sciendo

CAIM
ISSN 2038-0909

Commun. Appl. Ind. Math. 13 (1), 2022, 39–56 DOI: 10.2478/caim-2022-0004

Research Article

Deep learning based non-intrusive load monitoring with low resolution data from smart meters

Marco Manolo Manca[1] **, Luca Massidda**[2*]

[1]University of Cagliari, Department of Mathematics and Computer Science, Via Ospedale 72, 09124 Cagliari (CA), Italy; marcom.manca@unica.it

[2]CRS4, Center for Advanced Studies, Research and Development in Sardinia, loc. Piscina Manna ed. 1, 09050 Pula (CA), Italy; luca.massidda@crs4.it

[*]Email address for correspondence: luca.massidda@crs4.it

Communicated by Elena De Angelis

Received on 05 02, 2022. Accepted on 09 16, 2022.

Abstract

A detailed knowledge of the energy consumption and activation status of the electrical appliances in a house is beneficial for both the user and the energy supplier, improving energy awareness and allowing the implementation of consumption management policies through demand response techniques. Monitoring the consumption of individual appliances is certainly expensive and difficult to implement technically on a large scale, so non-intrusive monitoring techniques have been developed that allow the consumption of appliances to be derived from the sole measurement of the aggregate consumption of a house. However, these methodologies often require additional hardware to be installed in the domestic system to measure total energy consumption with high temporal resolution. In this work we use a deep learning method to disaggregate the low frequency energy signal generated directly by the new generation smart meters deployed in Italy, without the need of additional specific hardware. The performances obtained on two reference datasets are promising and demonstrate the applicability of the proposed approach.

Keywords: energy disaggregation; non-intrusive load monitoring; NILM; deep learning; smart meter; Chain 2

AMS subject classification: 68T10

1. Introduction

Non-Intrusive Load Monitoring (NILM), or load disaggregation, aims to identify the operating state (on/off) and precise energy consumption of individual electrical loads, considering only the aggregate consumption of these loads as input.

Implementing an effective NILM technique for domestic and industrial use is a valuable resource for both the consumer and the utility. For a domestic user, for example, the knowledge of the consumption of household appliances over time promotes greater awareness of energy consumption, allowing an informed choice of consumption habits [1]. It can be used to rationally distribute and/or reduce load throughout the day, to effectively manage home automation systems, to participate in demand response programs [2,3]. For utility companies on the other hand, the knowledge of the habits and needs of individual users allows the offer of personalized services, a better segmentation of users [1] and a better scheduling of supply [4,5], greater accuracy in the prediction of consumption, and facilitates the implementation of demand response strategies.

The NILM technique was introduced by Hart's pioneering work in the mid-1980s [6]. In fact, he was the first to use measured active and reactive power time series to estimate the on/off status of individual devices. Research on this topic has had a strong push in the last decade, due to new requirements in power grid management systems and thanks to the developments in machine learning and deep learning techniques.

Recent work on the subject can be divided into two main groups, distinguished by the sampling frequency of the signals used for analysis. The first group of these papers is based on the availability of power data measured at high sampling rates (from 50Hz to over kHz), obtained through dedicated hardware, with which to obtain the identification of switching events [7,8]. The second group, which also represents the most current line of research, is based on lower frequency aggregate load analysis methods, obtained with dedicated hardware or directly from the smart meter with sampling periods ranging from 1s to 1 hour. This approach has been shown to be effective in obtaining, sometimes with reasonable accuracy, an estimate of the energy consumed by appliances with higher power demand even if used occasionally [9–12].

NILM studies have tested both supervised and unsupervised approaches, depending on the information available and the predictive algorithm implemented. Supervised methods require a data set labeled with sub-metered devices, which is not always available. Unsupervised methods can be implemented without any prior knowledge of the environment, but the user is usually required to match the patterns identified by the algorithm to the appliances.

Early NILM studies were primarily based on hidden Markov models (HMM), which are typically used for probabilistic modeling of time series data [13,14], these techniques are commonly inefficient when the number of devices increases and also suffer from high computational complexity [1].

A different approach is that of optimization methods, where the main idea is to find the optimal combination of the individual devices that make up the aggregated signal. The increased complexity resulting from a large number of devices and the loss of temporal continuity are two major drawbacks to using these techniques [15]. Thermal type loads, moreover, can be effectively disaggregated using Bayesian techniques [16].

Machine learning-based (non deep) approaches require manual feature construction using domain expert knowledge. These solutions have lower computational complexity than deep learning approaches and in some cases have yielded encouraging results. Numerous techniques for both regression and classification are used, such as K Nearest Neighbours [17], Support Vector Machines [18,19], and especially tree-based methods that achieve performance comparable to newer deep learning techniques [20–22]

All machine learning methods require the practitioner to define descriptive features of the load that are then processed by regression or classification algorithms. In the deep learning approach, feature synthesis becomes an integral part of the machine learning process; in fact, the first layers of the deep network are used for processing the raw signal to obtain meaningful features that are then processed by subsequent layers.

A widely used approach is Convolutional Neural Networks (CNN), in which the first layers consist precisely of convolution layers that operate as filters on the raw signal [23–26]. Subsequent layers use other types of layers such as pooling layers, normalization layers, and fully connected layers. The most common approach involves the use of 1d convolution modules that receive as input a time sequence of the aggregated load signal and return as output, the value of the load of one or more appliances, or the identification of its activation state, for a portion of the input sequence or even for just the midpoint of the interval.

Recurrent Neural Networks, are a type of deep learning architecture designed for temporal sequences, in which the basic concept is that information about previous states is part of the input for the next state. The most recent versions of this type of network are represented by the Gated Recurrent Unit (GRU) and Long Short Term Memory (LSTM) architectures. Several architectures with different complexity have been tested in the literature for NILM, in which usually the first layers are convolutional layers anyway, with performance comparable to that obtained with convolutional networks [23,27–29].

Another architecture studied in the literature is autoencoders (AE), an architecture consisting of an encoder, which compresses the input signal to a representation on a latent space and then reconstructs the signal itself using a portion of the network called a decoder. The latent space representation of the aggregate consumption signal can be used to extract information about the consumption of individual appliances [23,30,31].

More recent developments for processing sequential data include architectures that implement the attention mechanism, which, unlike recurrent architectures, uses all previous states of an encoder to construct the input for the decoder, in an overall sequence-to-sequence configuration [32,33].

In [34] we used techniques from semantic image segmentation to build a deep learning model using convolutional networks for load disaggregation on a low frequency signal. The model proved to be very effective in disaggregating the loads of a house even in the case, more interesting from the application point of view, in which the neural network is applied to the load of a house not present in the dataset used for the training of the model. We have also experimented with network ensembling techniques to further improve the already good accuracy achieved by the network [35]. One of the peculiarities of the proposed model is the use of data with a reduced sampling rate (1 min) for both training and testing, in order to allow a possible application of the technique to measurements coming directly from the smart meter without the use of dedicated monitoring hardware.

Most of the techniques proposed in the literature in fact assume the availability of a measurement of total consumption of the user with a sampling rate higher than that generally obtainable from smart meters installed by the distributor, so they require a minimum of intrusiveness for the user as it is often necessary to equip the house with a measurement system of total instantaneous power dedicated to the disaggregation application.

Our interest is instead to evaluate the applicability of the methodology developed to the measurement signal obtainable from new generation smart meters being deployed throughout Europe. In particular, in this work we want to verify the applicability of the model for a *Chain 2* signal, the channel of communication on conveyed waves (PLC-C) adopted in the second generation smart meters distributed in Italy [36,37]. The *Chain 2* signal combines a quarter-hourly consumption measurement with a signal for load variation events; a measurement of the instantaneous load is sent every time it crosses pre-set threshold values. In this paper, we will demonstrate that the approach we have developed is also applicable to this type of signal by allowing both user-side and utility-side load disaggregation without the need for dedicated measurement hardware. We will apply the neural network to two datasets in which the aggregated signal is filtered to reproduce a *Chain 2* type signal and we will verify the achievable performance. To our knowledge this is the first work to propose a disaggregation on the signal directly obtained from a commercial smart meter.

The article is structured as follows: the problem is formulated in Section 2.1, the methodology proposed for its solution is described in Section 2.2, Section 3 describes numerical experiments conducted on a reference dataset, the results of which are presented in Section 4; conclusions are drawn in Section 5 where some possible further developments of this research are presented.

2. Materials and Methods

2.1. *Problem formulation*

If $y(t)$ represents the total active electrical power consumed at instant t, and we denote by $y_i(t)$ the active power absorbed by the appliance with index i at the same instant, the total load can be expressed as the sum of the absorptions of the individual appliances and an unmeasured portion:

$$y(t) = \sum_{i=1}^{N} y_i(t) + e(t) \quad \forall t \in (0, T) \tag{1}$$

where N is the number of appliances considered and $e(t)$ is the unidentified residual load.

The problem is to get the values of $y_i(t)$ when only the measure of $y(t)$ is known, that is to get an approximation of $F(y(t))$:

$$[y_1(t), y_2(t), \ldots, y_i(t), \ldots, y_N(t)] = F(y(t)) \tag{2}$$

where F is the operator that, when applied to the total active power, returns N distinct values that are the best estimate of the power absorbed by individual appliances. Note that, in general, $y_i(t)$ does not represent all electrical appliances, but a fraction of all those present in a house. The unknown term $e(t)$ thus takes into account loads due to unmonitored devices. The task of finding an approximation of the operator F can be set up as a supervised learning problem when simultaneous measurements of aggregate load and consumption of individual appliances are available.

If, as in our case, we are primarily interested in cumulative consumption and activation times, the estimated consumption $\hat{y}_i(t)$ of individual devices can be approximated by a function that is constant over the activation period of the device:

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

where p_i is the average consumption of the appliance i and the $\hat{a}_i(t)$ is an estimate of the state of activation of the individual appliance at the time t, which has a unit value if the appliance is in operation and zero value otherwise.

Therefore, the method we propose seeks to obtain the most accurate possible estimate of the state of activation of the appliances starting from the aggregate load,

$$[\hat{a}_1(t), \hat{a}_2(t), \ldots, \hat{a}_i(t), \ldots, \hat{a}_N(t)] = F_a(y(t)) \tag{4}$$

and obtains an estimate of consumption using Equation 3 after knowing the average nominal consumption of the appliances examined.

2.2. *Methodology*

In [34] we proposed a methodology to obtain the F_a using a convolutional neural network architecture we called Temporal Pooling NILM (TP-NILM) an adaptation of the network called PSPNet (Pyramid Scene Parsing Network) proposed by Zhao et al. in [38] for the semantic segmentation of images.

The general scheme follows the classical approach to semantic image segmentation, where we have an encoder, which allows to increase the feature space of the signal at the cost of a reduction in its temporal resolution, and a decoder module that reconstructs an estimate of the activation state of the device at the same resolution as the original signal. In addition to these, there is a module called Temporal Pooling that performs feature aggregation at different resolutions, generating a temporal context, which spans long periods without completely losing resolution in the signal description.

The network layout is shown in Figure 1.

Unlike other approaches presented in the literature, in this case the model tries to recognize only the activation periods of the appliances under consideration and does not try to reconstruct the detail of their absorption over time. The task is less related to the characteristics of the individual appliance, so a model trained on a few users can provide reasonable results even on a user, and on an appliance, that does not enter the training set, but that retains absorption characteristics typical of the appliance class to which it belongs.

The network weights are obtained by gradient descent optimization. The loss function is a binary Cross-Entropy applied to each of the output channels that measures the difference between the activations estimated by the network $\hat{a}_i(t)$ and the actual activations $a_i(i)$ for each device examined and for each instant of the time period under consideration. The network was implemented using the PyTorch library [39].

2.3. *Chain 2 protocol*

In 2019, the EU revised its energy policy framework through the Clean Energy for All Europeans package, marking a significant step toward implementing the Union's energy strategy. The importance of smart metering in delivering energy efficiency and empowering consumers by enabling their active participation in the electricity market is reaffirmed, also through the coupling of smart meters with consumer energy management systems. The directive states, among other things, that smart meters must send information about consumption to end-users: historical consumption and near real-time consumption [40].

The second generation Smart Meters distributed in Italy provide the possibility of communication with a user device through the new channel called *Chain 2* on power line (PLC-C), compliant with standard [41]. Some works have already described the potential of this technology and presented the communication properties between the in-home device and the smart meter [36].

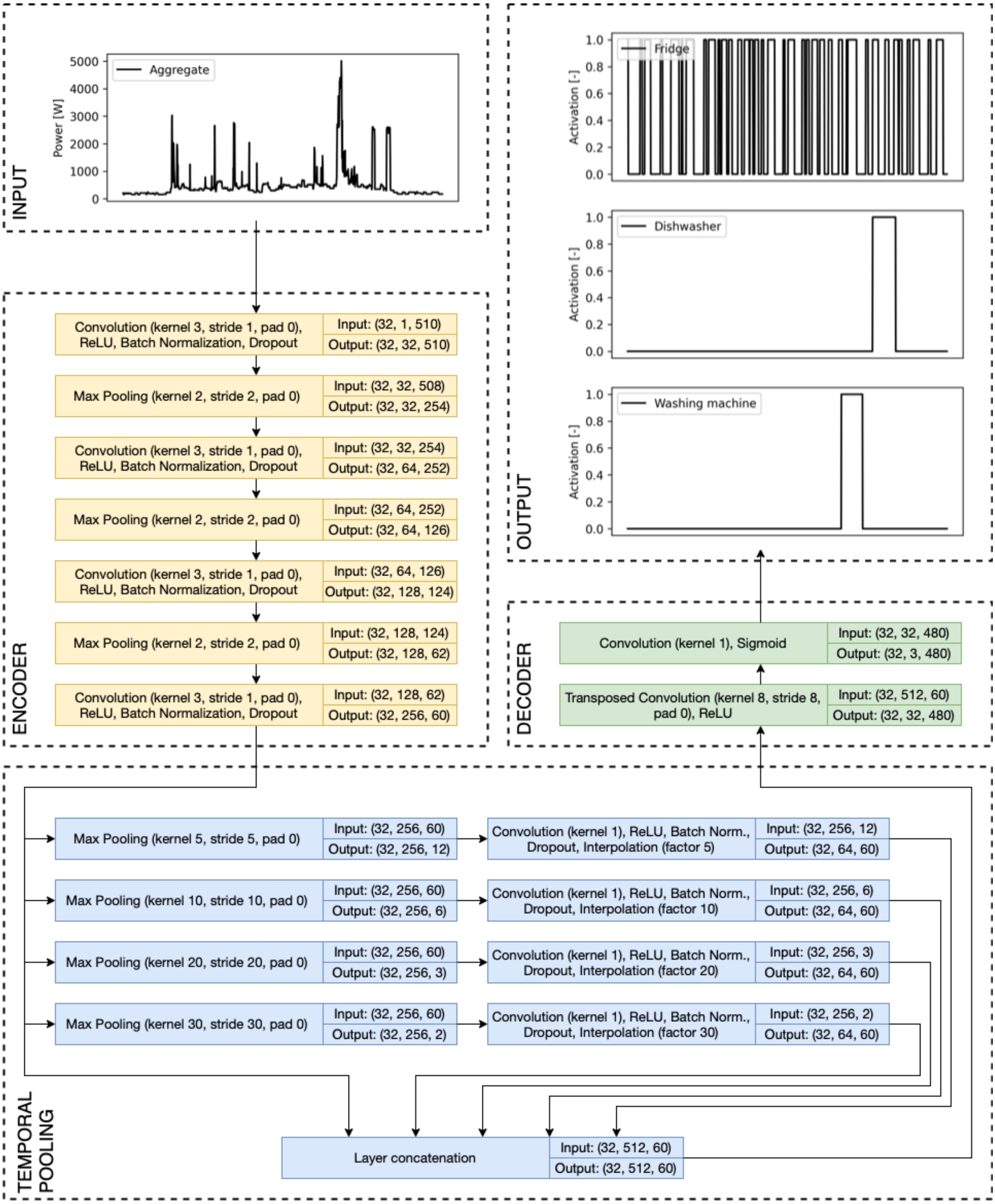

Figure 1. Outline of the network architecture used

The transmission of data occurs both synchronously and asynchronously; these types of transmission are used for communication between the smart meter and the in-home device enabling the user to monitor consumption. In the first type the power signal is sent every 15 minutes to the device, while in the second one the last active power sample is sent to the device when fixed thresholds are exceeded. The operation of sample acquisition by the in-home device is described in detail in standard [42].

In Figure 2 the signal of the power measurements at 1 min intervals is shown together with the synchronous signal, sent every quarter of an hour, and with the asynchronous signal, sent each time the power crosses a pre-defined thresholds, uniformly distributed with a constant step equal to 500 W for the example. The data volume to be communicated is clearly lower with respect to the power measurements at a constant sampling rate.

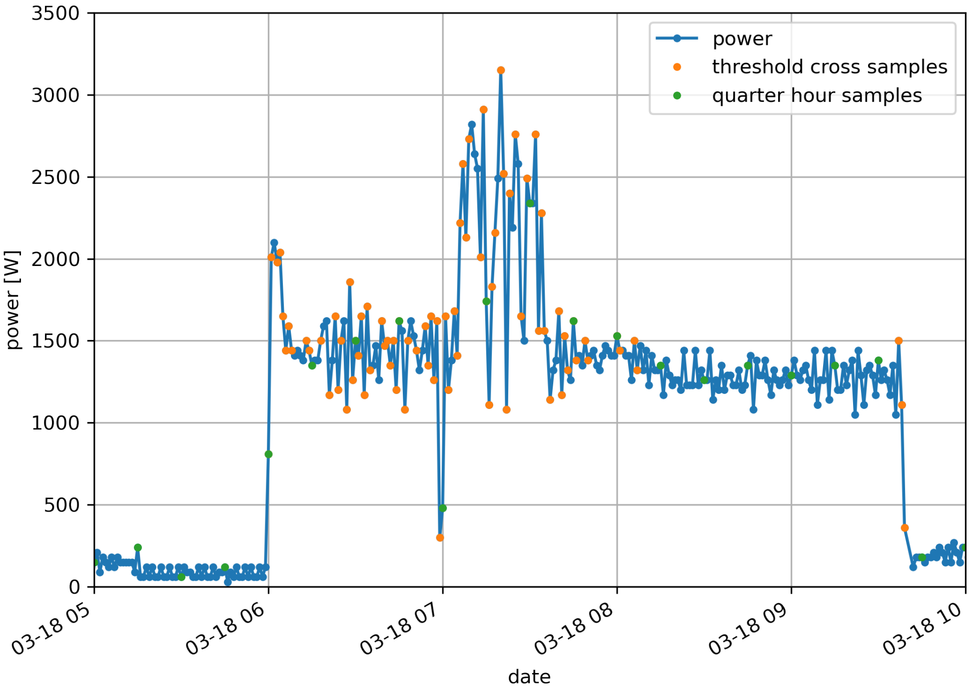

Figure 2. *Chain 2* filtering on load signal with power thresholds at 500 W intervals. Blue dots represent the load measurements with 1 min sampling rate, green dots are the power measurements at the constant rate of 15 min, orange dots are the power measurements sent to the user each time the power absorbed crosses a threshold, in this case the thresholds are set to [500, 1000, ..., 3500].

3. Experimental setup

We apply the TP-NILM architecture to two reference datasets, the UK-DALE dataset [43] and the REFIT dataset [44], to evaluate the accuracy in the disaggregation of an always-on appliance, such as the refrigerator, and two household appliances whose activation can be deferred, such as dishwasher and washing machine, which are therefore interesting in a perspective of energy management. The aggregate load for these dataset is available at a constant sampling rate. The time series is filtered to simulated a *Chain 2* signal. A time series with a constant sampling rate of 1 min is then reconstructed from the filtered signal and used as the input for the neural network. We then test the effectiveness of the disaggregation for different values of the preset thresholds for the *Chain 2* filter.

3.1. *UK-DALE dataset*

The UK-DALE dataset contains the aggregate and individual appliance signals from 5 different households in the United Kingdom. All aggregate loads were sampled at a frequency of 1Hz, unlike the individual appliance signals where the frequency was 1/6Hz.

The duration of the recordings is not the same in all the houses in the dataset, and the appliances in each house are also different. For this reason, after a careful analysis of the dataset, it was decided to use only houses 1, 2 and 5 for the purpose of the experiment being the only houses equipped with all the appliances considered.

3.2. *REFIT dataset*

Similarly to the previous case, the REFIT dataset contains the aggregate and individual appliance electricity consumption of 20 homes in United Kingdom. In this case the data is collected with a uniform frequency of 8 Hz for both aggregate and individual appliance consumption.

The timeline is not the same for all the houses in the dataset, but the records manage to cover a time span of two years. The richness of the data therefore allowed a considered choice of houses to be used for the experiments proposed in this work. A total of 15 houses were used, but unlike the previous case they were divided among the appliances of interest. The latter are however the same as those chosen for the UK-DALE case, i.e. fridge, dishwasher and washing machine.

3.3. *Chain 2 filtering*

The power measurements for the *Chain 2* signal useful for disaggregation purposes are of two types: an instantaneous load signal emitted at a constant rate every 15 min, and a signal in which the value of instantaneous power is transmitted whenever the power itself crosses fixed power values. For the second signal the possible values of the instantaneous load are divided into bins of fixed width (300 W is the default setting for the smart meter). Every time the instantaneous consumption belongs to a different bin than the last value communicated, the measurement of instantaneous power is transmitted.

Algorithm 3.1 Calculate *Chain 2* signal

```
Input: aggregate, threshold
Output: filtered_aggregate
filtered_aggregate=aggregate
l=0
for i=0:length(filtered_aggregate) do
  if ⌊filtered_aggregate[i] ÷ threshold⌋ ≠ l then
    l= ⌊filtered_aggregate[i] ÷ threshold⌋
  else
    filtered_aggregate[i]=NaN
  end if
end for
filtered_aggregate[every 15 min]=aggregate[every 15 min]
```

The aggregate consumption signal for households is filtered to reproduce a *Chain 2* type signal according to the algorithm 3.1.

The threshold value, i.e., bin width, is a smart meter parameter configurable by the utility, and clearly influences the amount of data transmitted and the richness of the signal useful for a disaggregation algorithm.

Figures 3 and 4 show the data volume of aggregated, filtered and unfiltered, signals from some houses used in this work in the UK-DALE and REFIT case respectively. It is evident from the figures that in the cases examined, the threshold value has a limited influence on the overall volume of data, especially when compared to the amount of unfiltered signal data.

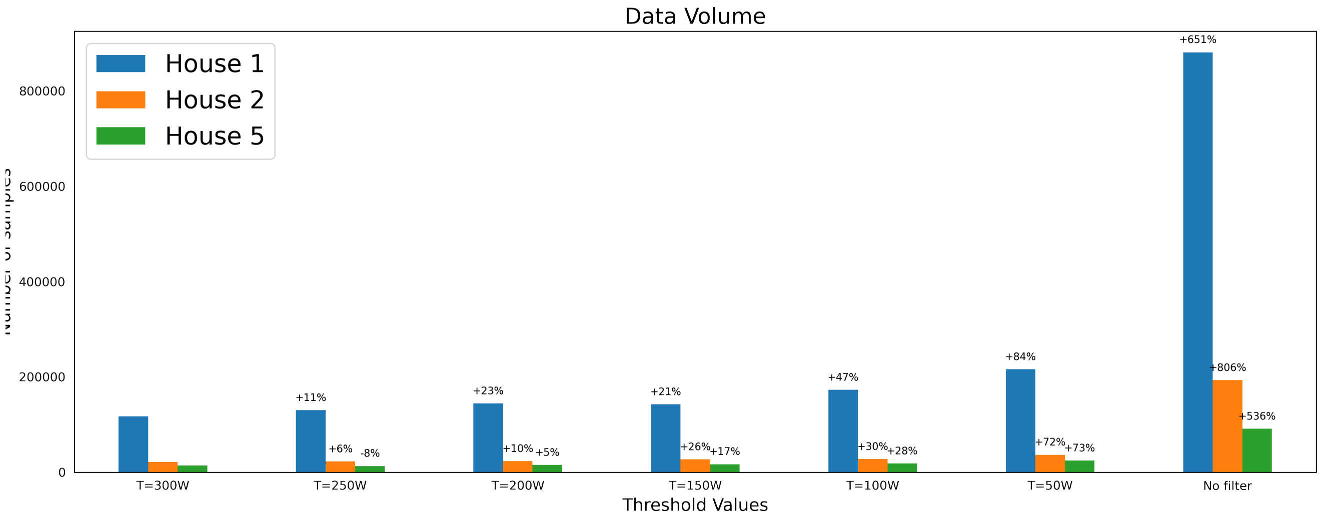

Figure 3. Number of samples for different threshold values of the *Chain 2* filter in UK-DALE dataset. A logaritmic scale was used and the percentages represent the increase in the number of samples compared to the number obtained with the default threshold value.

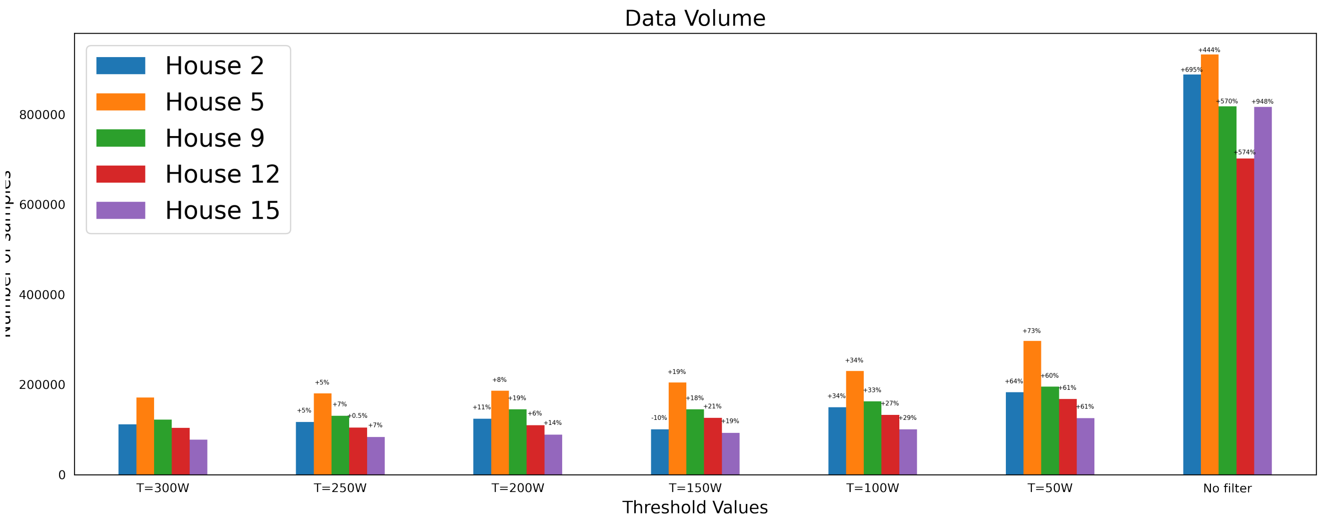

Figure 4. Number of samples for different threshold values of the *Chain 2* filter in REFIT dataset. A logaritmic scale was used and the percentages represent the increase in the number of samples compared to the number obtained with the default threshold value.

3.4. *Preprocessing*

Before training and using the neural network, the data from the two datasets were subjected to a pre-processing phase. As a first step, the power values of each appliance were extracted and sub-sampled at a frequency of 1/60 Hz, using the average power value in each time interval considered. Consequently, the aggregated signal was also resampled with the same approach and using the same frequency value.

The aggregated signal was then filtered to generate a *Chain 2* signal, and the output of the filtering was resampled over a period of 1 minute, propagating the last valid observation forward to the next valid one, in order to make it homogeneous and usable with the network.

The activation status of the individual appliances was then calulated with a threshold method, similarly to [23], since it appears to provide good results in terms of accuracy for both datasets [45]: each appliance is considered to be in operation when a certain threshold is exceeded for a fixed time interval, the Table 1 describes the thresholds and time intervals chosen, for UK-DALE dataset, taken from [23] and Table 2 the values for REFIT dataset as proposed in [45].

Table 1. Parameters used to derive the activation status of household appliances from measurements of the absorbed power for UK-DALE

	Fridge	**Dishwasher**	**Washing Machine**
Max. power (W)	300	2500	2500
Power threshold (W)	50	10	20
OFF duration (min)	0	3	30
ON duration (min)	1	30	30

Table 2. Parameters used to derive the activation status of household appliances from measurements of the absorbed power for REFIT

	Fridge	**Dishwasher**	**Washing Machine**
Max. power (W)	300	2500	2500
Power threshold (W)	20	30	20
OFF duration (min)	1	27	13
ON duration (min)	10	23	27

To facilitate convergence and lower the computational cost, the data were normalised by dividing each power value by a fixed value of 2000W. In addition, the time sequence of values was divided into windows of length 512 minutes, used as input to the neural network, which provides an output of 8 hours (480 minutes). The 32 minutes of difference between input and output constitute a leading and a trailing edge of 16 minutes each, due to the convolution filters for which no padding was applied.

Finally, in order for the input values to have zero mean, the average power value in each considered interval is subtracted from the signal values.

3.5. *Training and Testing*

The techniques described so far have the main goal of predicting the activation status of three fixed appliances, fridge, washing machine and dishwasher, by knowing the aggregate signal of a household. For this purpose the Convolutional Neural Network shown in Figure 1 was trained.

To evaluate the effectiveness of the proposed methodology, the network was tested on the two datasets described above, UK-DALE and REFIT.

In both cases, the network was tested on houses not contained in the training set. In this way, it is possible to assess the generalisation capacity of the model presented, i.e. the ability of the network to disaggregate the consumption of appliances in house that do not belong to the training set and for which the time series of the appliance consumption are used for the sole purpose of performance assessment.

Regarding the UK-DALE dataset, the network was trained using the training portion of House 1 and House 5 and it was tested using the entire time series of House 2. Table 3 summarizes which portions of the dataset were used for the training, validation and testing phases.

Table 3. Training, Validation and Testing dataset composition for the UK-DALE dataset case.

	Training	**Validation**	**Testing**
House 1	80 %	20 %	-
House 2	-	-	100 %
House 5	80 %	20 %	-

Concenrning the REFIT dataset, in accordance with [28], it was decided to use different houses for each appliance in the different phases of training, validation and testing. Table 4 summarizes which houses were chosen.

Table 4. Training, Validation and Testing dataset composition for the REFIT dataset case.

	Training (80%) and Validation (20%)	**Testing**
Fridge	2, 5, 9, 12	15
Dishwasher	1, 3, 5, 6, 7, 9, 10, 11, 13, 15, 16 18, 20	2
Washing Machine	1, 2, 3, 5, 7, 8, 9, 10, 11, 13, 15, 16, 17, 18, 19, 20, 21	6

Note that during the training phase the input signal was not filtered with the *Chain 2* protocol.

In both cases the Adam optimisation algorithm was used to optimise the parameters of the neural network with a gradient descent approach, the learning rate was set at $5 \cdot 10^{-5}$ and the batch size at 32. In order to avoid the overfitting an early stopping strategy was adopted.The parameters described were not the only ones used, but they were the ones that gave good results in terms of convergence times and accuracy obtained.

3.6. *Postprocessing*

The output of the network is an array of values that varies in the range $(0, 1)$ and represents the probability that the appliance is turned on. The appliance is rated as switched on if the probability exceeds an arbitrated threshold which has been set at 0.5. An estimate of the power consumption related to the appliance is also calculated, it is a constant value equal to the average power consumption of each appliance multiplied by the duration of its operation.

3.7. *Performance evaluation*

The performance of the network was evaluated both for identifying the activation status of each device and for estimating power consumption. Several metrics were used for this purpose.

Let $a_i(t)$ be the activation state of the i-th appliance at time t, which is equal to 1 if the appliance is on or 0 if it is off, and let $y_i(t)$ the value of the power consumed at the same time. Then we can denote with $\hat{a}_i(t)$ and $\hat{y}_i(t)$ the model predictions of the quantities described above. The metrics for evaluating the model's predictions refer to a time series t_k with $k \in [0, N_s)$

We can also define by True Positive (TP) the number of times the model correctly predicts the activation state of an appliance when it is switched on, by True Negative (TN) the number of times the appliance is correctly evaluated as switched off. While we can define with False Positive (FP) the number of instances in which the appliance is evaluated as on when it is off and with False Negative (FN), on the contrary, the number of times it is evaluated as off despite being on.

Then we can define the following metrics.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F_1 = 2\frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{9}$$

As can be seen, the Equation 5 of Accuracy represents the ratio of the number of correctly evaluated instances to the total number of predicted instances. On the other hand Precision, in Equation 6, is defined as the ratio between TP and the total number of times the appliance is evaluated to be switched on, while Recall, as described in Equation 7, defines the ratio between the number of times the appliance is rated in operation and the total number of times it is actually in operation. The F1 measurement defined in Equation 8 is the weighted average of Precision and Recall, varying in a range between $[0, 1]$ and high values of this score indicate a better ability to identify the state of the appliance. Finally the Matthews Correlation Coefficient is defined in Equation 9, whose values are in the range $[-1, 1]$. A value equal to 1 denotes an exact classification, 0 denotes a random prediction while a value equal to -1 indicates a totally wrong classification.

Note that the metrics defined so far have been used to assess performance in terms of predicting the activation state of an appliance, thus relating to a classification problem. Regarding the estimation of the energy consumed, the accuracy of the proposed methodology has been measured in terms of Mean Absolute Error (MAE) and Signal Aggregate Error (SAE), defined by the equations below.

$$MAE_i = \frac{1}{N_s} \sum_j \hat{y}_i(t_j) - y_i(t_j) \tag{10}$$

$$SAE_i = \frac{\sum_j \hat{y}_i(t_j) - \sum_j y_i(t_j)}{\sum_j y_i(t_j)} \tag{11}$$

MAE, defined in Equation 10, measures how much on average the estimated power deviates from the measured power at each instant, while SAE, defined in Equation 11, estimates the relative error of the predicted energy over the entire time interval considered.

The methodology we propose estimates instantaneous consumption by assuming a constant value during the appliance's activation cycle; this is a good approximation in the case of appliances with on/off operation such as the fridge, while it will result in higher values for MAE in the case of appliances with variable absorption during their activation, such as the dishwasher and washing machine. The MAE metric is of little significance to us; we report it to allow easier comparison with other methods proposed in the literature.

4. Results

The performance of the TP-NILM network was examined for the two UK-DALE and REFIT datasets, evaluating its accuracy both on the original signal and on the one filtered with the *Chain 2* protocol with different threshold values. The results for the three appliances considered for the UK-DALE dataset are shown in Table 5, Table 6 reports the performance obtained on the REFIT dataset.

In both datasets it is possible to see that the accuracy depends on the value of the threshold chosen for filtering. The maximum levels of accuracy are reached with the unfiltered signal, as expected, however filtering according to the *Chain 2* scheme still allows an effective disaggregation as long as the threshold values are not too high.

The results show that the proposed approach enables to obtain a very accurate disaggregation of the consumption of the appliances considered for both datasets, even using a low frequency sampling and a *Chain 2* type signal. The threshold, set for the event based component of the signal, is also a threshold for the appliances that can be disaggregated, an appliance that has a power consumption, when turned on, lower than or close to the threshold value cannot be disaggregated, because the consumption signal

Table 5. TP-NILM performances on House 2 of the *UK-DALE* dataset. NF (No Filter) represents network performances using the original aggregate signal.

	Prec.	Rec.	Acc.	F1	MCC	MAE	SAE
Fridge							
Krystalakos et al. [27] GRU	0.46	0.75	0.60	0.57	-	51	0.26
Krystalakos et al. [27] s2p	0.42	0.74	0.54	0.53	-	51	0.29
Yue et al. [32]	-	-	0.81	0.77	-	25.49	-
Rafiq et al. [46]	-	-	-	0.87	19.61	0.467	
Zhou et al. [47]	0.75	0.73	-	0.74	-	39.33	-
Song et al. [28]	-	-	-	0.94	-	22.35	0.12
Piccialli et al. [33]	-	-	-	0.87	-	13.24	-
Puente et al. [48]	0.97	0.96	-	0.97	-	-	-
TP-NILM No filter	0.87	0.89	0.90	0.89	0.79	17.35	−0.05
TP-NILM T=50W	0.87	0.83	0.89	0.85	0.76	18.36	−0.05
TP-NILM T=100W	0.70	0.70	0.77	0.70	0.52	27.67	0.01
TP-NILM T=150W	0.78	0.76	0.83	0.77	0.64	23.07	−0.03
TP-NILM T=200W.	0.67	0.68	0.75	0.67	0.47	29.51	0.02
TP-NILM T=250W	0.67	0.66	0.75	0.67	0.46	29.68	−0.01
TP-NILM T=300W	0.64	0.63	0.73	0.63	0.41	31.60	−0.01
Dishwasher							
Krystalakos et al. [27] GRU	0.62	0.42	0.97	0.50	-	24	0.07
Krystalakos et al. [27] s2p	0.47	0.43	0.96	0.45	-	21	0.07
Yue et al. [32]	-	-	0.97	0.67	-	16.18	-
Rafiq et al. [46]	-	-	-	0.81	15.27	0.323	
Zhou et al. [47]	0.88	0.86	-	0.87	-	118.1	-
Song et al. [28]	-	-	-	0.87	-	20.95	0.69
Piccialli et al. [33]	-	-	-	0.72	-	7.26	-
Puente et al. [48]	0.73	0.90	-	0.81	-	-	-
TP-NILM No filter	0.82	0.85	0.99	0.83	0.83	30.95	0.04
TP-NILM T=50W	0.81	0.87	0.99	0.84	0.84	30.89	0.07
TP-NILM T=100W	0.81	0.85	0.99	0.83	0.82	31.08	0.05
TP-NILM T=150W	0.82	0.86	0.99	0.84	0.83	30.82	0.05
TP-NILM T=200W	0.81	0.85	0.99	0.83	0.82	30.56	0.05
TP-NILM T=250W	0.79	0.87	0.99	0.83	0.83	31.63	0.10
TP-NILM T=300W	0.80	0.82	0.99	0.81	0.80	31.45	0.02
Washing machine							
Krystalakos et al. [27] GRU	0.22	0.54	0.96	0.31	-	30	0.58
Krystalakos et al. [27] s2p	0.26	0.55	0.97	0.35	-	17	0.28
Yue et al. [32]	-	-	0.97	0.33	-	6.98	-
Rafiq et al. [46]	-	-	-	0.77	14.42	0.512	
Zhou et al. [47]	0.74	0.99	-	0.85	-	55.90	-
Song et al. [28]	-	-	-	0.88	-	12.27	0.26
Piccialli et al. [33]	-	-	-	0.69	-	6.57	-
Puente et al. [48]	0.43	0.40	-	0.41	-	-	-
TP-NILM No filter	0.76	0.92	0.99	0.83	0.84	9.57	0.21
TP-NILM T=50W	0.78	0.87	0.99	0.82	0.82	9.24	0.12
TP-NILM T=100W	0.82	0.81	0.99	0.82	0.82	8.64	−0.01
TP-NILM T=150W	0.89	0.65	0.99	0.75	0.75	7.88	−0.28
TP-NILM T=200W	0.88	0.53	0.99	0.66	0.68	7.89	−0.40
TP-NILM T=250W	0.91	0.41	0.99	0.56	0.61	7.43	−0.55
TP-NILM T=300W	0.92	0.20	0.99	0.32	0.43	7.15	−0.78

generated by it is filtered in the processing of the *Chain 2* signal. In fact, it can be seen that the accuracy in the disaggregation of the fridge, falls for threshold values higher than 50W, since the absorption of the

Table 6. TP-NILM performances on House 15 (*Fridge* case), House 2 (*Dishwasher* case) and House 6 (*Washing Machine* case) of the *REFIT* dataset.

	Prec.	**Rec.**	**Acc.**	**F1**	**MCC**	**MAE**	**SAE**
			Fridge				
Song et al. [28]	-	-	-	0.95	-	20.15	0.13
Pan et al. [49]	-	-	-	-	-	16.77	0.10
Murray et al. [24] CNN	-	-	0.77	0.93	-	8.56	-
Murray et al. [24] GRU	-	-	0.64	0.85	-	13.30	-
D'Incecco et al. [25] CNN	-	-	-	-	-	20.02	0.33
TP-NILM No filter	0.81	0.79	0.89	0.80	0.73	9.37	−0.02
TP-NILM T=50W	0.69	0.71	0.83	0.70	0.58	14.05	0.03
TP-NILM T=100W	0.61	0.64	0.78	0.62	0.48	17.33	0.06
TP-NILM T=150W	0.61	0.64	0.79	0.63	0.48	17.21	0.05
TP-NILM T=200W.	0.59	0.63	0.78	0.61	0.45	18.15	0.07
TP-NILM T=250W	0.59	0.63	0.78	0.61	0.45	18.06	0.06
TP-NILM T=300W	0.57	0.60	0.76	0.58	0.42	19.08	0.06
			Dishwasher				
	Prec.	**Rec.**	**Acc.**	**F1**	**MCC**	**MAE**	**SAE**
Jiang et al. [26]	-	-	-	0.60	-	17.66	-
Cimen et al. [29]	-	-	-	0.78	-	13.35	-
Song et al. [28]	-	-	-	0.88	-	12.34	0.60
Pan et al. [49]	-	-	-	-	-	4.80	0.17
Murray et al. [24] CNN	-	-	0.83	0.82	-	82.74	-
Murray et al. [24] GRU	-	-	0.85	0.82	-	73.53	-
D'Incecco et al. [25] CNN	-	-	-	-	-	12.26	0.26
TP-NILM No filter	0.79	0.86	0.98	0.82	0.81	61.91	0.09
TP-NILM T=50W	0.79	0.86	0.98	0.82	0.81	62.18	0.09
TP-NILM T=100W	0.78	0.86	0.98	0.82	0.81	62.39	0.10
TP-NILM T=150W	0.78	0.86	0.98	0.82	0.81	62.70	0.11
TP-NILM T=200W	0.78	0.86	0.98	0.82	0.81	62.97	0.11
TP-NILM T=250W	0.78	0.86	0.98	0.82	0.81	62.98	0.12
TP-NILM T=300W	0.77	0.87	0.98	0.82	0.80	63.38	0.12
			Washing machine				
	Prec.	**Rec.**	**Acc.**	**F1**	**MCC**	**MAE**	**SAE**
Jiang et al. [26]	-	-	-	0.75	-	3.91	-
Cimen et al. [29]	-	-	-	0.94	-	8.70	-
Song et al. [28]	-	-	-	0.89	-	8.88	0.22
Pan et al. [49]	-	-	-	-	-	5.88	0.13
Murray et al. [24] CNN	-	-	0.72	0.79	-	71.99	-
Murray et al. [24] GRU	-	-	0.69	0.86	-	79.33	-
D'Incecco et al. [25] CNN	-	-	-	-	-	16.85	2.61
TP-NILM No filter	0.79	0.89	0.99	0.84	0.84	8.54	0.13
TP-NILM T=50W	0.87	0.89	0.99	0.88	0.87	7.96	0.03
TP-NILM T=100W	0.90	0.87	0.99	0.89	0.88	7.66	−0.03
TP-NILM T=150W	0.91	0.79	0.99	0.85	0.85	7.38	−0.13
TP-NILM T=200W	0.92	0.66	0.99	0.77	0.78	6.99	−0.28
TP-NILM T=250W	0.93	0.54	0.99	0.69	0.71	6.74	−0.42
TP-NILM T=300W	0.95	0.40	0.99	0.57	0.62	6.36	−0.58

same, when the compressor is active, is slightly higher than 100W, typically. For the dishwasher, in both datasets, maximum accuracy can be obtained for threshold values up to 300W. For the washing machine, in both datasets, the maximum tolerable threshold is 100W; a threshold of 300W results in a significant deterioration of the performance of the proposed algorithm. Among the various metrics proposed we note the good values for the estimate of the activation state, the MAE has instead modest values, as it was deliberately chosen not to have an exact estimate of the instantaneous absorption as this information has little practical relevance. Much more important is the estimate of the total energy absorbed, whose accuracy is measured by the SAE has excellent values for the three appliances for thresholds up to 100W.

Therefore, the *Chain 2* signal can be directly used for load disaggregation as long as the chosen threshold is commensurate with the power level of the household appliance, and disaggregation can then be realized directly from smart meter measurements without the need to install additional measurement hardware.

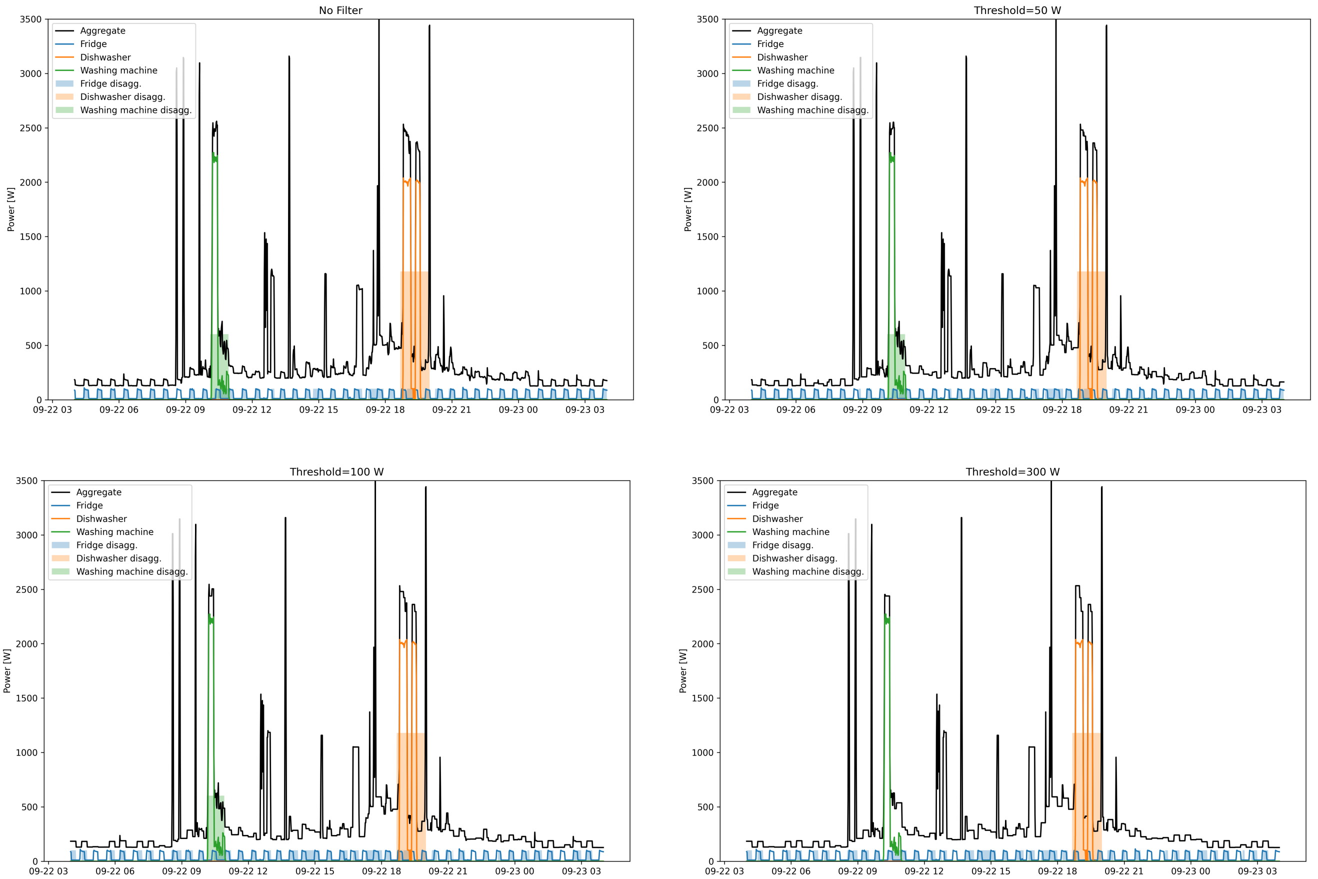

Figure 5. Example of load disaggregation for the UK-DALE dataset case. The graphs show the aggregate load and the load of each appliance as solid lines while the disaggregated estimate of the loads is represented by a shaded area. The first plot represents the estimate for the original (unfiltered) input signal while the other three are related to the threshold values chosen to perform the experiments.

Figure 5 shows an example application of the network for the UK-DALE case, the portion of the load signal was chosen so that the three appliances examined were simultaneously activated. The figures show the aggregate signals and actual consumption of the appliances superimposed on the estimate of activations and average consumption obtained with the proposed algorithm. As said, the algorithm allows to obtain a direct estimate of the activation status starting from the aggregate signal only, eventually filtered with the Chain 2 protocol. The estimate of the absorbed power is obtained, as said, considering a constant average power for the activation period, as is also evident from the graphical representation. In this example we can see the degradation of the disaggregation estimate for the refrigerator already for low thresholds, the good accuracy in the disaggregation of the dishwasher, independent of the threshold value adopted, and the sudden degradation in the estimation of the absorption of the washing machine for thresholds above 100W.

Tables 5 and 6 also report the results obtained from the most recent works in the literature that have examined the two datasets. This is not a fair comparison, as all the cited methods use the signal at the maximum sampling rate, which would not be possible to obtain from a smart meter, and therefore require additional measurement hardware for their application. Nevertheless, the performance of our proposed architecture is aligned with the performances of the best results in the recent literature.

In the comparison with the literature, our architecture obtains the worst performance according to the F1 metric in the case of the refrigerator, for both datasets. The reason is related precisely to the sampling of the signal, which we recall, even in the unfiltered case is limited to a rate of 1 min in our case, this signal does not allow the network to effectively detect the initial transient due to the compressor switching on. However, it should be noted that the performance in terms of SAE for the refrigerator are nonetheless excellent. The value of F1 is lower than the best results in the literature also in the case of the washing machine in the REFIT dataset, which is compensated by the excellent performance in terms of SAE in this case as well.

5. Conclusions

In this work we have verified the possibility to realize a disaggregation of the loads of some household appliances of interest, starting from the signal that is directly provided by the second generation smart meters deployed in Italy, without the need of additional hardware for the measurement of the aggregated load, therefore with a fully non-intrusive procedure. We applied a deep learning methodology based on convolutional neural networks to two reference datasets, for which we filtered the aggregate load signal simulating the *Chain 2* protocol, used in smart meters. The results show that with a reasonable choice of filter parameters an accurate disaggregation is achieved with a significant reduction in the volume of data transmitted, thus realizing an effectively non-intrusive consumption monitoring, feasible both on user and utility side. These results also highlight how rich in information the electrical signal of consumption can be, and emphasise the need for special attention in the protection of sensitive information that can be derived from it. The limitations we see are typical of supervised approaches, which are highly dependent on the available data. The acquisition of labeled datasets is certainly expensive, but in principle the algorithm proves to be valid, and it is possible to think of its scalability when more complete datasets were available to facilitate the training of models.

Author contributions

Conceptualization, L.M.; methodology, M.M. and L.M.; software, M.M. and L.M.; validation, M.M. and L.M.; formal analysis, L.M.; investigation, M.M.; resources, L.M.; data curation, M.M.; writing—original draft preparation, M.M.; writing—review and editing, L.M.; visualization, M.M.; supervision, L.M.; project administration, L.M.; funding acquisition, L.M.

All authors have read and agreed to the published version of the manuscript.

Funding

The research for this article was carried out partly within the PhD project of Manolo Manca funded by CRS4 (Università di Cagliari, XXXV Ciclo, Decreto Rettorale 837/2019 del 23/07/2019) and partly within the project Tessuto Digitale Metropolitano funded by Regione Autonoma della Sardegna, Delibera 66/14 del 13.12.2016, Progetto Complesso area ICT.

Supplementary material

The software and the data with which the results of the paper can be reproduced is available in the GitHub repository https://github.com/lmssdd/nilm_chain2

References

1. G. F. Angelis, C. Timplalexis, S. Krinidis, D. Ioannidis, and D. Tzovaras, Nilm applications: Literature review of learning approaches, recent developments and challenges, *Energy and Buildings*, p. 111951, 2022.

2. A. R. Al Ali, I. A. Zualkernan, M. Rashid, R. Gupta, and M. Alikarar, A smart home energy management system using iot and big data analytics approach, *IEEE Transactions on Consumer Electronics*, vol. 63, no. 4, pp. 426–434, 2017.
3. D. M. Han and J. H. Lim, Smart home energy management system using ieee 802.15.4 and zigbee, *IEEE Transactions on Consumer Electronics*, vol. 56, no. 3, pp. 1403–1410, 2010.
4. F. M. Wittmann, J. C. López, and M. J. Rider, Nonintrusive load monitoring algorithm using mixed-integer linear programming, *IEEE Transactions on Consumer Electronics*, vol. 64, no. 2, pp. 180–187, 2018.
5. K. C. Armel, A. Gupta, G. Shrimali, and A. Albert, Is disaggregation the holy grail of energy efficiency? the case of electricity, *Energy Policy*, vol. 52, pp. 213–234, 2013.
6. G. W. Hart, Nonintrusive appliance load monitoring, *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.
7. E. Elhamifar and S. Sastry, Energy disaggregation via learning powerlets and sparse coding., in *AAAI*, pp. 629–635, AAAI, 2015.
8. S. Gupta, M. S. Reynolds, and S. N. Patel, Electrisense: single-point sensing using emi for electrical event detection and classification in the home, in *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pp. 139–148, ACM, 2010.
9. G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and R. Cepeda, Privacy for smart meters: Towards undetectable appliance load signatures, in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pp. 232–237, IEEE, 2010.
10. A. Prudenzi, A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recordings at meter panel, in *Power Engineering Society Winter Meeting, 2002. IEEE*, vol. 2, pp. 941–946, IEEE, 2002.
11. K. Basu, V. Debusschere, A. Douzal-Chouakria, and S. Bacha, Time series distance-based methods for non-intrusive load monitoring in residential buildings, *Energy and Buildings*, vol. 96, pp. 109–117, 2015.
12. K. Basu, V. Debusschere, S. Bacha, U. Maulik, and S. Bondyopadhyay, Nonintrusive load monitoring: A temporal multilabel classification approach, *IEEE Transactions on Industrial informatics*, vol. 11, no. 1, pp. 262–270, 2015.
13. 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*, pp. 747–758, SIAM, 2011.
14. 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*, AAAI, 2012.
15. A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. E. Rizzoli, A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring, *Applied energy*, vol. 185, pp. 331–344, 2017.
16. L. Massidda and M. Marrocu, A bayesian approach to unsupervised, non-intrusive load disaggregation, *Sensors*, vol. 22, no. 12, p. 4481, 2022.
17. F. Hidiyanto and A. Halim, Knn methods with varied k, distance and training data to disaggregate nilm with similar load characteristic, in *Proceedings of the 3rd Asia Pacific Conference on Research in Industrial and Systems Engineering 2020*, pp. 93–99, ACM, 2020.
18. M. Singh, S. Kumar, S. Semwal, and R. Prasad, Residential load signature analysis for their segregation using wavelet - svm, in *Power Electronics and Renewable Energy Systems*, pp. 863–871, Springer, 2015.
19. F. Gong, N. Han, Y. Zhou, S. Chen, D. Li, and S. Tian, A svm optimized by particle swarm optimization approach to load disaggregation in non-intrusive load monitoring in smart homes, in *2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2)*, pp. 1793–1797, IEEE, 2019.
20. M. Hasan, D. Chowdhury, M. Khan, Z. Rahman, *et al.*, Non-intrusive load monitoring using current shapelets, *Applied Sciences*, vol. 9, no. 24, p. 5363, 2019.

21. Z. Xiao, W. Gang, J. Yuan, Y. Zhang, and C. Fan, Cooling load disaggregation using a nilm method based on random forest for smart buildings, *Sustainable Cities and Society*, vol. 74, p. 103202, 2021.

22. X. Wu, Y. Gao, and D. Jiao, Multi-label classification based on random forest algorithm for non-intrusive load monitoring system, *Processes*, vol. 7, no. 6, p. 337, 2019.

23. 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*, pp. 55–64, ACM, 2015.

24. 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)*, pp. 8330–8334, IEEE, 2019.

25. M. D'Incecco, S. Squartini, and M. Zhong, Transfer learning for non-intrusive load monitoring, *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1419–1429, 2019.

26. J. Jiang, Q. Kong, M. D. Plumbley, N. Gilbert, M. Hoogendoorn, and D. M. Roijers, Deep learning-based energy disaggregation and on/off detection of household appliances, *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 15, no. 3, pp. 1–21, 2021.

27. O. Krystalakos, C. Nalmpantis, and D. Vrakas, Sliding window approach for online energy disaggregation using artificial neural networks, in *Proceedings of the 10th Hellenic Conference on Artificial Intelligence*, pp. 1–6, ACM, 2018.

28. J. Song, H. Wang, M. Du, L. Peng, S. Zhang, and G. Xu, Non-intrusive load identification method based on improved long short term memory network, *Energies*, vol. 14, no. 3, p. 684, 2021.

29. H. Çimen, N. Çetinkaya, J. C. Vasquez, and J. M. Guerrero, A microgrid energy management system based on non-intrusive load monitoring via multitask learning, *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 977–987, 2020.

30. M. Valenti, R. Bonfigli, E. Principi, and S. Squartini, Exploiting the reactive power in deep neural models for non-intrusive load monitoring, in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2018.

31. A. Faustine, L. Pereira, H. Bousbiat, and S. Kulkarni, Unet-nilm: A deep neural network for multi-tasks appliances state detection and power estimation in nilm, in *Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring*, pp. 84–88, ACM, 2020.

32. Z. Yue, C. R. Witzig, D. Jorde, and H. A. Jacobsen, Bert4nilm: A bidirectional transformer model for non-intrusive load monitoring, in *Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring*, pp. 89–93, ACM, 2020.

33. V. Piccialli and A. M. Sudoso, Improving non-intrusive load disaggregation through an attention-based deep neural network, *Energies*, vol. 14, no. 4, p. 847, 2021.

34. L. Massidda, M. Marrocu, and S. Manca, Non-intrusive load disaggregation by convolutional neural network and multilabel classification, *Applied Sciences*, vol. 10, no. 4, p. 1454, 2020.

35. L. Massidda, M. Marrocu, and S. Manca, Non-intrusive load disaggregation via a fully convolutional neural network: improving the accuracy on unseen household, in *2020 2nd IEEE International Conference on Industrial Electronics for Sustainable Energy Systems (IESES)*, vol. 1, pp. 317–322, IEEE, 2020.

36. R. Terracciano, V. Galdi, V. Calderaro, D. Pappalardo, G. Ceneri, and A. O. Piti, Demand side management services for smart buildings with the use of second generation smart meter and the chain-2 of e-distribuzione, in *2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I CPS Europe)*, pp. 1–6, IEEE, 2020.

37. D. Serra, D. Mardero, L. Di Stefano, and S. Grillo, Post-metering value-added services for low voltage electricity users: Lessons learned from the italian experience of chain 2, *Applied Energy*, vol. 304, p. 117806, Dec 2021.

38. H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, Pyramid scene parsing network, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2881–2890, IEEE, 2017.

39. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035, Curran Associates, Inc., 2019.

40. S. Vitiello, N. Andreadou, M. Ardelean, and G. Fulli, Smart metering roll-out in europe: Where do we stand? cost benefit analyses in the clean energy package and research trends in the green deal, *Energies*, vol. 15, no. 7, p. 2340, 2022.

41. C. Staff, Cei-en 50065-1, signalling on low-voltage electrical installations in the frequency range 3 khz to 148.5 khz part 1: General requirements, frequency bands and electromagnetic disturbances, *CEI Standards*, 2012.

42. C. Staff, Cei ts 13-82:2017-08, sistemi di misura dell'energia elettrica - comunicazione con i dispositivi utente, parte 2: Modello dati e modello applicativo., *CEI Standards*, 2012.

43. 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, p. 150007, 2015.

44. 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.

45. P. Laviron, X. Dai, B. Huquet, and T. Palpanas, Electricity demand activation extraction: From known to unknown signatures, using similarity search, in *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*, pp. 148–159, ACM, 2021.

46. H. Rafiq, X. Shi, H. Zhang, H. Li, and M. K. Ochani, A deep recurrent neural network for non-intrusive load monitoring based on multi-feature input space and post-processing, *Energies*, vol. 13, no. 9, p. 2195, 2020.

47. G. Zhou, Z. Li, M. Fu, Y. Feng, X. Wang, and C. Huang, Sequence-to-sequence load disaggregation using multiscale residual neural network, *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2020.

48. C. Puente, R. Palacios, Y. González-Arechavala, and E. F. Sánchez-Úbeda, Non-intrusive load monitoring (nilm) for energy disaggregation using soft computing techniques, *Energies*, vol. 13, no. 12, p. 3117, 2020.

49. Y. Pan, K. Liu, Z. Shen, X. Cai, and Z. Jia, Sequence-to-subsequence learning with conditional gan for power disaggregation, in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3202–3206, IEEE, 2020.