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.

Published in:
I E E E Industrial Electronics Magazine

DOI (link to publication from Publisher):
10.1109/MIE.2020.3023075

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Çimen, H., Garcia, E. J. P., Kolbæk, M., Cetinkaya, N., Vasquez, J. C., & Guerrero, J. M. (2021). Smart-Building Applications: Deep Learning-Based, Real-Time Load Monitoring. *I E E E Industrial Electronics Magazine*, *15*(2), 4-15. Article 9310683. https://doi.org/10.1109/MIE.2020.3023075

Smart-Building Applications

Deep Learning-Based, Real-Time Load Monitoring

HALIL ÇIMEN,
EMILIO JOSÉ PALACIOS-GARCÍA,
MORTEN KOLBÆK,
NURETTIN ÇETINKAYA,
JUAN C. VASQUEZ, and
JOSEP M. GUERRERO

Digital Object Identifier 10.1109/MIE.2020.3023075
Date of current version: 24 December 2020

Google's Director of Research Peter Norvig said, "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 foresight about the future. With the development of 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 benefits 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, with which they may be able to make more conscious decisions.

Previous studies have reported that, if instantaneous energy-consumption data are given to consumers as feedback, approximately 20% of energy savings can be achieved per household [1]. To reach this target, more detailed data on the electricity consumed by each appliance are needed. Smart meters cannot meet this need since they can read only the total electricity consumption.

1932-4529/21©2021IEEE

To overcome this issue, appliance load monitoring (ALM) is frequently used. ALM is used to monitor individual appliances in households by using sensors. Nonintrusive 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.

Currently, 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 that 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 the advantages of DL-based analysis are explained in detail. Also, a multilabel DL approach, which can save training time and reduce the need for model storage, is presented and tested in real time. Because the studies in the literature were carried out offline, the online analysis capacity of recent DL models was tested in a laboratory environment. In this way, the accuracy difference between offline and online implementations is revealed.

NILM

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

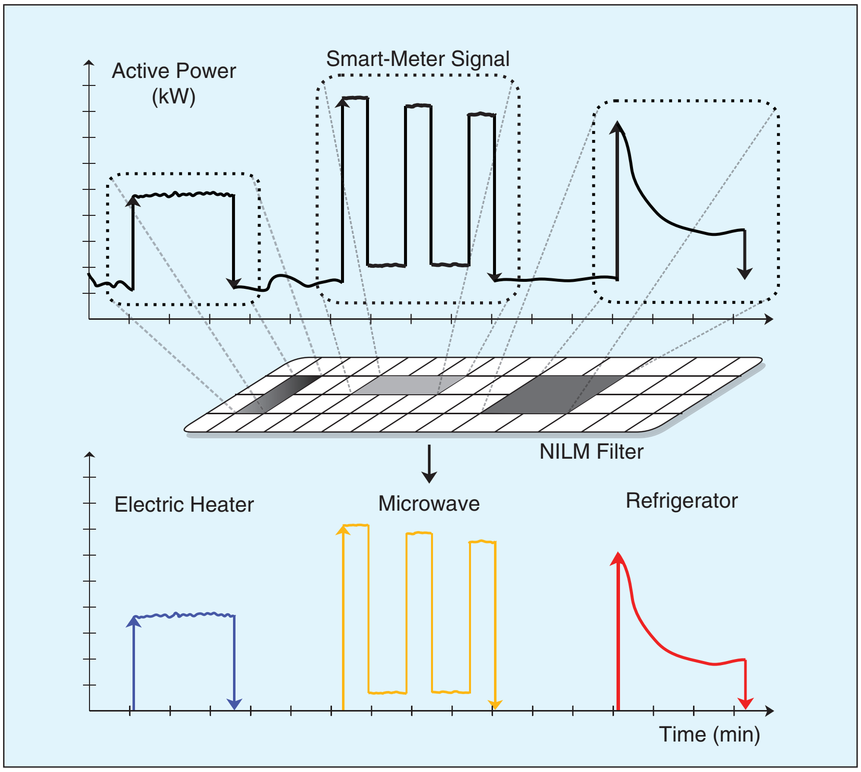

FIGURE 1 – An example of disaggregated data of residential appliances.

However, the main disadvantages are the need for comprehensive installation, communication infrastructure, maintenance, and updating. All of these features make the ILM a high-cost system, besides the data privacy breach. Users can be conservative in sharing 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 *nonintrusive*. Aggregated data are analyzed by various signal processing or pattern recognition methods to obtain the appliance-level disaggregated data. An example of data disaggregation is shown in Figure 1.

With a successful NILM analysis, real-time and statistical information about the appliances, their daily usage rate, and users' daily consumption behaviors 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 Figure 2.

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 by 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 data sets and the increase of data obtained from smart meters.

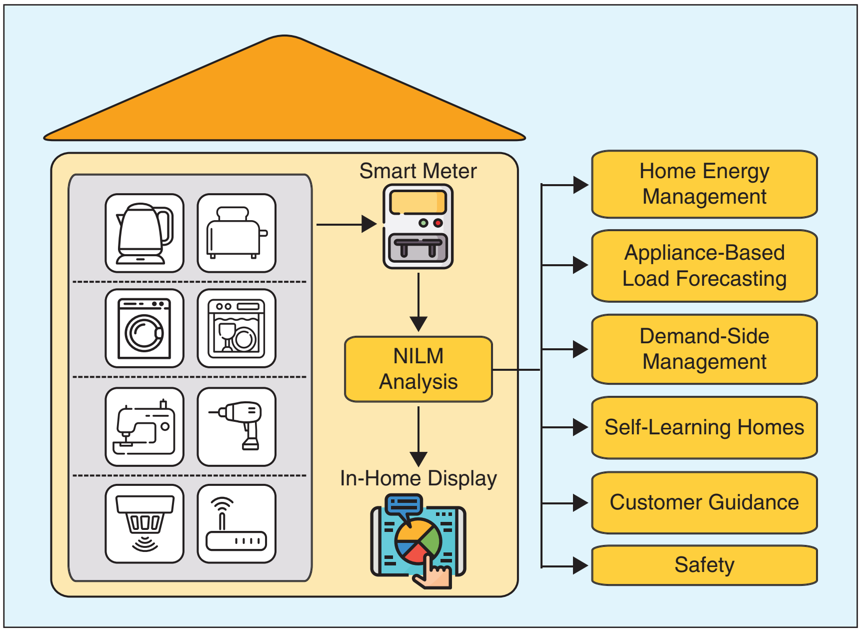

FIGURE 2 – The general NILM structure of a household case.

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 the power consumption of each appliance. The NILM problem can be formulated in simple form as follows:

$$p_{\text{agg}}(t) = \sum_{n \in N} s_n(t) \cdot p_n(t) + e(t), \quad (1)$$

where $p_{\text{agg}}(t)$ is the aggregated active power for sample t; and s_n and p_n are the status (on: one/off: zero) and 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 the *appliance signature*. To systematically address the NILM problem, appliances need to be classified. Hart [7] categorized appliances and divided them into three types according to their signatures. The types of appliances and their general signatures are shown in Figure 3. Type 1 appliances have only on/off states [e.g., a toaster (TO) or kettle (KT)]. On the contrary, type 2 appliances are those that have multiple states [e.g., a washing machine (WM) or tumble dryer (TD)]. Type 3 appliances consume variable power and do not have a specific state or periodic operation.

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 nontrivial 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 low (1 Hz and lower) and high resolution (higher than 1 Hz). There is a tradeoff between them. High-resolution data provide more detailed information but at high hardware cost. Low-resolution data provide limited information but are cost effective. It is more realistic to perform NILM analysis using low-resolution active power data since they are already available from smart meters. Detailed information on NILM analysis using high-resolution data can be found in [8] and [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 appliances that are turned on or off. Energy disaggregation is the process of estimating the energy consumption of appliances individually. A high-accuracy energy disaggregation might also provide information about the load identification.

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 reformulating (1) as follows:

$$\tilde{S}(t) = \underset{S}{\operatorname{argmin}} \left| p_{\text{agg}}(t) - \sum_{n \in N} s_n(t) \cdot \bar{p}_n \right|. \quad (2)$$

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 submetering data or using the appliance manual.

However, this method is not practical. First, 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 the *base load* and should be estimated by a prediction method or a statistical approach. Second, as the number of appliances increases, the length of the vector increases, and the solution space grows exponentially. In addition, appliances consuming similar power cannot be distinguished [10], [11].

Therefore, pattern recognition-based approaches, such as the 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 their robust analysis capabilities [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, achieve results similar or even (often) better than what would have been obtained with hand-engineered features. 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]. An illustrative comparison of ML and DL for the NILM application is shown in Figure 4.

DL techniques 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 three different DL models are frequently used: convolutional neural network (CNN), recurrent neural network (RNN), and autoencoder (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, the technique is called *S2S* if a sequence is estimated at the output and *S2P* if a single point is estimated.

Another CNN-based model, Wavenet, which was originally developed for raw audio generation, is implemented in [22]. The advantage of this model is that it can analyze longer input sequences with fewer parameters. It can be suitable for long-term-operating appliances, such as a dishwasher (DW). In [23] and [24], energy disaggregation is performed by using the AlexNet and Visual Geometry Group-16 models, which were originally developed for image classification.

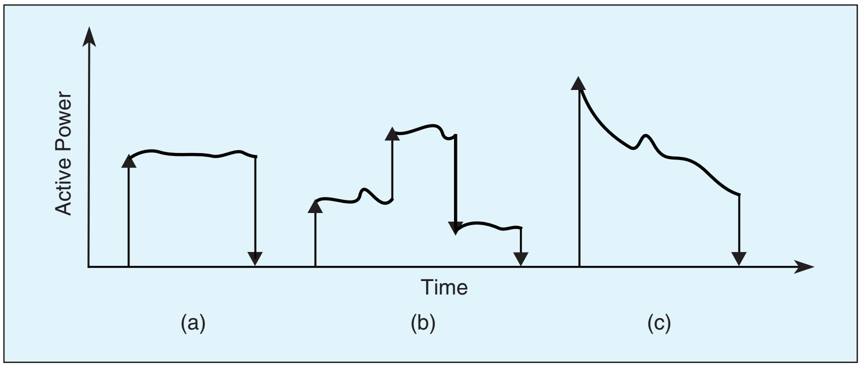

FIGURE 3 – The types of appliances: (a) type 1: two states, (b) type 2: multiple states, and (c) type 3: variable.

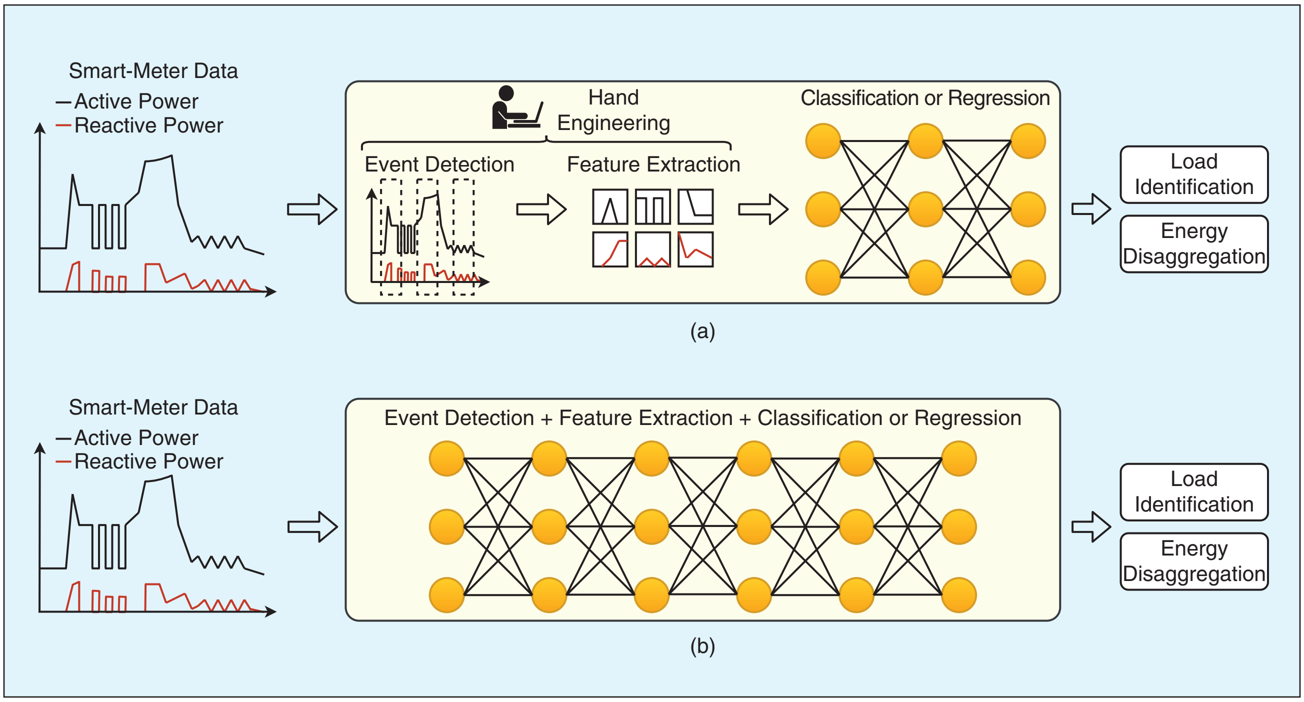

FIGURE 4 – A comparison of NILM with (a) ML and (b) DL.

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 past and 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, future data are mostly linked to 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. Therefore, it can be trained more quickly than LSTM. If a model can be trained faster, experiments can be conducted more efficiently, and, ultimately, the chance of finding good hyperparameters increases, which usually leads to better performance. In [25] and [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 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 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 AE (dAE) to filter noises. Although successful results were obtained for type 1 appliances, they were insufficient for type 2. 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.

Multilabel Convolutional GRU Architecture

When the studies mentioned earlier are examined, it is seen that each DL method has its advantages and disadvantages, yet, 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 article, a real-time load identification is performed using a convolutional GRU (C-GRU) model. The model architecture is shown in Figure 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 are 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. 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, and two bidirectional GRUs as well as two fully connected layers. For the convolutional layer, the filter size and number of filters are selected as three and 64, respectively. The GRU layers have 256 nodes, while the first fully connected layer has 128 nodes. The hyperbolic tangent is used in all hidden layers as the activation function.

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 article, multilabel 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 multilabel 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 adaptive moment estimation are used as the loss function and optimizer, respectively. This architecture is designed for supervised learning in which the input is aggregated data read from the smart meter and the 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 is assumed that the appliance is on.

Real-Time Evaluation of Different DL Models

The studies in the literature are conducted offline using publicly available data sets. During offline analysis, NILM is performed more easily since the whole energy-consumption period (past, current, and future) is available. In the online analysis, however, the appliances must 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 yet been addressed.

The most important factor affecting the real-time analysis is undoubtedly the selected window size w. In the literature, it is recommended that the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

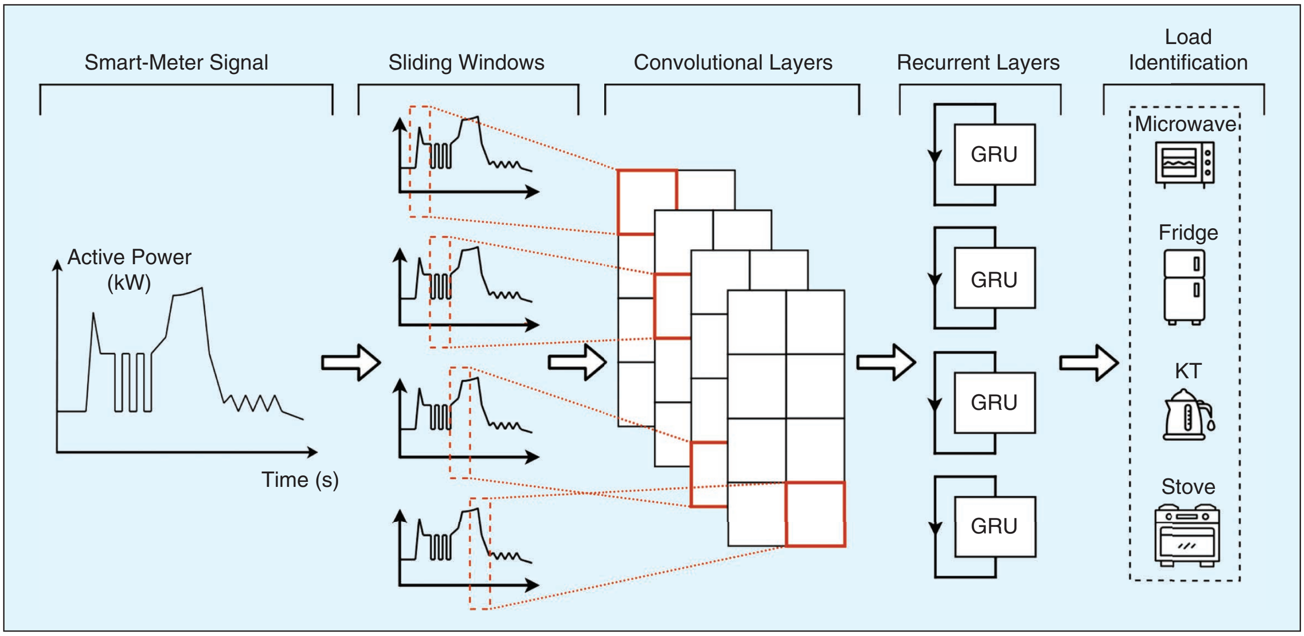

FIGURE 5 – The model architecture for real-time load identification.

window size be determined according to the operation cycle of the analyzed appliance [25]. For example, the window size should be selected as relatively long for appliances with a long operating time, such as DWs, to analyze their 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 Figure 6.

In Figure 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 to be as short as possible to instantly detect the appliance operation. In this article, the iterations are progressed with 60-s intervals.

Another important parameter, window size, should be chosen wisely. Since the proposed model has a multilabel classification structure, only one window size should be selected for appliances with both long and short operating times. Considering that short-term appliances, such as microwaves (MWs) and TOs, operate for an average of 5–10 min and long-term appliances, such as DWs, operate for an

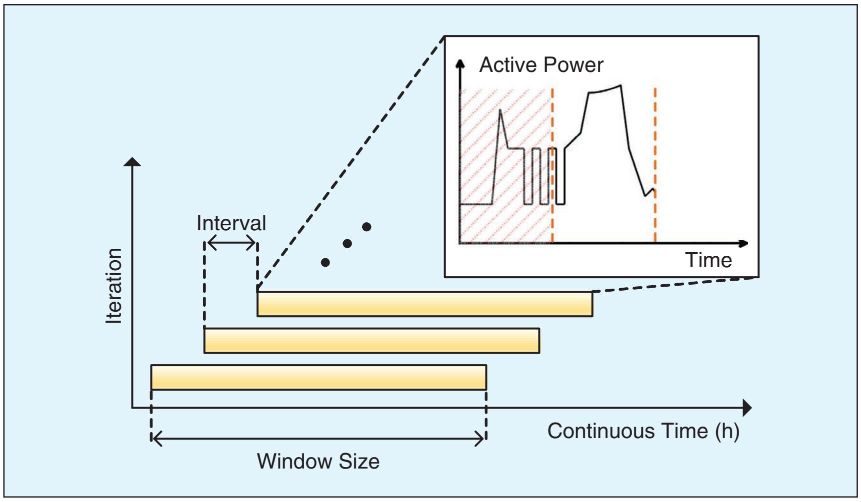

FIGURE 6 – The online analysis process.

average of 1 h, an average window size of 256 samples (approximately 20 min) that can be suitable for both types of appliances is determined for analysis.

Domestic appliances are basically divided into two groups: controllable and noncontrollable. The 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 article, two thermostatically controlled loads—refrigerator (FR) and heater (HE)—and seven deferrable loads—MW, KT, coffee maker (CM), DW, TD, WM, and TO—are taken into consideration for real-time identification. In addition, appliances such as WMs, DWs, and TDs (around 1.8 kW) as well as HEs, MWs, KTs (around 1 kW) have similar power consumptions or peak points. Thus, it will be 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 Figure 7.

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 data set contains

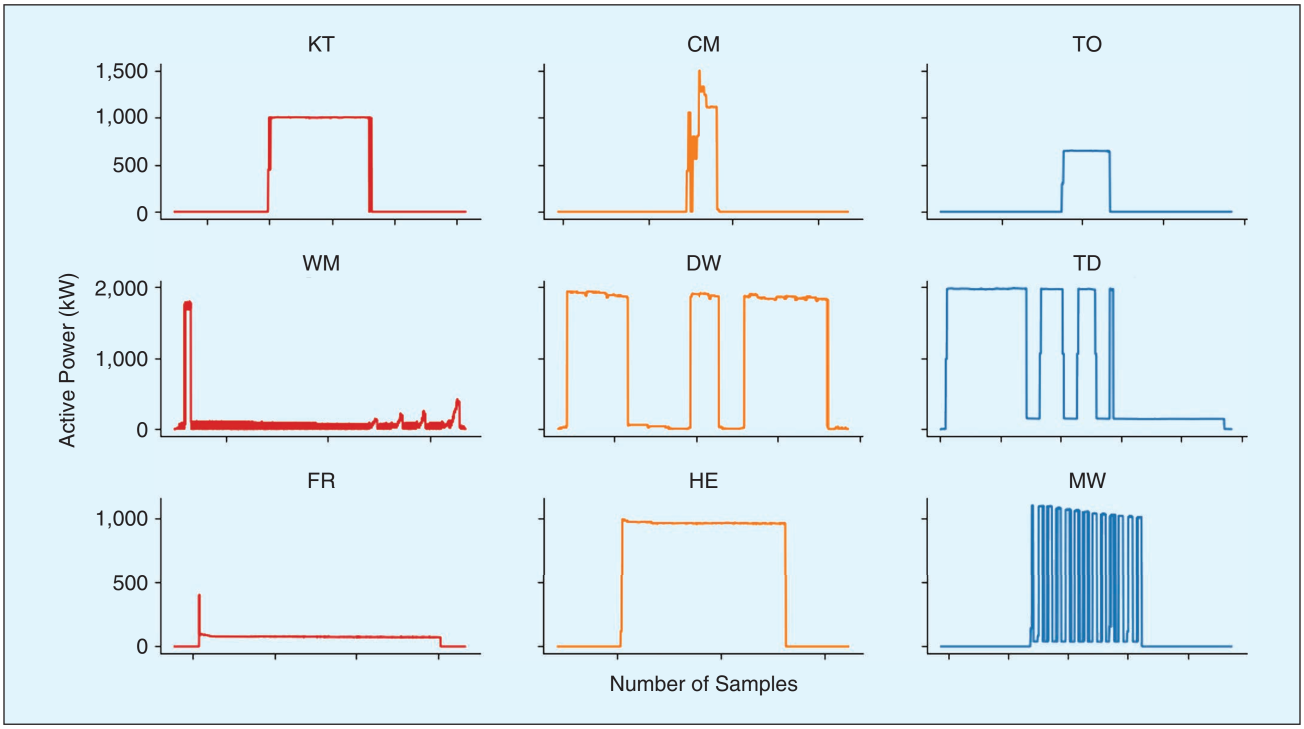

FIGURE 7 – The signatures of the target appliances.

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 are modified to include missing data, the model can also handle missing points that will occur during online analysis.

Second, the usage frequency of the target appliance should be analyzed. For example, if a household's aggregated data covers one 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 data set, 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 data set as new data. In this article, 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 data set.

In the final step, the sampling frequency of target appliances and aggregated power consumption should be adjusted for proper supervised learning. The frequency of the data read from the sensors is not regular and changes between 5 and 10 s. First, an up-sampling 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 s since the data with 1-s 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.

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 Figure 8.

A Dell workstation with a six-core Intel Xeon CPU at 3.60 GHz, 32 GB of random-access memory (RAM), and a dedicated GPU NVIDIA Quadro P600 running on the Community Enterprise Operating System was used for the training and initial tests. In addition, the final trained networks were deployed on a Windows 10 laptop with an i5 (second-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 the Keras library.

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

$$\text{recall} = \frac{\text{TP}}{\text{TP}+\text{FN}},$$
$$\text{precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$$
$$F_1 = 2\times\frac{\text{precision}\times\text{recall}}{\text{precision}+\text{recall}},$$
$$\text{accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}, \quad (3)$$

where true positive (TP) and true negative (TN) indicate that the model correctly predicts that the appliance is on and off, respectively. False positive

(FP) and false negative (FN) are outputs where the model incorrectly predicts that 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 TO is used only once or twice a day. The DL model will achieve an accuracy of more than 99%, even if it predicts that the TO 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 the preferred metric because it is interpreted as a weighted average of precision and recall. The F_1 score comparison of online and offline analysis results for different types of DL models is shown in Table 1. 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 four appliances were operated simultaneously with different combinations.

The results can be evaluated from three 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. This 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 1 appliances used in this experiment have short operating times of around 2–4 min. Since the window size is determined around 20 min, their consumption may be perceived as small spikes in this window. For this reason, the success of the analysis is between 65% and 82%. Type 2 appliances are long-running and multistate appliances, which makes their signature distinctive. Analysis success is higher than for type 1 appliances because more precise connections can be established among the past, current, and future. The energy consumption of type 3 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 analyses. 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 cycles of the appliances are completed. Therefore, they are not sufficiently detected, or the wrong appliances are considered active. As new data are read, however, the success rate increases. An average of 5% accuracy loss can be reported between online and offline analysis.

In addition, the analysis success for WMs and MWs was higher than for other appliances. The main reason behind this is their distinctive signatures. As seen in Figure 7, most appliances somehow have a rectangular energy-consumption profile. However, since WMs and MWs 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 Figure 9. In Figure 9(a), using different window sizes affects the F_1 score. Since GRU and LSTM have long-term memory, accuracy increases with increasing 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 the 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 an *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 h. As seen in Figure 9(a), the F_1 score difference between the multilabel C-GRU and appliance-specific approach is very small. However, there is a big difference in 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 the household or building level. This implies the use of embedded edge-computing systems or even existing home or building energy-management systems 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 periods are longer. Although RNN is trained in a shorter time compared to GRU and LSTM, its analysis success remains insufficient. GRU can be trained faster than LSTM, and slightly better results can be achieved.

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 MWs, DWs, and CMs for simple aggregated data examples are shown in Figure 10.

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 multilabel C-GRU model is proposed that makes it possible to train and test multiple appliances with a single DL model. In this way, both significant time saving

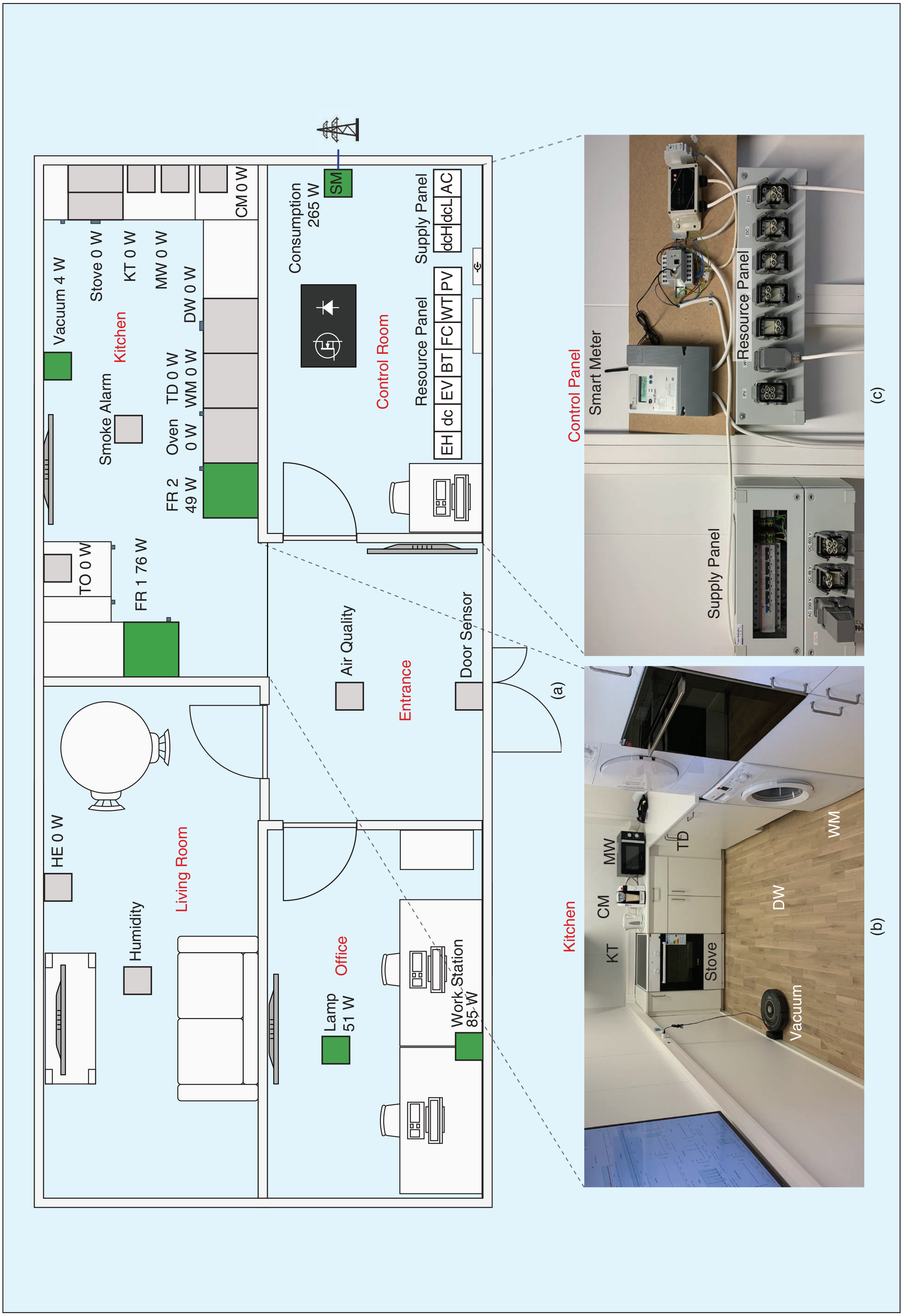

FIGURE 8 – The IoT-MGLab general overview: (a) an overall map, (b) the kitchen, and (c) the control panel. ac: 230 V ac; BT: battery; dc: dc connection; dcH: 400 V dc; dcL: 48 V dc; EH: electrical heating; EV: electric vehicle; FC: fuel cell; PV: photovoltaic panels; SM: smart meter; WT: wind turbine.

TABLE 1 – AN 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 1	KT	0.714	0.116	0.694	0.738	**0.822**	0.62	0	0.597	0.701	**0.755**
	CM	0.678	0.084	0.508	0.665	**0.732**	0.522	0	0.358	0.592	**0.678**
	TO	0.549	0.161	0.351	**0.682**	0.651	0.395	0	0.219	0.651	**0.661**
Type 2	WM	0.938	0.954	0.893	0.94	**0.962**	0.924	0.897	0.914	**0.952**	0.939
	DW	0.662	0.72	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 3	FR	0.679	0.661	0.675	0.764	**0.777**	0.688	0.653	0.69	**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

Values in bold type represent the best results obtained for offline and online analysis.

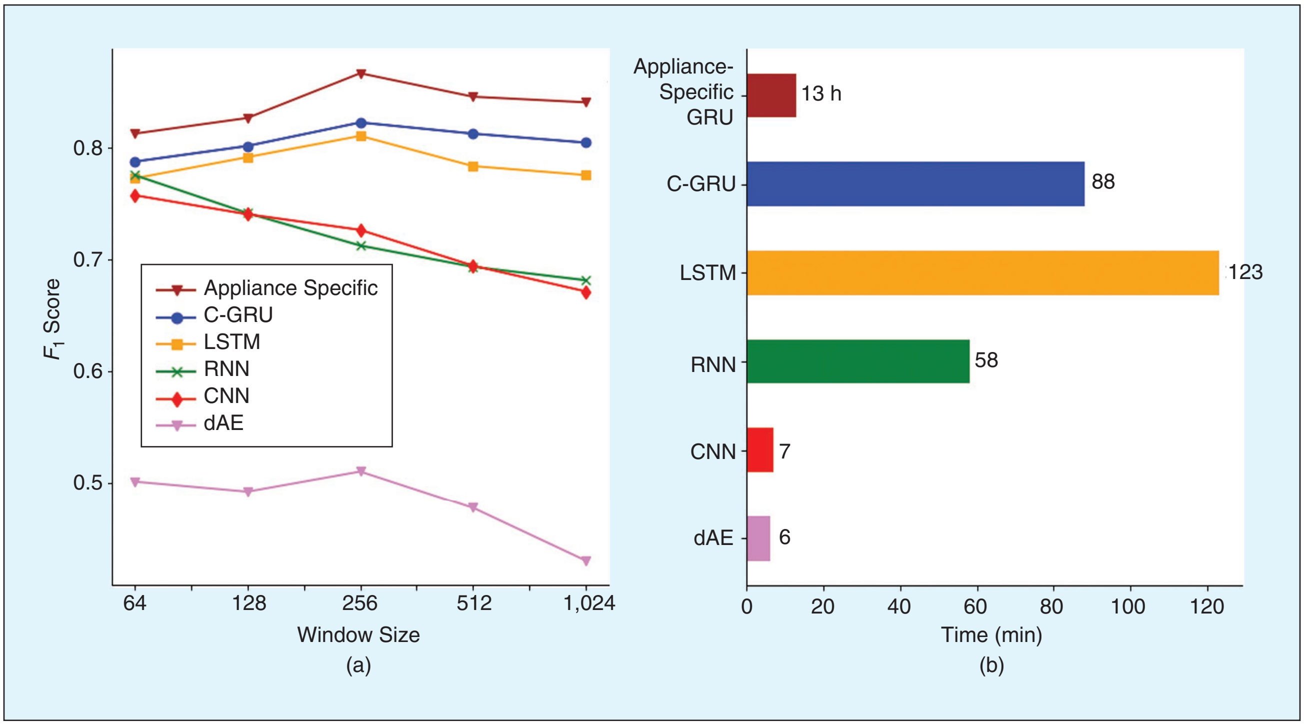

FIGURE 9 – The (a) effect of window size selection and (b) comparison of the training times.

is achieved, and the need for data storage can be reduced, which are critical factors for the integration of such algorithms at the 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 LSTM and GRU models outperformed CNN and dAE approaches. Therefore, it is recommended that new DL models to be developed be compared with recurrent-based techniques. In this regard, C-GRU is trained faster than LSTM, and slightly better results are obtained compared 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, WMs and MWs 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 the consumption profile rather 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 and offline analysis successes of DL models. This difference should be considered for the real-time load identification required for demand-response applications. In addition, the difference may increase with a greater number of analyzed appliances. This increase can be mitigated by using either more robust DL models or postprocessing, which is the approach of refining the results with the help of various optimization algorithms. Accuracy rates

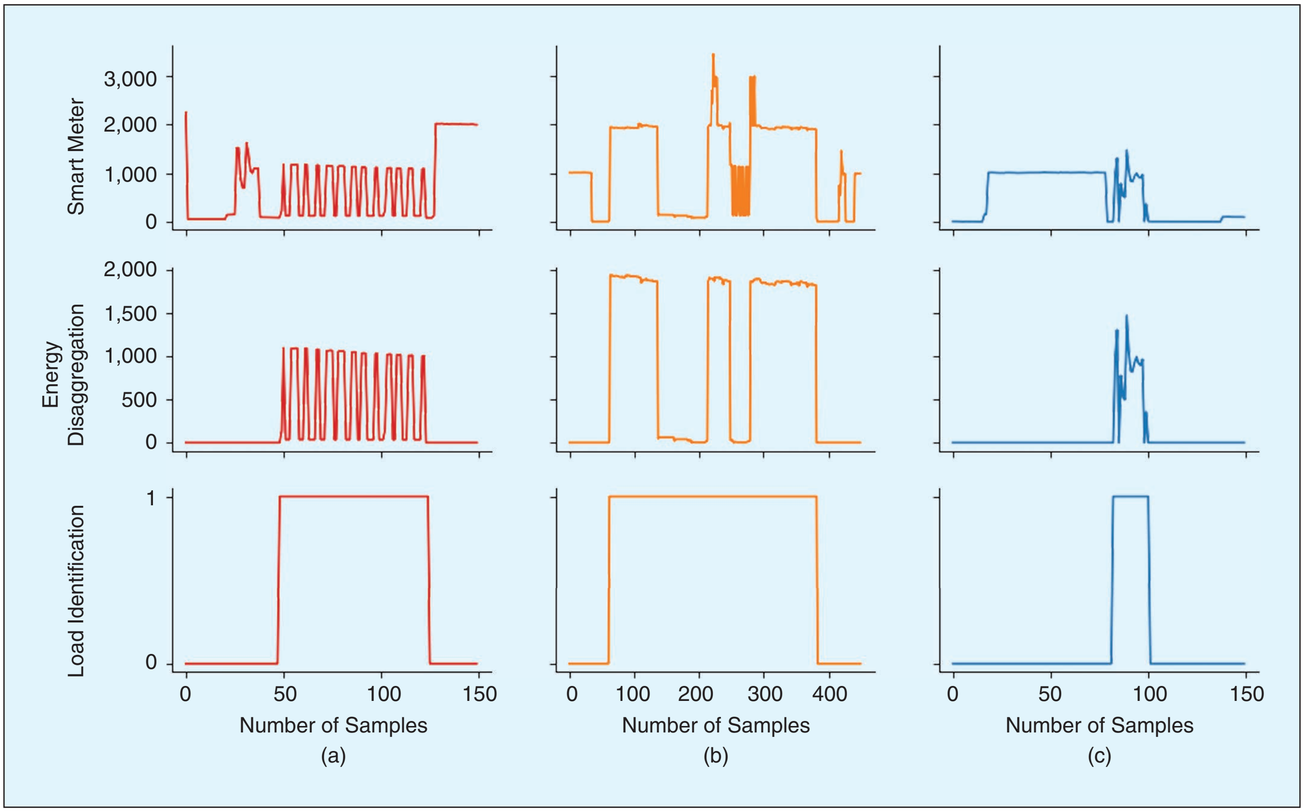

FIGURE 10 – The results of energy disaggregation and load identification for the (a) MW, (b) DW, and (c) CM.

can be improved by reanalyzing the outputs of DL models. In 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.

Acknowledgments

This work was supported by the Scientific and Technological Research Council of Turkey BIDEB-2214 International Doctoral Research Fellowship Program, VILLUM FONDEN under the VILLUM Investigator Grant 25920: Center for Research on Microgrids (www.crom.et.aau.dk), and the Aalborg University Talent Program 2016 with the research project "The Energy Internet—Integrating the Internet of Things Into the Smart Grid" (771116).

Biographies

Halil Çimen (hcimen@ktun.edu.tr) earned B.Eng. and M.Sc. degrees in electrical electronics engineering from Selçuk University, Turkey, in 2011 and 2016, respectively, and is currently working toward his Ph.D. degree at Konya Technical University, Konya, Turkey. From April 2019 through July 2020, he was an academic visiting scholar with the Department of Energy Technology at Aalborg University, Denmark. His research interests include distributed generation, nonintrusive load monitoring, and machine learning-supported energy management. During his research at Aalborg University, he gained experience in applying deep learning methods to load-monitoring studies as part of the Energy Internet project.

Emilio José Palacios-García (emiliojose.palaciosgarcia@kuleuven.be) earned a Ph.D. degree in advanced computation and energy in 2018. He is a postdoctoral research associate with the Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium, and part of the EnergyVille Research Centre, Genk, Belgium. From 2018 to 2020, he worked as a postdoctoral researcher with the Department of Energy Technology at Aalborg University, Denmark. His research interests include stochastic modeling, smart meters, and artificial intelligence methods applied to energy management. He is a Member of IEEE and the IEEE Industrial Electronics Society.

Morten Kolbæk (mok@es.aau.dk) earned a B.Eng. degree in electronic design from Aarhus University in 2013, an M.Sc. degree in signal processing and computing from Aalborg University in 2015, and a Ph.D. degree from Aalborg University in 2018 for the thesis "Single-Microphone Speech Enhancement and Separation Using Deep Learning." He is a postdoctoral researcher in the Section for Signal and Information Processing in the Department of Electronic Systems, Aalborg University, Aalborg, Denmark. His research interests include single-microphone algorithms for speech enhancement and multitalker speech separation, machine learning (deep learning in particular), and intelligibility improvement of noisy speech. He is a Member of IEEE.

Nurettin Çetinkaya (ncetinkaya@ktun.edu.tr) earned B.S., M.Sc., and Ph.D. degrees in electric and electronics engineering from Selçuk University, Konya, Turkey, in 1993, 1998, and 2005, respectively. He is an associate professor with the Department of Electrical and Electronics Engineering, Konya Technical University, Konya, Turkey. Until 2005, he worked as a research assistant at Selçuk University. He also worked as an advisor in the Konya organized industrial zone. His research interests include power system planning, smart grids, energy efficiency, renewable energy, power system economics, and management. He received publication awards given by the Scientific and Technological Research Council of Turkey.

Juan C. Vasquez (juq@et.aau.dk) earned a Ph.D. degree from the Polytechnic University of Catalonia (currently referred to as BarcelonaTech), Spain. In 2019, he became a professor in the Department of Energy Technology, Aalborg University, Aalborg, Denmark, and he is the codirector of the Villum Center for Research on Microgrids, Aalborg, Denmark. His research interests include operation, control, and energy management applied to ac/dc microgrids as well as the integration of the Internet of Things, energy Internet, digital twins, and blockchain solutions. He was recognized as a "Highly Cited Researcher" in 2017 and was the recipient of the Young Investigator Award in 2019. He is a Senior Member of IEEE and a member of the IEEE Industrial Electronics Society.

Josep M. Guerrero (joz@et.aau.dk) earned a Ph.D. degree from the Technical University of Catalonia, Barcelona, in 2003. He is a full professor with the Department of Energy Technology, Aalborg University, Aalborg, Denmark, where he is the director of the Center for Research on Microgrids. His research interests include distributed energy-storage systems, hierarchical and cooperative control, energy-management systems, smart metering, and the Internet of Things for ac/dc microgrid clusters. He was recognized as an Institute for Scientific Information "Highly Cited Researcher" by Thomson Reuters from 2014 to 2019 and a Villum Investigator in 2019. He is a Fellow of IEEE and a member of the IEEE Industrial Electronics Society.

REFERENCES

[1] E. Aydin, D. Brounen, and N. Kok, "Information provision and energy consumption: Evidence from a field experiment," *Energy Econ.*, vol. 71, pp. 403–410, Mar. 2018. doi: 10.1016/j.eneco.2018.03.008.

[2] L. Li, K. Ota, and M. Dong, "Deep learning for smart industry: Efficient manufacture inspection system with fog computing," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4665–4673, 2018. doi: 10.1109/TII.2018.2842821.

[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 Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3794–3803, 2019. doi: 10.1109/TIE.2018.2856193.

[4] W. Li, H. Li, Q. Wu, X. Chen, and K. N. Ngan, "Simultaneously detecting and counting dense vehicles from drone images," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9651–9662, 2019. doi: 10.1109/TIE.2019.2899548.

[5] W. Peng, Z. Ye, and N. Chen, "Bayesian deep-learning-based health prognostics toward prognostics uncertainty," *IEEE Trans. Ind. Electron.*, vol. 67, no. 3, pp. 2283–2293, 2020. doi: 10.1109/TIE.2019.2907440.

[6] A. Aboulian et al., "NILM dashboard: A power system monitor for electromechanical equipment diagnostics," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1405–1414, 2018. doi: 10.1109/TII.2018.2843770.

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

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

[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. 16,838–16,866, 2012. doi: 10.3390/s121216838.

[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 Trans. Smart Grid*, vol. 10, no. 6, pp. 6472–6480, 2019. doi: 10.1109/TSG.2019.2906012.

[11] J. Liang, S. K. Ng, G. Kendall, and J. W. Cheng, "Load signature study—Part I: Basic concept, structure, and methodology," *IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 551–560, 2009. doi: 10.1109/TPWRD.2009.2033799.

[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 Trans. Smart Grid*, vol. 9, no. 4, pp. 3362–3372, 2018. doi: 10.1109/TSG.2016.2631238.

[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 Trans. Smart Grid*, vol. 9, no. 4, pp. 3079–3090, 2016. doi: 10.1109/TSG.2016.2626389.

[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 Trans. Smart Grid*, vol. 7, no. 6, pp. 2575–2585, 2015. doi: 10.1109/TSG.2015.2494592.

[15] M. A. Mengistu, A. A. Girmay, C. Camarda, A. Acquaviva, and E. Patti, "A cloud-based online disaggregation algorithm for home appliance loads," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3430–3439, 2019. doi: 10.1109/TSG.2018.2826844.

[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. 55,937–55,952, Mar. 2020. doi: 10.1109/ACCESS.2020.2981969.

[17] S. M. Tabatabaei, S. Dick, and W. Xu, "Toward non-intrusive load monitoring via multi-label classification," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 26–40, 2016. doi: 10.1109/TSG.2016.2584581.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. doi: 10.1038/nature14539.

[20] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for non-intrusive load monitoring," presented at the 32nd AAAI Conf. Artific. Intell., 2018.

[21] K. Chen, Q. Wang, Z. He, K. Chen, J. Hu, and J. He, "Convolutional sequence to sequence non-intrusive load monitoring," *J. Eng.*, vol. 2018, no. 17, pp. 1860–1864, 2018. doi: 10.1049/joe.2018.8352.

[22] J. Jiang, Q. Kong, M. Plumbley, and N. Gilbert, "Deep learning based energy disaggregation and on/off detection of household appliances," 2019, arXiv:1908.00941.

[23] G. Cui, B. Liu, W. Luan, and Y. Yu, "Estimation of target appliance electricity consumption using background filtering," *IEEE Trans. Smart Grid*, vol. 10, no. 6, 2019. doi: 10.1109/TSG.2019.2892841.

[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 Trans. Smart Grid*, vol. 11, no. 1, 2019. doi: 10.1109/TSG.2019.2918330.

[25] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ.*, 2015, pp. 55–64. doi: 10.1145/2821650.2821672.

[26] L. Mauch and B. Yang, "A new approach for supervised power disaggregation by using a deep recurrent LSTM network," in *Proc. 2015 IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, pp. 63–67. doi: 10.1109/GlobalSIP.2015.7418157.

[27] D. Murray, L. Stankovic, V. Stankovic, S. Lulic, and S. Sladojevic, "Transferability of Neural Network approaches for low-rate energy disaggregation," in *Proc. ICASSP 2019-2019 IEEE Int. Conf. Acoustics, Speech Signal Process. (ICASSP)*, pp. 8330–8334. doi: 10.1109/ICASSP.2019.8682486.

[28] T. Sirojan, B. T. Phung, and E. Ambikairajah, "Deep neural network based energy disaggregation," in *Proc. 2018 IEEE Int. Conf. Smart Energy Grid Eng. (SEGE)*, pp. 73–77. doi: 10.1109/SEGE.2018.8499441.

[29] J. Zhang, X. Chen, W. W. Ng, C. S. Lai, and L. L. Lai, "New appliance detection for nonintrusive load monitoring," *IEEE Trans. Ind. Informat.*, vol. 15, no. 8, pp. 4819–4829, 2019. doi: 10.1109/TII.2019.2916213.