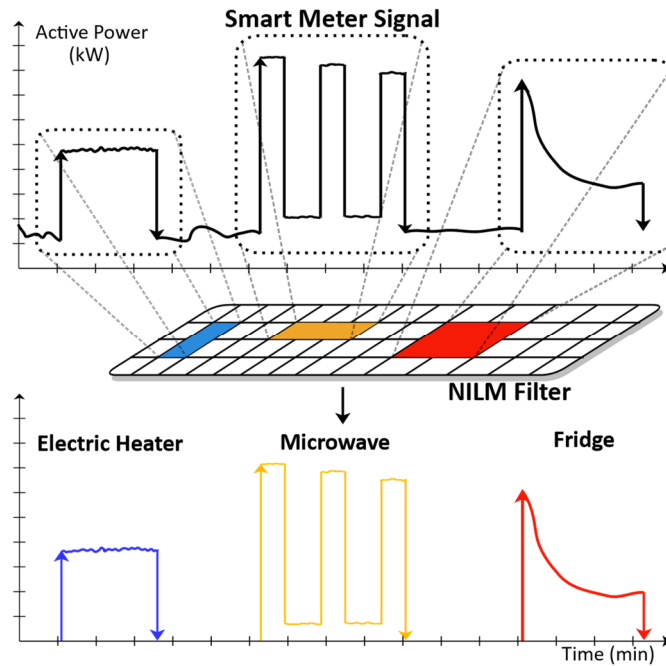

Fig. 1. An example of disaggregated data of residential appliances

With a successful NILM analysis, real-time and statistical information about the appliances, their daily usage rate, and users' daily consumption behavior can be easily obtained. Using the obtained data, many different actions such as home energy management, appliance-based load forecasting, and demand response can be taken by the utility and consumption side. The general NILM structure and some of its benefits are depicted in Fig. 2.
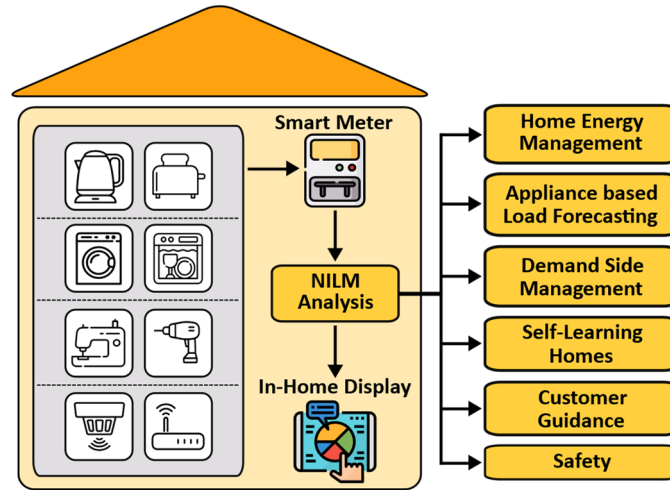
Fig. 2. General NILM structure of a household case

The NILM is of great interest in the private sector and academia. Today, there are more than 40 companies offering energy disaggregation products. Each company provides solutions with its hardware/software and they do not share detailed information about their methods. Academic studies began in 1992 with a study by George W. Hart [7] and although many years have passed since the first study, the desired level of success has not been achieved yet. For this reason, it attracts great interest in academia. In recent years, studies have gained momentum with the sharing of public datasets and the increase of data obtained from smart meters.

3. METHODOLOGY

NILM can be considered as a signal separation, which is the process of recovering source signals by separating a mixed-signal measured from a single sensor. For the NILM problem,

the mixed signal is aggregated data, and the source signals are power consumption of each appliance. The NILM problem can be formulated in simple form as follows:

$$p_{agg}(t) = \sum_{n \in N} s_n(t).p_n(t) + e(t) \qquad (1)$$

where, $p_{agg}(t)$ is the aggregated active power for sample t, s_n and p_n are the status (on-1/off-0) and the instantaneous active power consumption of the appliance n for sample t, respectively. N is the number of appliances and e is the measurement error or noise. Although (1) is a simple equation, the fact that there are many appliances with different working principles makes it difficult. Each appliance has its load pattern, which is called "Appliance Signature". To systematically address the NILM problem, appliances need to be classified. Hart [7] categorized appliances and divided them into 3 types according to their signatures. The types of appliances and their general signatures are shown in Fig. 3. Type-I appliances have only on/off states (e.g., toasters, kettles). On the contrary, Type-II appliances are those who have multi-states (e.g., washing machine, tumble dryer). Type-III appliances consume variable power and do not have a specific state or periodic operation.
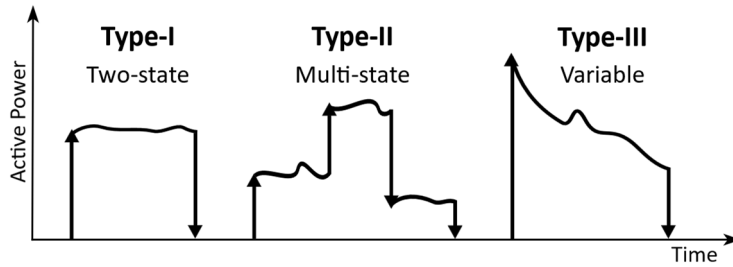
Fig. 3. Types of appliances

The most important factor directly affecting NILM success is the characteristics of the data. Active power is the most commonly used data type. However, analyzing the appliances consuming similar active power or activated simultaneously is a non-trivial task. Therefore, the use of additional features such as reactive power can facilitate the analysis. The second important characteristic is the resolution which can be divided into two: Low (1Hz and lower)

and High resolution (higher than 1Hz). There is a trade-off between them. High-resolution data provide more detailed information, but at high hardware cost. Low-resolution data provide limited information, but it is cost-effective. It is more realistic to perform NILM analysis using low-resolution active power data since it's already available from the smart meters. Detailed information on NILM analysis using high-resolution data can be found in [8, 9].

The ultimate goal of NILM studies can be classified under two main titles: Load identification and Energy disaggregation. Load identification is the instant detection and recognition of the appliances that are turned on or off. Energy disaggregation is the process of estimating the energy consumption of the appliances individually. A high accuracy energy disaggregation might also provide information about the load identification.

4. DATA-DRIVEN LOAD MONITORING STUDIES

Optimization or pattern recognition-based approaches are frequently preferred in the field of NILM. Given the optimization-based approach, a minimization problem can be written by re-formulating (1) as follows:

$$\tilde{S}(t) = \arg \min_{S} \left| p_{agg}(t) - \sum_{n \in N} s_n(t).\bar{p}_n \right| \tag{2}$$

A status vector, $\tilde{S} = \{s_1, s_2, ..., s_N\}$, is created that estimates whether appliances are operating or not for sample t. To minimize the difference between the aggregated power and sum of appliance-level consumption, the best possible appliance combination is tried to be found by using different status vectors, which are obtained combinatorially. The average energy consumption, $\bar{p}_n$, can be obtained by analyzing the sub-metering data or using the appliance manual. However, this method is not practical. Because either the power consumption of all appliances must be known in advance, which might not be possible in practice, or the power consumption of the appliances that will not be analyzed should be defined as "base-load" and it

should be estimated by a prediction method or a statistical approach. Secondly, as the number of appliances increases, the length of the vector increases, and the solution space grows exponentially. Besides, appliances consuming similar power cannot be distinguished [10, 11]. Therefore, pattern recognition-based approaches such as Hidden Markov Model (HMM) and Machine Learning (ML) are preferred. A traditional HMM [12] and its variants [13-15] are implemented to improve the analysis accuracy. Despite achieving reasonable results, the biggest disadvantage is that the complexity increases exponentially as the number of appliances increases. Various ML algorithms such as support vector machine, k-nearest neighbor, and decision trees are performed in the NILM field due to its robust analysis capability [16, 17]. The performance of ML methods depends on manually extracted features. However, it is often not possible to predict which features are more effective. Especially in complex systems, where feature extraction means a long time and huge human effort. DL models, if provided enough data, achieves results similar or even (often) better than what would have been achieved when hand-engineered features are used. Since DL model training scales well with the amount of data, DL models can usually utilize much more data than traditional non-DL models. This enables the models to utilize these large quantities of data and ultimately achieve state of the art performance [18, 19]. Illustrative comparison of ML and DL for the NILM application is shown in Fig. 4.
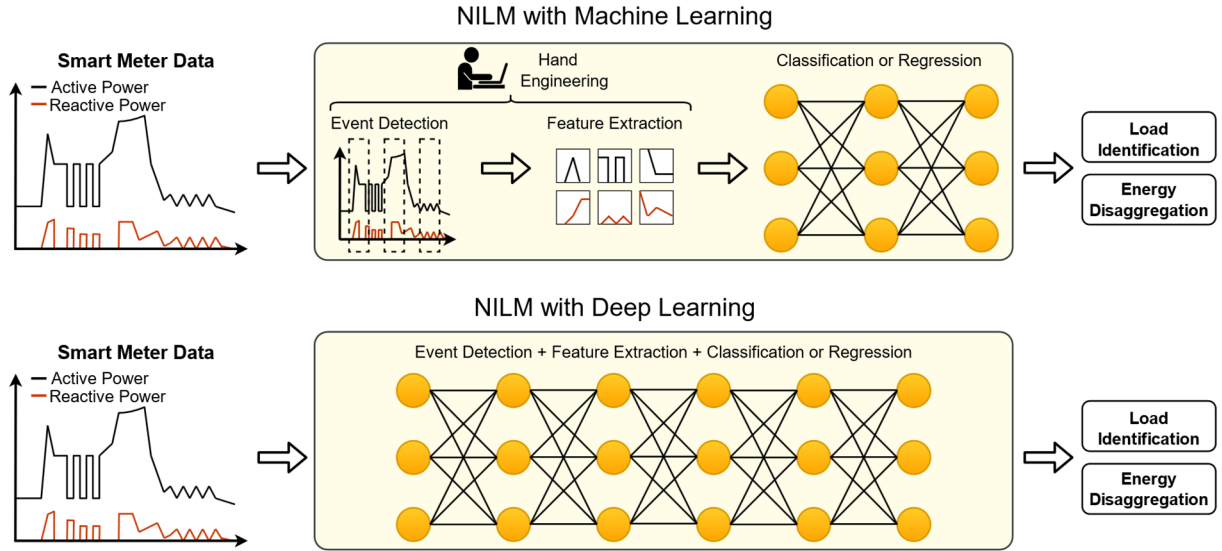
Fig. 4. Illustrative comparison of ML and DL

DL can be adapted to NILM since they can easily learn from the smart meter data. When the literature is investigated, it can be seen that 3 different DL models are frequently used. These are Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Auto Encoder (AE).

CNN stands out especially for its high performance in image classification [18]. When analyzing a large image, it uses a large number of small convolution kernels to produce simple concepts. By combining them, more complex concepts are obtained and the hierarchical features representing data are extracted. In the literature, two different approaches are used for CNN-based NILM analysis, sequence-to-point (S2P) [20], and sequence-to-sequence (S2S) [21]. Both of these methods use the same input data. However, it is called S2S if a sequence is estimated at the output, or S2P if a single point is estimated. Another CNN-based model, Wavenet, which is originally developed for raw audio generation, is implemented in [22]. The advantage of this model is that it can analyze longer input sequences with less parameters. It can be suitable for long-term operating appliances such as a dishwasher. In [23, 24], energy disaggregation is performed by using AlexNet and VGG-16 models, which are originally developed for image classification. These models are adapted for NILM with some

modifications and promising results are obtained. While all of these methods have advantages over each other, CNN is not capable of detecting time-dependent changes since it cannot make a connection between the past and the future data.

RNN can analyze sequence models or time-series. For the image processing field, all inputs and outputs are independent of each other. But in the case of time-series, the future data is mostly linked to the past data. The reason it is called recurrent is that it performs the same task for each element of an array based on the previous outputs. The RNN can evaluate the current input based on past data thanks to its memory. However, long sequence analysis weakens the learning capacity of RNN. Two RNN-based methods, Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) have been developed to mitigate this problem. Although LSTM and GRU are two similar models, the number of total training parameters of GRU is less since it does not have a separate memory cell. So it can be trained faster than LSTM. If a model can be trained faster, experiments can be conducted faster and ultimately the chance of finding good hyperparameters increases, which usually leads to better performance. In [25, 26], an LSTM model is implemented and promising results are obtained. An energy disaggregation model combination of CNN and GRU is proposed in [27]. The authors aimed to improve the energy consumption estimation results using GRU's time analysis capability.

The third method, AE consists of an encoder and a decoder. The encoder expresses the input data as a concentrated vector representation, which contains the distinctive features of the input. The decoder reconstructs this vector representation to the desired format. Considering the NILM, the aggregated data can be considered as noisy input (the noise here is the energy consumption of appliances other than the target appliance). The energy consumption of the target appliance is the decoded output. In [25], the authors proposed a denoising autoencoder (dAE) in order to filter noises. Although successful results were obtained for Type-I appliances,

they were insufficient for Type-II. A new AE model combined with CNN is proposed in [28]. The obtained results show that AE can be considered in the solution of the NILM problem.

a. Multi-label Convolutional GRU Architecture

When the studies mentioned above are examined, it is seen that each DL method has its advantages and disadvantages against each other and somehow they yield similar results. However, all of these studies are done offline and it is uncertain how these methods will behave in online applications. In this paper, a real-time load identification is performed using a convolutional GRU (C-GRU) model. The model architecture is shown in Fig. 5.

The input data are the active power values read from the smart meter. Since there is a large amount of data (over the months), the input and output should be split using the sliding windows. Assuming that the selected window size is w, the input data is split as *(t:t+w-1)* from the starting of sample t by shifting with a certain step for each time. When sliding windows are set, they are evaluated by 1D convolutional layers to obtain high-level features, which are given as an input to the GRU. Afterward, GRU layers evaluate the data as dependent on historical data and identify the actively operating appliances. In order to improve the performance, they can be used with bidirectional layers, which make it possible to analyze the time-series forward and backward. Ultimately a larger model is obtained with access to more context. The designed model consists of one input, one convolutional, two bidirectional GRUs, and two fully connected layers. For the convolutional layer, filter size and number of filters are selected 3 and 64, respectively. The GRU layers have 256 nodes, while the first fully connected layer has 128 nodes. Hyperbolic tangent is used in all hidden layers as the activation function.
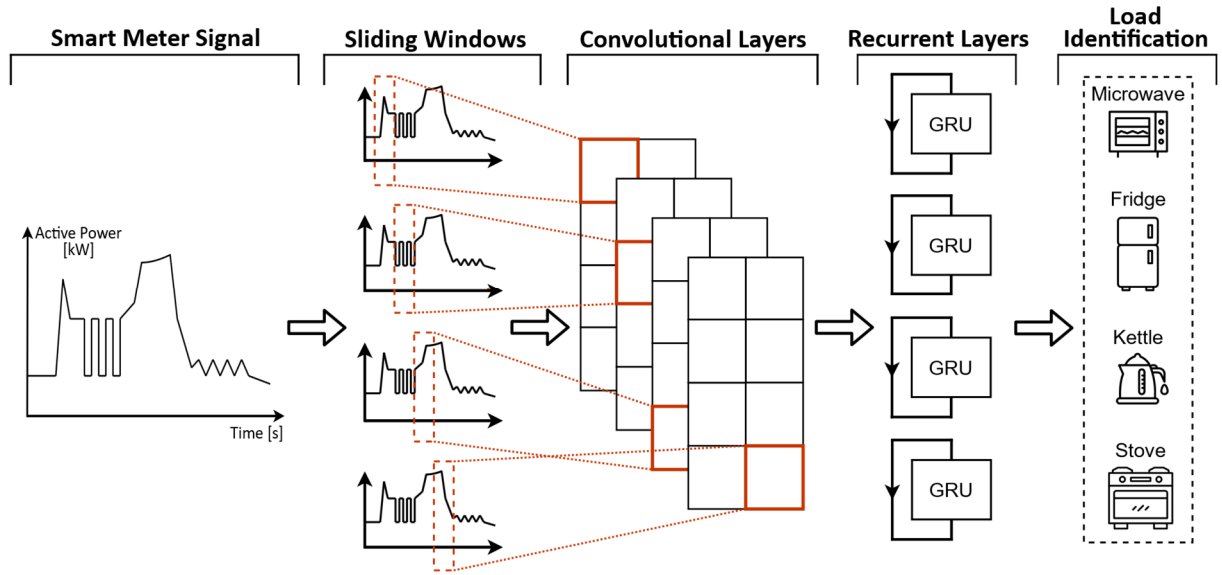
Fig. 5. Model architecture for real-time load identification

When studies in the literature are examined, it is seen that an individual DL model is trained for each appliance. Considering that DL models are trained with a large amount of data, it is clear that the training period may be very long. In this paper, multi-label appliance classification, which is capable of analyzing multiple appliances with a single DL model, has been proposed to reduce training time. Considering that there are more than 20 appliances in a household, it is obvious that this approach will significantly save time. For multi-label classification, the number of nodes and activation function of the output layer are selected as the number of appliances and sigmoid, respectively. Binary cross-entropy and Adam are used as loss function and optimizer, respectively. This architecture is designed for supervised learning in which input is aggregated data read from the smart meter and output is the status (on/off) information of target appliances which we want to analyze. The on/off status is determined according to a predetermined threshold. If the energy consumption of an appliance is higher than the threshold, it's assumed that the appliance is on.

5. REAL-TIME EVALUATION OF DIFFERENT DL MODELS

The studies in the literature are conducted offline using publicly available datasets. During offline analysis, NILM is performed more easily, since the whole energy consumption period (past, current, and future) is available. But in the online analysis, the appliances need to be detected instantly with only current and past data. Therefore, how big the gap will be between the accuracy rates of online and offline applications has not been addressed yet.

The most important factor affecting the real-time analysis is undoubtedly the selected window size w. In the literature, it is recommended to determine the window size according to the operation cycle of the analyzed appliance [25]. For example, the window size should be selected relatively long for appliances with long operating time such as dishwashers to analyze its entire cycle. However, this is not possible during the online analysis. Unlike the offline, the online analysis should be performed without waiting for the appliance to complete its cycle. For this reason, an analysis interval is defined as shown in Fig. 6.

As shown in Fig. 6, a certain number of samples is read from the smart meter depending on the window size and it is evaluated using the DL model for each iteration. The next iteration should be analyzed after a certain interval, which should be chosen as short as possible to instantly detect the appliance operation. In this paper, the iterations are progressed with 60-second intervals. Another important parameter, window size, should be chosen wisely. Since the proposed model has a multi-label classification structure, only one window size should be selected for both appliances with long and short operating time. Considering that short-term appliances such as microwaves and toasters operate for an average of 5-10 minutes, and long-term appliances such as dishwashers operate for an average of 1 hour, an average window size of 256 samples (approx. 20 minutes) that can be suitable for both types of appliances is determined for analysis.
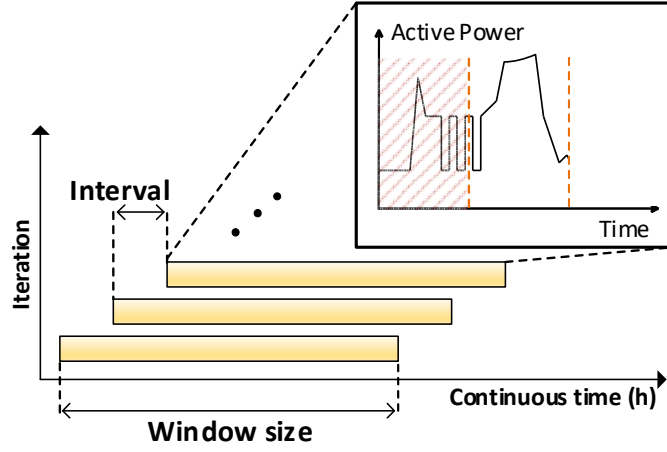
Fig. 6. Online analysis process

Domestic appliances are basically divided into two groups controllable and non-controllable. Analysis of controllable appliances, which can be classified as thermostatically controlled and deferrable loads, is more important to support both energy-saving and demand-side management applications. In this paper, two thermostatically controlled loads, which are fridge (FR) and heater (HE), and seven deferrable loads, which are microwave (MW), kettle (KT), coffeemaker (CM), dishwasher (DW), tumble dryer (TD), washing machine (WM) and toaster (TO), are taken into consideration for real-time identification. Besides, appliances such as WM, DW, TD (around 1.8kW), and HE, MW, KT (around 1kW) have similar power consumptions or peak points. Thus, it will make possible to observe the effect of the presence of appliances in the same range on NILM analysis. Signatures of the target appliances are shown in Fig. 7.
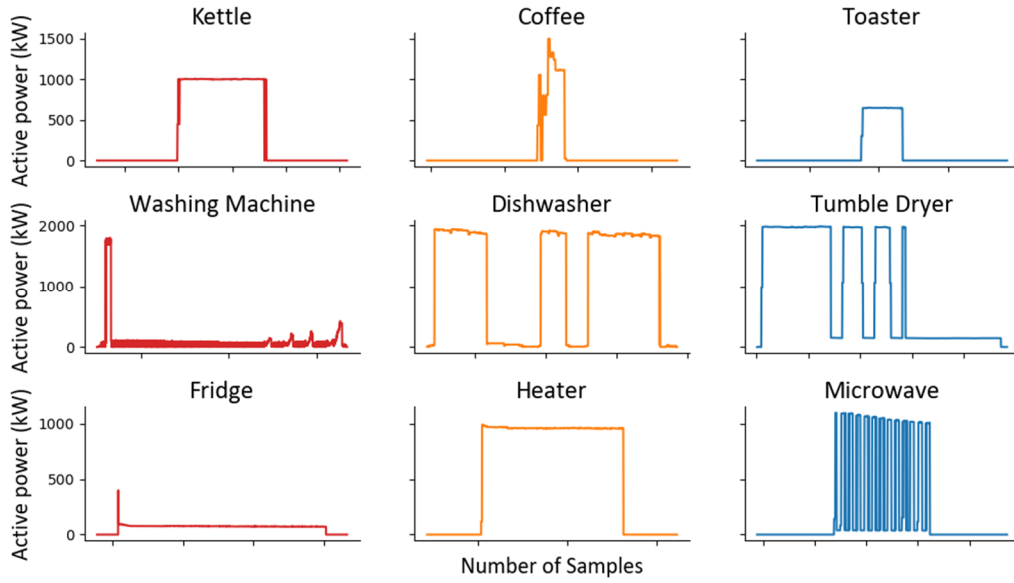
Fig. 7. Signatures of the target appliances

Appliance-level data and aggregated data are obtained with the help of the prosumer meter and smart plugs. If a successful analysis is desired, it should be ensured that the dataset contains good quality observations and is large enough to extract the necessary features. However, real-world data may not always be sufficient. Therefore, the data should be examined first, and missing values should be corrected with filling forward, which fills the gaps based on the corresponding value in the previous sample, for both training and testing data. However, if the training data is modified to include missing data, the model can also handle missing points that will occur during online analysis. Secondly, the usage frequency of the target appliance should be analyzed. For example, if a household's aggregated data covers 1 month, and the target appliance was used only once during that period, sufficient information cannot be extracted [29]. To mitigate this problem, synthetic data generation, which is a method to augment the data by using the existing dataset, is used. For an image classification problem, original images are modified using different techniques such as rotation, scaling, and cropping the picture. The modified images are added to the dataset as new data. In this paper, signatures of different appliances are randomly combined to create a new synthetic consumption profile. In this way,

the number of load patterns that belongs to the target appliance is increased in the dataset. In the final step, the sampling frequency of target appliances and aggregated power consumption should be adjusted for a proper supervised learning. The frequency of the data read from the sensors is not regular and changes between 5-10 seconds. First, an upsampling with filling forward was applied to convert these data to 1 Hz so that all data are simultaneous. Then the data were resampled to 5 seconds since the data with 1-second resolution require extra hardware to store and extra time for training. The data are standardized by subtracting the mean and dividing it by the standard deviation to increase the learning capacity of the model.
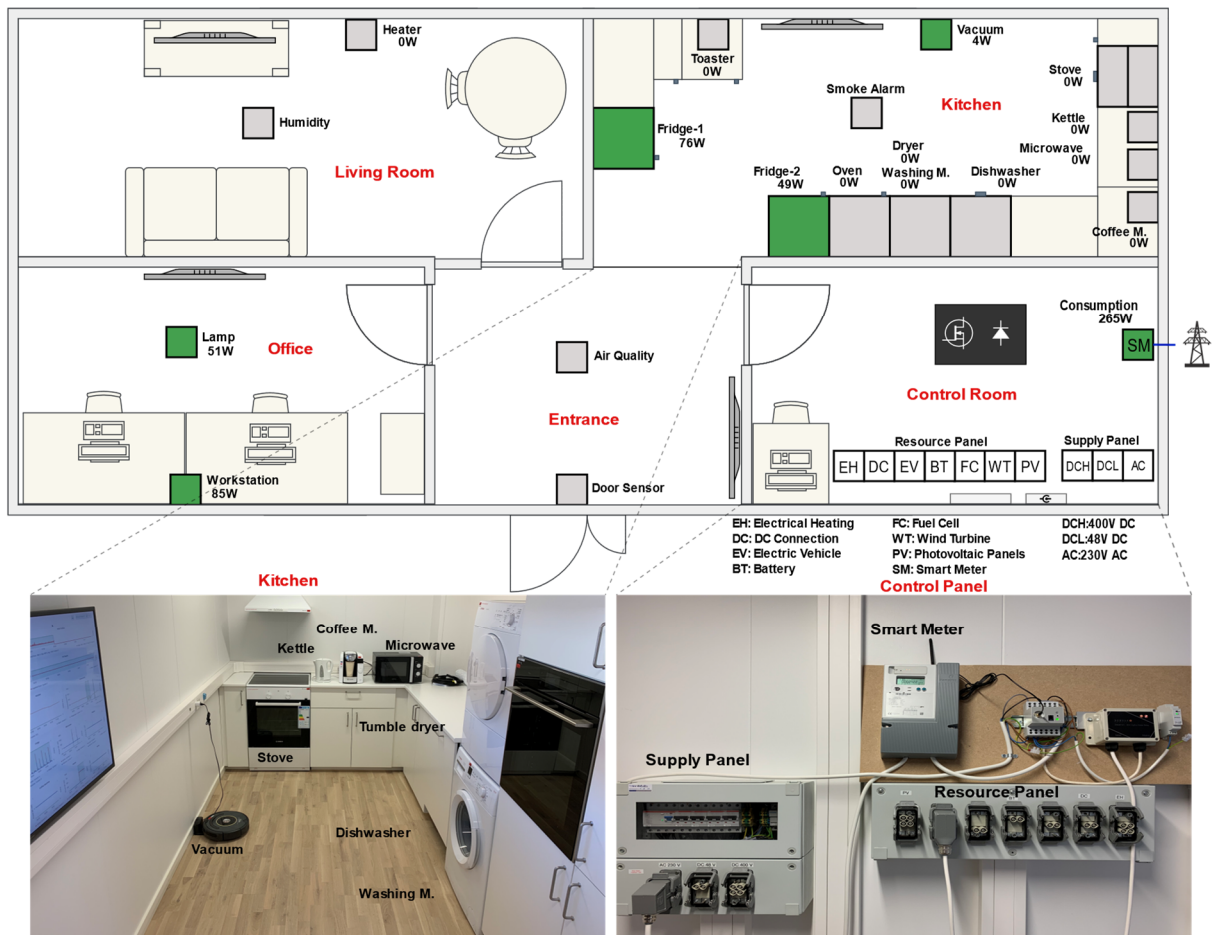

Fig. 8. IoT-MGLab general overview

The developed DL model has been tested at the IoT-Microgrid Living Laboratory (IoT-MGLab) at the Department of Energy Technology, Aalborg University. An overview of the laboratory is shown in Fig. 8.

A Dell Workstation with a 6-core Intel Xeon CPU at 3.60 GHz, 32 GB of RAM and a dedicated GPU NVIDIA Quadro P600 running on CentOS was used for the training and initial tests. In addition, the final trained networks were deployed on a Windows 10 laptop with a i5 (2nd Generation) CPU at 2.4 GHz and 6 GB of RAM for the online evaluation. This laptop was connected to the central data collection system of the IoT-MGLab from which it obtained the real-time measurements used in the identification of the appliances. The DL models are implemented in Python using Keras library.

To obtain more realistic results, the experiment is repeated 10 times. The results are averaged and evaluated using 4 different metrics as shown below:

$$
recall = \frac{TP}{TP + FN} \quad , \quad precision = \frac{TP}{TP + FP}
$$
$$
F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad , \quad accuracy = \frac{TP + TN}{TP + TN + FP + FN}
$$

(3)

where, TP (true positive) and TN (true negative) indicate that the model correctly predicts the appliance is on and off, respectively. FP (false positive) and FN (false negative) are outputs where the model incorrectly predicts the appliance is on and off, respectively. Considering the metrics, the accuracy score can be a misleading indicator in cases of unbalanced appliance signatures. For example, a toaster is used only once or twice a day. The DL model will achieve an accuracy of over 99%, even if it predicts that the toaster is off all day. However, precision and recall can give more realistic results because they mostly analyze the periods during which the appliance is on. In the literature, the F-1 score is generally preferred metric because it is interpreted as a weighted average of precision and recall. The F-1 score comparison of online

and offline analyzing results for different types of DL models are shown in Table I. To analyze the problem from a wider perspective, CNN-based S2S [20], dAE, LSTM [25], RNN, and C-GRU models were compared. RNN, LSTM and C-GRU models have the same configuration except for recurrent layers. During each experiment, at least 4 appliances were operated simultaneously with different combinations.

Table I. F-1 score comparison of online and offline analysis results

Appliances and Types		Offline Analysis					Online Analysis				
		CNN	dAE	RNN	LSTM	C-GRU	CNN	dAE	RNN	LSTM	C-GRU
Type I	KT	0.714	0.116	0.694	0.738	**0.822**	0.620	0.000	0.597	0.701	**0.755**
	CM	0.678	0.084	0.508	0.665	**0.732**	0.522	0.000	0.358	0.592	**0.678**
	TO	0.549	0.161	0.351	**0.682**	0.651	0.395	0.000	0.219	0.651	**0.661**
Type II	WM	0.938	0.954	0.893	0.940	**0.962**	0.924	0.897	0.914	**0.952**	0.939
	DW	0.662	0.720	0.695	**0.794**	0.773	0.677	0.638	0.748	**0.755**	0.703
	DR	0.509	0.735	0.681	**0.846**	0.831	0.498	0.716	0.586	0.759	**0.761**
Type III	FR	0.679	0.661	0.675	0.764	**0.777**	0.688	0.653	0.690	**0.733**	0.698
	HE	0.878	0.624	0.899	0.933	**0.935**	0.821	0.426	0.719	**0.868**	0.825
	MW	0.931	0.526	0.942	0.933	**0.943**	0.908	0.392	**0.921**	0.907	0.892

The results can be evaluated from 3 different aspects. Considering the DL models, recurrent-based models outperformed CNN and dAE models. The secret behind this success is the memory capability of recurrent-based models. On the other hand, the CNN model gives better results than the dAE model since it has a deeper structure than dAE. It shows that if CNN-based load identification analysis is desired, deeper CNN models should be designed. If we compare recurrent-based models, the success rate of RNN is lower due to the limited capacity to analyze long sequences. However, LSTM and GRU give comparable results for long sequences. Slightly better results were obtained with the C-GRU model. The second aspect is the appliance types and signatures. Type-I appliances used in this experiment have short operating times around 2-4 minutes. Since the window size is determined around 20 minutes, their consumption may be perceived as small spikes in this window. For this reason, the success of the analysis is between 65-82%. Type-II appliances are long-running and multi-state appliances, which makes their signature distinctive. Analysis success is higher than Type-I as more precise connections can be established between the past, current, and future. The energy

consumption of Type-III appliances is not constant since their set points can vary according to the user's knob setting. Thanks to the generalization capacity of DL models, the analysis success is high despite the use of different set points.

The third aspect is the comparison of online and offline analysis. For almost every appliance, online analysis success was observed to be lower. The most obvious reason for this is that analysis is requested before the operation cycle of the appliances is completed. Therefore, they are not sufficiently detected or wrong appliances were considered active. But as new data are read, the success rate has increased. An average of 5% accuracy loss can be reported between online and offline analysis. Besides, the analysis success of WM and MW are higher than other appliances. The main reason behind this is their distinctive signatures. As seen in Fig. 7, most appliances somehow have a rectangular energy consumption profile. However, since WM and MW have a constantly changing and dynamic load profile, they can be analyzed by the models with higher accuracy.

The effect of window size selection and comparison of the training times of the models can be seen in Fig. 9.
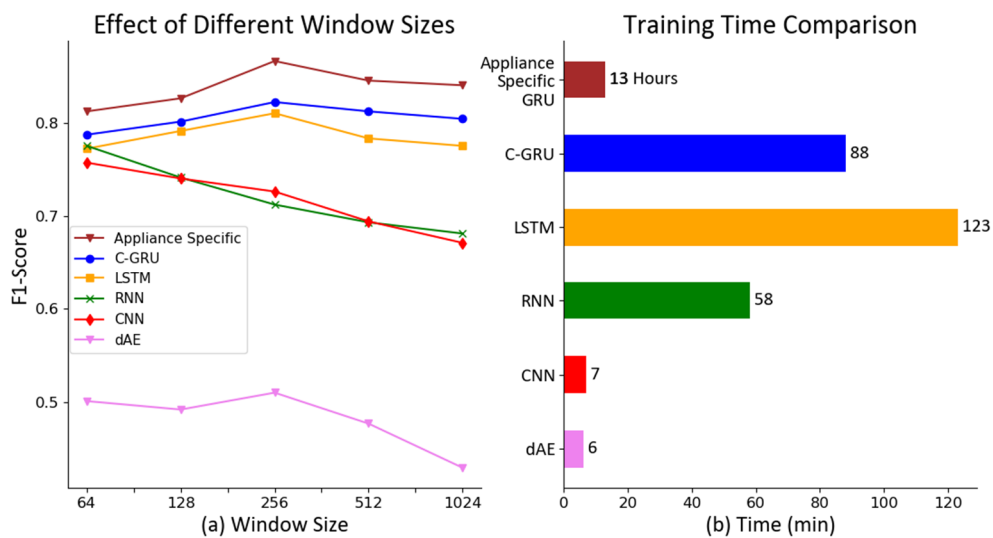

Fig. 9. The effect of window size selection, and comparison of the training times

As can be seen from Fig. 9 (a), using different window sizes affects the F-1 score. Since GRU and LSTM have long-term memory, accuracy increases with increasing the window size. However, analysis success may decrease, as very long window sizes can make it difficult to remember historical data. Since RNN cannot analyze long sequences, its performance decreases rapidly and the model gets worse results than C-GRU and LSTM. The obtained results from CNN and dAE models are not good enough for real-time analysis. More accurate results can be obtained if an individual model is trained for each appliance, which is called appliance-specific model. In this case, nine different models need to be trained for nine different appliances, the total training time of which takes about 13 hours. As seen in Fig. 9 (a), the F-1 score difference between the multi-label C-GRU and appliance-specific approach is very small. However, there is a big difference between training time. Other disadvantages of the appliance-specific model are that each trained model takes up extra space on the hard drive and must be run separately, which requires extra hardware. This can be a significant constraint since NILM algorithms will potentially be deployed at household or building level. This implies the use of embedded edge-computing systems or even existing home or building energy management systems (HEMS, BEMS) with limited computational resources.

Considering the training times of other models, it is seen that CNN and dAE are trained faster since their trainings are done based on matrix multiplication. Because GRU, LSTM and RNN are memory-based models, their training period is longer. Although RNN is trained in a shorter time compare to GRU and LSTM, its analysis success remains insufficient. GRU can be trained faster and slightly better results can be achieved than LSTM.

Also, the same C-GRU model can be used for energy disaggregation. Only the activation function of the output layer should be changed to linear, and the training loss function to mean squared error. The obtained results for MW, DW, and CM for simple aggregated data examples are shown in Fig. 10.
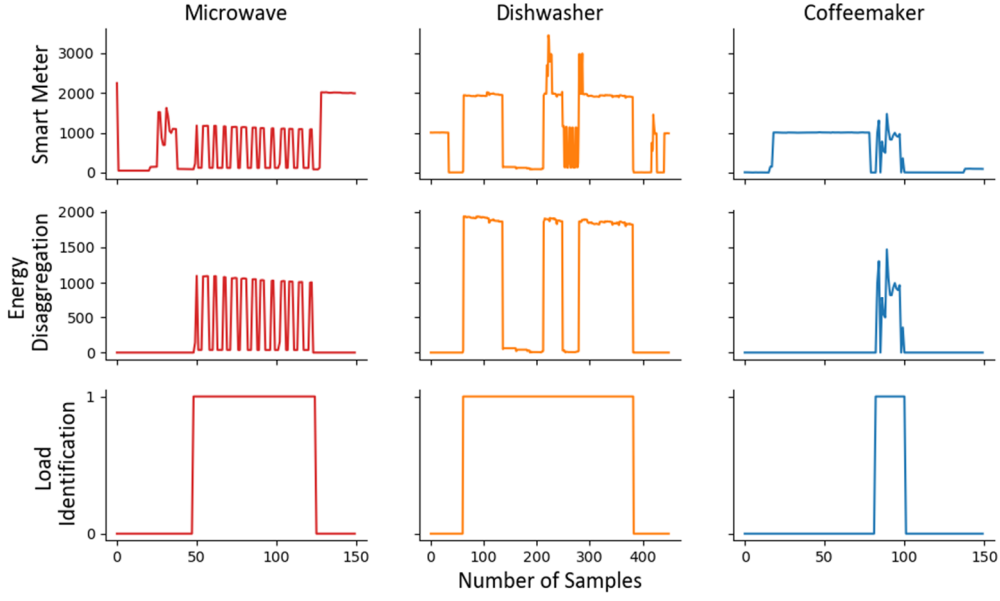
Fig. 10. Results of energy disaggregation and load identification

6. CONCLUSION

In this article, the NILM technique is introduced and applications of the recent DL methods in the NILM field are explained. In addition, a multi-label C-GRU model is proposed, which makes it possible to train and test multiple appliances with a single DL model. In this way, both significant time saving is achieved and the need for data storage can be reduced, which are critical factors for the integration of such algorithms at household or building level. The proposed model is tested in real-time and the results are compared with up-to-date DL models. Recurrent-based models LSTM and GRU outperformed CNN and dAE models. Therefore, it is recommended to compare the new DL models to be developed with recurrent-based models. Regarding to them, C-GRU is trained faster than LSTM and slightly better results are obtained compare to RNN and LSTM.

The majority of appliances used in the experiment somehow have a rectangular energy consumption profile and are similar to each other. However, WM and MW have a more distinctive and dynamic load profile. For this reason, they have been identified with higher accuracy. According to our perception, since DL models analyze consumption profile rather

20

than the state of appliances, appliance types should be redefined in more detail, considering the similarity and difference of the consumption profiles rather than the state transitions of the appliances.

Finally, it has been observed that there is an average 5% difference between the online-offline analysis success of DL models. This difference should be considered for real-time load identification required for demand response applications. In addition, the difference may increase with increasing number of analyzing appliances. This increase can be mitigated using either more robust DL models or post-processing. Post-processing is the approach of refining the results with the help of various optimization algorithms. Accuracy rates can be increased by re-analyzing the outputs of DL models. In the upcoming years, great advances can be made in the energy sector by combining load monitoring algorithms with security and energy management, especially in residential, industrial, and naval uses.

ACKNOWLEDGEMENTS

REFERENCES

[1]   E. Aydin, D. Brounen, and N. Kok, "Information provision and energy consumption: Evidence from a field experiment," *Energy Economics,* vol. 71, pp. 403-410, 2018.
[2]   L. Li, K. Ota, and M. Dong, "Deep Learning for Smart Industry: Efficient Manufacture Inspection System With Fog Computing," *IEEE Transactions on Industrial Informatics,* vol. 14, no. 10, pp. 4665-4673, 2018.
[3]   C. Shi, G. Panoutsos, B. Luo, H. Liu, B. Li, and X. Lin, "Using Multiple-Feature-Spaces-Based Deep Learning for Tool Condition Monitoring in Ultraprecision Manufacturing," *IEEE Transactions on Industrial Electronics,* vol. 66, no. 5, pp. 3794-3803, 2019.

[4]   W. Li, H. Li, Q. Wu, X. Chen, and K. N. Ngan, "Simultaneously Detecting and Counting Dense Vehicles From Drone Images," *IEEE Transactions on Industrial Electronics,* vol. 66, no. 12, pp. 9651-9662, 2019.

[5]   W. Peng, Z. Ye, and N. Chen, "Bayesian Deep-Learning-Based Health Prognostics Toward Prognostics Uncertainty," *IEEE Transactions on Industrial Electronics,* vol. 67, no. 3, pp. 2283-2293, 2020.

[6]   A. Aboulian *et al.*, "NILM dashboard: A power system monitor for electromechanical equipment diagnostics," *IEEE Transactions on Industrial Informatics,* vol. 15, no. 3, pp. 1405-1414, 2018.

[7]   G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE,* vol. 80, no. 12, pp. 1870-1891, 1992.

[8]   M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE transactions on Consumer Electronics,* vol. 57, no. 1, pp. 76-84, 2011.

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

[10]   K. He, D. Jakovetic, B. Zhao, V. Stankovic, L. Stankovic, and S. Cheng, "A generic optimisation-based approach for improving non-intrusive load monitoring," *IEEE Transactions on Smart Grid,* 2019.

[11]   J. Liang, S. K. Ng, G. Kendall, and J. W. Cheng, "Load signature study—Part I: Basic concept, structure, and methodology," *IEEE transactions on power Delivery,* vol. 25, no. 2, pp. 551-560, 2009.

[12]   W. Kong, Z. Y. Dong, J. Ma, D. J. Hill, J. Zhao, and F. Luo, "An Extensible Approach for Non-Intrusive Load Disaggregation With Smart Meter Data," *IEEE Transactions on Smart Grid,* vol. 9, no. 4, pp. 3362-3372, 2018.

[13]   W. Kong, Z. Y. Dong, D. J. Hill, J. Ma, J. Zhao, and F. Luo, "A hierarchical hidden markov model framework for home appliance modeling," *IEEE Transactions on Smart Grid,* vol. 9, no. 4, pp. 3079-3090, 2016.

[14]   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.

[15]   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.

[16]   T. Le, H. Kang, and H. Kim, "Household Appliance Classification Using Lower Odd-Numbered Harmonics and the Bagging Decision Tree," *IEEE Access,* vol. 8, pp. 55937-55952, 2020.

[17]   S. M. Tabatabaei, S. Dick, and W. Xu, "Toward non-intrusive load monitoring via multi-label classification," *IEEE Transactions on Smart Grid,* vol. 8, no. 1, pp. 26-40, 2016.

[18]   A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[19]   Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature,* vol. 521, no. 7553, pp. 436-444, 2015.

[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 *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[21] K. Chen, Q. Wang, Z. He, K. Chen, J. Hu, and J. He, "Convolutional sequence to sequence non-intrusive load monitoring," *The Journal of Engineering,* vol. 2018, no. 17, pp. 1860-1864, 2018.

[22] J. Jiang, Q. Kong, M. Plumbley, and N. Gilbert, "Deep Learning Based Energy Disaggregation and On/Off Detection of Household Appliances," *arXiv preprint arXiv:1908.00941,* 2019.

[23] G. Cui, B. Liu, W. Luan, and Y. Yu, "Estimation of Target Appliance Electricity Consumption Using Background Filtering," *IEEE Transactions on Smart Grid,* 2019.

[24] W. Kong, Z. Y. Dong, B. Wang, J. Zhao, and J. Huang, "A Practical Solution for Non-Intrusive Type II Load Monitoring based on Deep Learning and Post-Processing," *IEEE Transactions on Smart Grid,* 2019.

[25] 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.

[26] L. Mauch and B. Yang, "A new approach for supervised power disaggregation by using a deep recurrent LSTM network," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2015, pp. 63-67: IEEE.

[27] D. Murray, L. Stankovic, V. Stankovic, S. Lulic, and S. Sladojevic, "Transferability of Neural Network approaches for low-rate energy disaggregation," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 8330-8334: IEEE.

[28] T. Sirojan, B. T. Phung, and E. Ambikairajah, "Deep Neural Network Based Energy Disaggregation," in *2018 IEEE International Conference on Smart Energy Grid Engineering (SEGE)*, 2018, pp. 73-77: IEEE.

[29] J. Zhang, X. Chen, W. W. Ng, C. S. Lai, and L. L. Lai, "New appliance detection for nonintrusive load monitoring," *IEEE Transactions on Industrial Informatics,* vol. 15, no. 8, pp. 4819-4829, 2019.