

A Secure Edge Monitoring Approach to Unsupervised Energy Disaggregation using Mean Shift Algorithm in Residential Buildings

Qi Liu^{a,b,c,d,1}, Francis Mawuli Nakoty^{b,1}, Xueyan Wu^{b,c}, Raphael Anaadumba^b, Xiaodong Liu^e, Yonghong Zhang^f, Lianyong Qi^{g,*}

^a*Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAET), Nanjing University of Information Science and Technology, No. 219, Ningliu Road, Nanjing, 210044*

^b*School of Computer and Software, Nanjing University of Information Science and Technology, No. 219, Ningliu Road, Nanjing, 210044, China*

^c*Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing, China*

^d*Engineering Research Center of Digital Forensics, Ministry of Education, China*

^e*School of Computing, Edinburgh Napier University, 10 Colinton Road, Edinburgh, EH10 5DT, UK*

^f*School of Automation, Nanjing University of Information Science and Technology, Nanjing, Jiangsu 210044, China*

^g*College of Information Science and Engineering, Qufu Normal University, Rizhao, Shandong China*

Abstract

Compared to Intrusive Load Monitoring which uses smart power meters at each level to be monitored, Non-Intrusive Load Monitoring (NILM) is an ingenious way that relies on signal readings at a single point to deduce the share of the devices that have contributed to the overall load. This reliable technique that guarantees the safety and privacy of individual users has recently become an increasingly popular topic, as it turns out to be a major solution to assist household users in the process of obtaining details of their electricity consumption. The detailed consumption promotes better management of the electrical power on the consumer side by helping to eliminate any waste of energy. In this paper,

*Corresponding author

Email addresses: qi.liu@nuist.edu.cn (Qi Liu), francisnakoty@yahoo.fr (Francis Mawuli Nakoty), yexiplvos@hotmail.com (Xueyan Wu), atimanaadumba200@gmail.com (Raphael Anaadumba), x.liu@napier.ac.uk (Xiaodong Liu), zyh@nuist.edu.cn (Yonghong Zhang), lianyongqi@gmail.com (Lianyong Qi)

¹Both authors are the first author due to equal contribution to this paper.

an edge gateway has been implemented to safely monitor the overall load in a smart energy system. A load separation method has been introduced based on events detected on a low-frequency power signal, which allows the consumption profile of On/Off and multi-state devices to be generated without relying on the knowledge of the cardinality of these devices. Following the extraction of significant features contained in the aggregate signal, an appliance profile recognition approach is presented based on the non-parametric Mean Shift algorithm. The ability of the proposed method to learn and deduce devices profile is validated using the Reference Energy Disaggregation Dataset (REDD) . The experimental results show that the proposed approach is efficient in detecting events of binary state and finite state appliances.

Keywords: Home Energy Management, Non-Intrusive Load Monitoring (NILM), Unsupervised Learning, Non-parametric Algorithm

1. Introduction

Electric power comes as a primary component in the daily life of households around the world. The perfect control of its production and management is then crucial to maintain a constant availability and respond effectively to demand. In light of the continual evolution of the electrification ratio and the exponential hike in the number of household appliances within residences, the establishment of an adequate method for an efficient energy conservation stands out as a matter of high necessity. On the journey towards a clean transition to renewable energies, smart-houses implementation [1] and the installation of solar panels on household roofs to balance between traditional and green energy have seen the light of day. These ingenious approaches are all aiming at reducing consumption expenses, conservation of depletable resources such as coal and gas used in electricity production, and partially limit carbon emissions into the environment. Related statistics published in [2] show that 89% of world's total primary energy consumption is generated from fossil sources (coal, gas, nuclear, oil) versus 11% from renewable sources (hydroelectric, solar, wind and

biomass) in 2018. In spite of the efforts made, the effective implementation of these solutions remains a forward-looking goal, there is a long way to go. For the time being, energy conservation remains a major alternative as not only
20 does it eliminate any wastage due to unnecessary use, but also benefits in terms of lower electricity bills and in helping the environment by reducing pollution related to energy consumption.

Electrical Load Separation or Energy Disaggregation is the process of dividing the overall power used into individual consumption of each appliance
25 that contributed to the total load. Based on the details of single appliance's consumption, users can better plan the use of their devices [3]. The itemized power consumption of each device can be obtained by two methods: intrusive and non-intrusive mode. The installation of smart meters on the main channel as well as on all subordinate circuits of the household to directly measure the
30 transmitted electrical load is the conventional method also known as intrusive load monitoring [4]-[5]. The direct access to consumption right on the wires underlies the excellent performance of this approach. The concern lies in the cost of implementation and maintenance which increases exponentially according to the number of devices supported or the architecture of the supervision
35 system. The privacy considerations also make the applicability of this solution challenging. The non-intrusive method employs a unique power meter usually located at the household's entry point to predict the consumption of appliances used in the household. Using mathematical formulas, analysis of the graph produced by the collected measurements or artificial intelligence-based approach,
40 this sophisticated methodology attempts at producing the details of the total consumption from an installation reduced to a single power meter and a total avoidance of intrusion. Though this approach appears to be much more complex or hardly achievable, it offers the benefit of being economically feasible.

NILM was first introduced by G. W. Hart[6] under the guiding principle
45 that the total energy entering a house is equivalent to the sum of the energy utilized by all the domestic devices increased by a negligible noise quota. NILM solutions are basically categorized into two groups depending on the approach

used to implement the solution. Therefore, there are supervised solutions and unsupervised solutions[7]. As the name implies, a supervised solution makes use of additional information during the energy disaggregation process. These are methods using a probabilistic approach, classification algorithms, regression or deep neural networks methods that first try to learn the pattern followed by the consumption profile of the devices and then use it to detect the use of the device in the total load at a later stage. Although this approach is an obvious way to go, it has the disadvantage of only supporting the devices on which the experiment was carried out. A new device would thus be undetectable by such a solution. Purely unsupervised solutions are mostly based on blind source separation, clustering of detected events or graph signal processing. This type of solution is highly valuable since the number of devices used or the knowledge of their signature is not required. This makes them more robust since they have no limitation on the number of devices supported.

Due to the unrealistic nature of having some training data available to perform in-depth analysis of energy used for a real household scenario, the unsupervised NILM is the main field in which this work aims to make a valuable contribution. The proposed approach is a low-complexity event-based solution that uses a non-parametric clustering algorithm for grouping events detected on a low-frequency signal. The validation of our algorithm with the REDD demonstrates its effectiveness.

The objectives of this contribution are to present an algorithm to detect events occurring on the main electrical channel of a household and to group them through an inductive method that can be scaled up to practical use since it does not require the cardinality of appliances used in the house. We then illustrate a generalizing module which makes it possible to obtain the consumption profiles of household appliances both of ON/Off and multi-state type.

The rest of this work is organized as follows: Section II presents the general NILM framework, Section III exposes the related work, Section IV encapsulates the details of the proposed solution, Section V outlines the solution validation experiments and the discussion of the result obtained and Section VI concludes

our contribution.

80 2. Related Terminologies

The main objective of NILM research remains the estimation of the contribution of each appliance in a habitation based solely on knowledge of the total incoming load. The overall load model can be formulated as follows:

$$P(t) = \sum_{i=1}^n P_i(t) + e(t) \quad (1)$$

in which n stands for the cardinality of appliances present in the household, $P(t)$ the overall load entering the household, $P_i(t)$ the individual consumption of appliance i at time t , and $e(t)$ the small margin of error or noise present in the circuits. Based on the above formulation of electrical load disaggregation, several approaches have been explored to overcome the aforementioned problem. 85 The steady state and transient state represent the commonly used examples. The steady state represents the operating phase during which the power used is almost constant. This characteristic is easily detected on low frequency data since the steady state is generally spread over several measurements. However, 90 steady state has the downside of being inefficient in differentiating devices having relatively identical levels of power consumption used simultaneously. The transient state, on the other hand, studies the high-definition curve produced when the device is switched on or off, or when moving from one operating mode to another. Detection of this type of feature requires high frequency sampling. 95 As such a feature is identified over a relatively short period of time, the risk of overlapping is low.

In order to better understand the diversity among household appliances, a classification into four distinct groups has been established according to their operating profile [8]-[9]. The operating profile depends on the internal composition of each appliance. Consequently, the design of a solution can be directed 100 toward particular types of appliances for an in-depth study of the case. The four classes of appliances are as follows:

- ON/OFF or type I devices: it groups devices that have only two operating modes such as a table lamp; such appliances are either on or off.
- 105 • Finite State Machine (FSM) or type II: They are devices that have more than two operating modes. We can take the case of a ventilator that works according to the speed of rotation set by the user.
- Continuously Variable Devices (CVD) or type III: Here are devices that have a random consumption profile that varies all the time. They usually
110 do not have a steady state and so no signature vector.
- Permanent Consuming Devices (PCD) or type IV: They are appliances which never go completely off or which continue to consume an insignificant amount of current even when they are not in use; this is the case of the majority of smart devices, which always wait for a signal from the
115 users to start working.

3. Related Works

One of the most recent approaches that have been utilized to tackle the energy disaggregation challenges is based on appliance usage patterns in the active power signal. With a spectral decomposition algorithm, [10] proposed a study
120 that estimates, according to the time and day, whether a device is switched on or off and the active load required in the following minutes. A cloud based using the Markov chain appliance modeling is used in [11] to build an online NILM solution. The core of the solution relies on the Factorial Hidden Markov model that disaggregate in real-time on the event detected from the transient's
125 features captured on the main signal. The cloud-based NILM architecture does not rely on training with the ground-truth readings from household appliances and supports a smart-grid usage. Another recent work makes use of Activities of Daily Livings (ADL) to build a classification method using a deep learning architecture for sensing the activation of specific major household devices [3].The

130 proposed approach is tested on the UK-DALE [12] dataset and the results confirm the effectiveness of the solution. Various adaptations of the Hidden Markov Models (HMM) have been evaluated for forecasting the possible combinations of operation state of residential appliances [13]-[14]. The study of [15] used a Factorial Hidden Markov Models (FHMM) to determine the device-specific load
135 models considering the overall power readings which obtained a disaggregation efficiency of 90% and 80% respectively for type I and type II devices. The research of Yong et al. [16] was directed toward the establishment of a system for event matching built on the improved Kuhn-Munkras algorithm. By a three-fold approach including the event detection, appliance signature establishment
140 and the load identification using the event matrix, they have obtained a satisfactory experiment result of the proposed approach on the publicly available REDD dataset.

Classification and clustering algorithms have also been used in the design of methods for electrical load separation. An evaluation of a modified version
145 of the k-nearest neighbor classifier (K-NN), a hierarchical clustering approach (HAC) and a custom build device matching (DevMat) approach is presented in [17]. The model presented in this paper is an event -based clustering approach that support instant detection. The study results show that high frequency features may greatly improve the load disaggregation process.

150 Using an auto-associative neural network, Lorena et al. [18] illustrate a deep learning based appliance load forecasting solution. The auto-associative neural networks are arranged in parallel and they compete against one another when a new input vector is entered and the closest recognition is accepted to identify the given electrical device. Validation experiments with the REDD and UK-DALE
155 datasets show that the proposed method achieves a performance higher than 95%. Subsequently, [8] presents an approach based on a novel signature using different reflection rate to emphasize on the variation of the aggregated power signal. The load separation model uses a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) architecture and an additional advanced deep
160 learning. The evaluation experiments with the UK-DALE dataset produced a

satisfactory result. Also, a Convolutional Neural Network (CNN) -based method is proposed by [7] for appliance consumption prediction based on the previous operation mode. The work of reference [19] is a real-time identification system of several devices from measurements of a smart meter by the means of a multi-
165 label consistent deep dictionary learning and deep transform learning. The study has introduced energy disaggregation as a multi-label classification issue. The hypothesis is that during training, no more than the total power signal data and the logs (ON/OFF status) of the various appliances are used. Yanchi et al. have mainly been interested by the junction between image classification
170 and NILM problem [20]. They used a pre-trained deep learning model on visual recognition dataset for analyzing the changes in the V-I trajectory. The findings of this research show that the proposed method improves the current results in the field and that a future generalization of the solution can be envisaged.

A new methodology to predict the energy usage of a family home, or group
175 of homes, based on electrical power disaggregation and graph spectral clustering is introduced by [21]. In this approach, the overall energy usage is divided into individual device signals and the power of each device is predicted independently. Subsequently, Marisa et al. describe the implementation of an algorithm for extracting electrical properties and pattern detection, which can be used to
180 estimate the power consumption of each appliance based on the house's overall electrical load signal [22]. Experiments demonstrated the effectiveness of the suggested solution. Kitisak et al. also presents in [23] an investigation on the impact of lowering the sampling rate using a Recurrent Neural Network architecture. Experience has shown that the result of disaggregating the overall reading
185 remains acceptable even when the interval between measurements increases to a given threshold.

Throughout this section, we can notice that several angles have been used for the establishment of electrical load separation systems. However, it should be noted that the elaboration of an adoptable solution must take into account
190 the complexity of implementation and a diversity of appliance types, while producing a high-performance system that will not decrease when the experimental

environment changes. The privacy concern during data transmission between different component of the designed system is a critical factor that should not be ignored [24]. Those are the fundamental constraints that impedes research in the field of NILM. Based on the model presented in [25], we propose an advanced energy disaggregation method that relies on the overall power collected at a low frequency. Considering the results of the validation experiments, we can conclude that the presented solution fulfills the requirements for a practical use.

4. The Proposed Algorithm

The process of obtaining an estimation of the electrical energy used by each individual appliance is described on Figure 1.

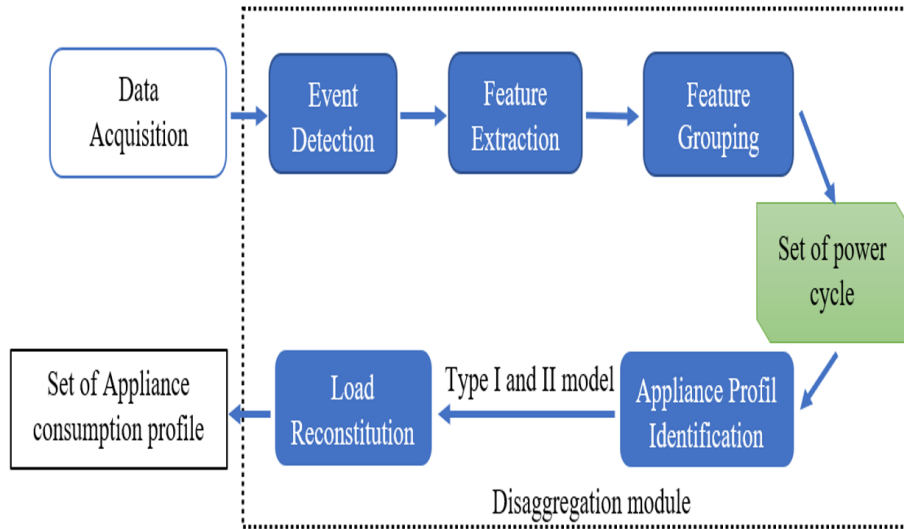


Figure 1: Energy disaggregation framework overview

4.1. Data Acquisition and Preprocessing

The collection of measurement data of the energy consumed involves a data-center based module where the different nodes of the system act to transform

and centralize the captured information from the electrical sensors. Various edge service architecture can be considered for the design of such a system [26] - [27]. The work of Xu et al. introduced a two-phase offloading optimization approach to enhance the security related to data exchange between the various units of an IoT architecture [28] - [29]. This paper explores a novel utility-aware task offloading method and a joint optimization method on the offloading problem. Blockchain architecture also constitutes a reliable approach to assure data integrity for an edge computing system [30]. At this stage, smart power meters are deployed in the circuits to measure the energy flow. For the purposes of experience, the total incoming load only goes beyond the sufficient data to deduce the energy of the appliances used in the household and furthermore for evaluating the proposed solution. The consumption of individual appliance or group of appliances is also required for an in-depth study of each type of appliance, for supplying input data for learning algorithms and additionally for the validation of the results after the load disaggregation process. Thus, the first layer of our edge computing architecture is made up of electrical sensors that will measure the voltage and current deployed in the circuits connected to each device. The data is then routed to the house's gateway for further investigation. The preprocessing consists in aligning the collected data in order to adjust the current and voltage data as well as to ensure the synchronization of the time axes of the various measurements. After computing the power corresponding to the collected data, an encryption using the twofish or advanced encryption standard algorithm is performed on the raw data in order to ensure a secure transfer to the long-term storage system located on the cloud [31].

The data is generally available in low frequency (sampling rate less than or equal to 1 hertz) or high frequency (up to KHz or MHz). The low frequency data allows to observe the steady state of the appliances while the high frequency data offers in addition the details of the transient state. Therefore, the features taken into account during data collection are of great importance as they determine the choice of approach to be used and increase the scope of study during the experimental period. Several features can be supported such as voltage, current,

active power, passive power, reactive power, real power, apparent power and more. A detailed description of electrical energy data collection system can be found in [32] - [33]. In terms of load separation system design, the use of the minimum feature which is the low-frequency power readings is considerably encouraged since the majority of smart meters available on the market support this configuration [34].

4.2. The Proposed Unsupervised NILM Algorithm

The proposed algorithm can be subdivided into three essential phases, namely: event detection and feature extraction, significant event clustering and finally the appliance power profile identification and load reconstitution. The flow-chart of the proposed dsolution is illustrated in on Figure 2.

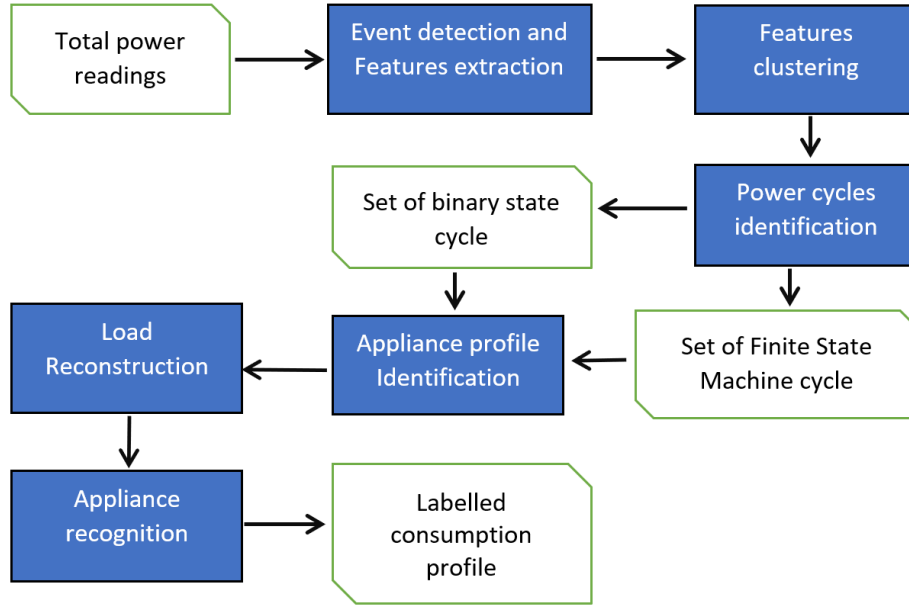


Figure 2: Proposed solution flow-chart

4.2.1. Event Detection and Feature Extraction

Starting from the aggregated signal of the house, this module is responsible
250 of tracking and retrieving significant changes that have taken place during a
defined time interval. The event detection module works in our case with the
assumption that only one event could occur during a time frame θ_t . We then
examine the time series to find obvious indications of the occurrence of an event
using a threshold value \mathcal{E} chosen according to the error β in the measurements
255 of the incoming load (power grid noise) and the minimum consumption P_{\min}
of the low consumption devices used in the household.

One approach for scanning the data is to compare the local variation from
the measurements over the time frame θ_t . If we consider $\theta_t = 5\text{s}$ for a signal
measured at the frequency of 1 Hz, the local average $\omega(t)$ of the power at time t
can be obtained by Eq. (2) and the local variation $\varphi(t)$ is deduced as presented
in Eq. (3). The change at time t is considered as an event if $\varphi(t)$ is equal or
greater than the threshold \mathcal{E} .

$$\omega(t) = \frac{1}{5} \sum_{i=0}^{n<5} P(t-i) \quad (2)$$

$$\varphi(t) = \sqrt{\frac{1}{5} \sum_{i=0}^{n<5} (P(t-i) - \omega(t))^2} \quad (3)$$

Whilst this method is ingenious and effective for the detection of ON/OFF
appliances, it has some limitations to handle events of relatively similar mag-
nitude that occur during the selected θ_t time interval or identifying a transient
phase that covers a time frame higher than θ_t . Due to the stochastic nature
of the occurrence of events, we have considered in addition to the local power
variation the direct difference Eq. (4) between two measurements, which makes
our algorithm more sensitive to major changes that occur. The basic formula
for this variation is as follows:

$$\Delta P(t) = |P(t) - P(t-1)| \geq \varepsilon \quad (4)$$

where $\Delta P(t)$ is the absolute value of the power difference of the power between

each pair of measurements, t is the time for which we check the occurrence of the event and \mathcal{E} is the threshold for a change to be considered as significant. The process used to detect the events that have taken place on the aggregate signal is presented at the level of the algorithm, as shown in Table 1.

Table 1: A SIGNIFICANT EVENT DETECTION ALGORITHM

Input: Time series of aggregated power $P(t)$, threshold ε and time frame θ_t .

Output: Set of significant events.

Begin

For each measurement in aggregated power reading

 Compute $\omega(t)$ and $\varphi(t)$ using θ_t

 Compute $\Delta P(t)$

If $\varphi(t) \geq \varepsilon$ then

 Register $P(k) = \varphi(t)$ as event

 Add the time t to $T(k)$

Else If $\Delta P(t) \geq \varepsilon$ then

 Register $P(k) = P(t) - P(t-1)$ as event

 Add the time t to $T(k)$

Else

 Time t is part of a steady state

End If

End For

End

Note: $P(k)$ is the feature vector of events and $T(k)$ the occurrence time.

The features obtained from event detection of the events include the amplitude of the transition and the time reference at which the event occurred. Figure 3 illustrates the frequency distribution of the identified features over a day's duration. We can notice an apparent parity between the events with positive and negative amplitudes. This is due to the fact that a positive amplitude event representing the switching on of the device or the beginning of a new

operating phase has a high chance of having an event with similar amplitude but negative marking the shutdown of the device or the end of this operating
270 phase. This explains the normal distribution of events around the zero value.

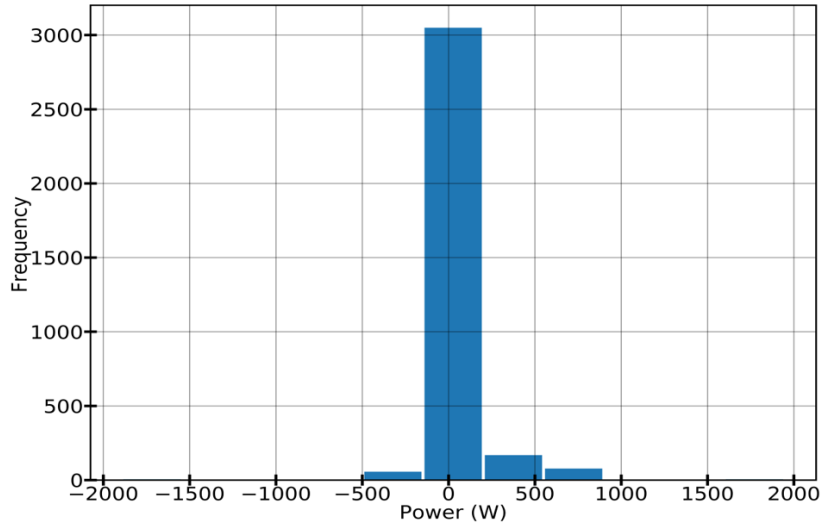


Figure 3: Distribution of the extracted features over one day

4.2.2. Significant Event Clustering

In this section, all the features are categorized using the absolute value of the event magnitude. The objective of this exercise is to implicitly group the events marking the beginning and the end of the steady states of the devices in
275 order to simplify the identification of power consumption cycles. The use of the absolute value of the selected variations constitutes a first level of reinforcement of the symmetrical character which must dominate the distribution of events. It also has the advantage of reducing the complexity level regarding the data to be clustered and thus the computation demand of the clustering algorithm.

280 Given the inability to know the number of devices and the number of operating phases involved, the grouping algorithm must be inductive enough to automatically identify the required number of clusters to effectively subdivide the events. With this in mind, we used an adaptation of the Mean Shift [35] algorithm which is a density-based algorithm that needs no prerequisite infor-

285 mation regarding the number of clusters, and does not constrain the shape of
the clusters. Considering an ensemble of data points, the algorithm iteratively
allocates every individual data point to the nearest cluster centroid. Since the
direction to the nearest cluster centroid is defined by the location of most of
the proximate points, at each iteration, every data point will move closer to the
290 location of the majority of the points, which is or will eventually result in the
center of the cluster. The output of the algorithm is in the form of a linking of
any point to its cluster.

Figure 4 shows a distribution of events in several clusters according to the
amplitude of the change that occurred.

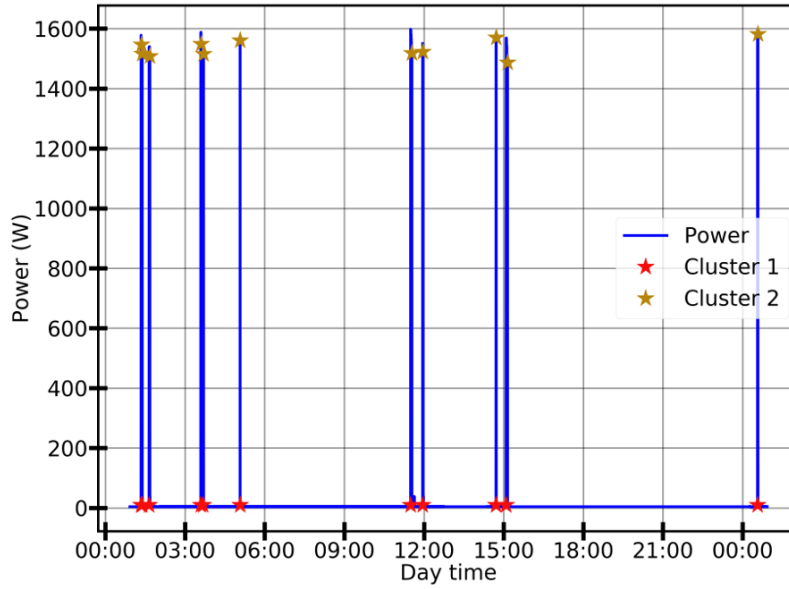


Figure 4: Example of event grouping on the microwave channel of house 1.

295 *4.2.3. Appliance Power Profile Identification and Load Reconstitution*

The guiding principles that lead our research are among others:

- a single event can occur during the time frame separating two measurements.
- all devices can have a power profile represented by more than one steady

300 state (ON/OFF devices as well).

- there may be interleaving between the operating phases of devices of different types.

At this stage, we have of the characteristics of the significant transitions that happened on the main channel grouped by class according to the event's amplitude similarity. The profile identification module takes care of identifying
305 the beginning and end of each steady state. Inside each cluster, positive values are supposed to represent the starting point of an operating phase while an exit transition from an operating phase is represented by an event with negative amplitude. It is then a question of finding all the pairs of events delimiting the
310 beginning and the end of the power cycles of a device as well as the level of consumption during this period. This process is two-fold:

a) ON/OFF case

During this process, the power cycles that are supposed to belong to the same device are tracked within each cluster obtained from the event grouping
315 step. After matching two events, the times t_1 and t_2 representing respectively the on and off event of the cycle are recorded in addition to the cluster center value as the amplitude of the steady state. The events that have not been matched are also retained for the next step which will consist in checking their membership as a multi-state device.

320 **b) Finite State Machine case**

Once the power cycles existing on the internal plane of each cluster have been identified, the next step will be to rescan the occurrence profile of the latter in order to detect events that fall immediately after each other or the ones that are separated by an intermediate cycle. This is the first step in the
325 identification of power cycles that could belong to the same device.

Devices whose energy profile is made up of operating phases that have different amplitudes will naturally see their events scattered over various groups according to their amplitudes. The example of microwaves can be considered in Figure 4. A more sophisticated search is made by looking for the missing part

330 of each cluster events that were not matched in the previous step.

This process starts with the high amplitude cluster points. By combining the centers of lower amplitude clusters, we look for the best set of events that satisfies the current case.

After finding the events that define the power cycles, we obtain a set of
335 operating mode delimited by a start time t_1 and an end time t_2 , an amplitude for each cycle and a marker that identifies all the cycles that have been assigned to the use of the same device by our algorithm. The next task consists in reconstructing both the consumption profile of the detected devices as well as at the main channel by summing up the consumption models detected. The
340 validation of this step is discussed in the second experiment presented in section 5. The output of the load reconstruction module consists of a set of consumption profile detected from the aggregated signal.

4.2.4. Appliance Recognition

The appliance recognition module aims at recognition from a detected consumption pattern, the device that generated it. This task is carried out in two phases, which are described as follows: i) at first, the similarity between a detected usage pattern and the actual consumption pattern of a device from the dataset is evaluated by using the maximum likelihood classifier. Thus, each profile is then attached to the device that obtained the highest similarity rate; ii) second, we compare the extracted power profile to the signal of the indexed device. An event is considered correctly assigned if it has been detected in the chosen θ_t time interval and with an amplitude which is approximately equal to the one recorded at the device level with an error δ bounded as defined in Eq. (5) below, where ε represents the power grid noise:

$$\mathbf{0} \leq \delta \leq \varepsilon \tag{5}$$

During this process, we report the values of True Positive, True Negative
345 False Positive and False Negative, the estimated state and the actual state for future evaluation of our algorithm in Section 5.

5. Validation of the Proposed Solution

This section provides the context for validating the proposed solution. The experiments were carried out on using the Reference Energy Disaggregation Dataset (REDD) [36]. The dataset is made up of the power readings collected from six real houses over three to nineteen days at a frequency of 1 hertz on the main channel and every 3 seconds on the subsequent level circuits in the low frequency folder. In addition to the overall signal, this data set also provides power readings of the consumption of appliances such as refrigerator, dishwasher, lighting, air conditioner, microwave, etc., which is a mixture of data from different types of appliance available for research. Due to the frequency of power readings and the set of appliances taken into account in the various houses, this dataset constitutes an adequate source of data for the execution of the evaluation experiments of our work.

For the purpose of validation experiments of the proposed algorithm, we have chosen $\varepsilon = 0.02\text{KW}$. The choice of this threshold is explained by the comparison results between the aggregated consumption and the sum of the appliances having contributed to the total load in Figure 5 and the difference between them (as shown in Figure 6). This allows to evaluate the amplitude of the interference (Figure 6) during data collection. The interference is obtained by calculating the difference between the aggregate power and the total consumption of devices for the times t when the readings are available for the main channel as well as for the devices in the subchannels. Although it is obvious that the threshold must be set to make a clear difference between changes due to unwanted electrical or electromagnetic energy that degrades the quality of signals and data, the low consumption appliances must also be taken into account.

5.1. Evaluation Metrics

To evaluate our work, we used a set of commonly used metrics to validate electrical energy disaggregation solutions presented in [3] and [11]. These are Precision Eq. (6) which is the positive predictive values, Recall Eq. (7) which

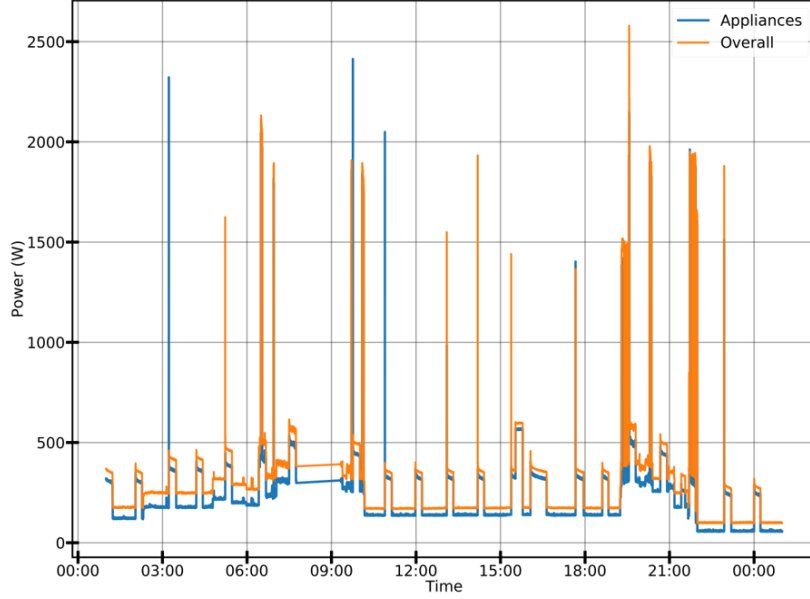


Figure 5: Evaluation of the overall load versus the sum of the consumption of the registered devices.

represents the true positive rate or sensitivity, the Disaggregation Accuracy Eq. (8), F-measure Eq. (9) the harmonic mean of Precision and Recall, and the Total Energy Correctly Assigned Eq. (10) which meters the quantity of power that has been properly allocated.

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

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

$$Accuracy(Acc.) = \frac{TP + FN}{TP + TN + FP + FN} \quad (8)$$

$$F - measure(F1) = 2 \times \frac{P \times R}{P + R} \quad (9)$$

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

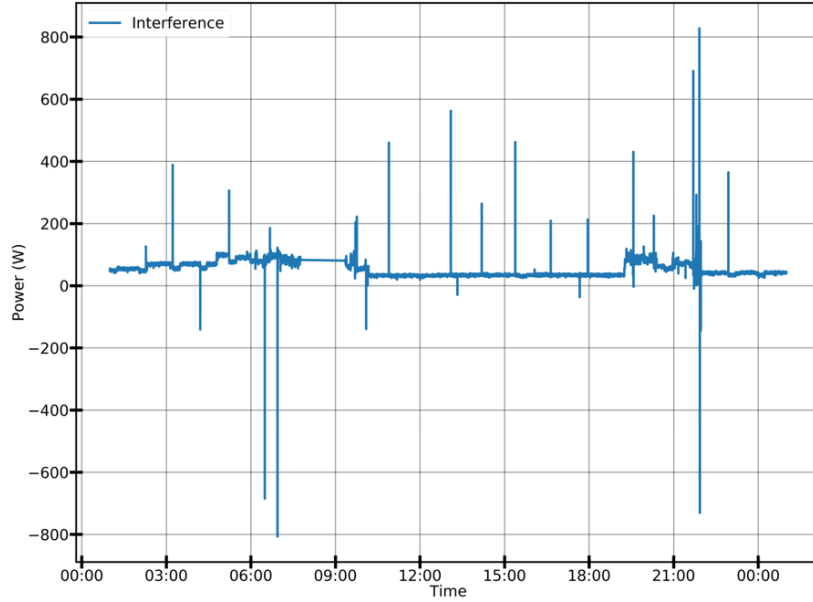


Figure 6: Interference on the aggregate signals

where TP stands for true-positives (properly predicted that the device is ON), TN denotes true-negative (properly predicted that the device is OFF), FP lies
 375 for false-positives (the device is predicted as ON but OFF), and FN for false-
 negatives (the device is ON but predicted OFF), $\hat{y}_t^{(i)}$ represents the isolated
 apparatus load, $\hat{y}_t^{(i)}$ represents the initial apparatus load, \bar{y}_t is the monitored
 aggregate signal, K is the cardinality of apparatus load and T is the number of
 samples.

380 5.2. Experimental Scenarios

The first experiment aims to validate the event detection and features ex-
 traction module using the aggregated signal. The scan of one day data was
 performed considering the time frame $\theta_t = 5$ seconds. The collected features
 obviously show the ability of the proposed method to identify significant events
 385 that took place during a day. The distribution of these events is shown in
 Figure 7. It is possible to observe a high concentration in the middle of the
 day, which is generally due to the daily activity of the household occupants.

We also see fairly high peaks during the night, which indicates the presence of special-purpose equipment. Further on, we also performed the grouping of the extracted features and the identification of the consumption profiles using the extracted features from one-day data of each house. The results obtained are presented in Table 2.

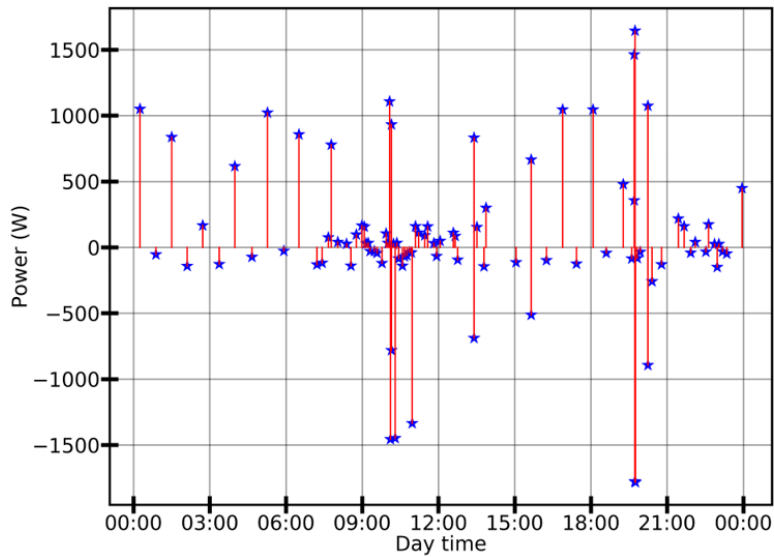


Figure 7: Events detected over a period of one day.

Table 2: Event Detection Evaluation Over One Day Consumption

House Number	P (%)	R (%)	F1(%)	Acc (%)	TECA (%)
1	97.43	99.09	98.25	95.21	90.62
2	99.33	96.34	97.81	98.10	95.02
3	97.68	96.67	97.18	95.15	89.91
4	96.38	98.76	97.56	94.98	87.37
5	98.35	88.03	92.91	98.37	78.41
6	97.08	88.37	92.52	96.94	92.04

P: Precision, R: Recall, F1: F-measure, Acc: Disaggregation Accuracy, TECA:

Total Energy Correctly Assigned.

The evaluation of the proposed solution using domain estimation metrics indicates relatively good performance depending on the complexity of the aggregated signal. It is also worth noting the efficiency of the reconstruction of the consumption profiles of the devices that contributed the total load during the 24-hour period considered.

The second experiment follows the validation of our solution's ability to extract essential characteristics from the aggregated load collected over a 24-hour period. Therefore, we propose here an evaluation based on data from houses 1 and 6, which have a wide variety of household appliances. Using an algorithm that goes through the data from these houses, we created the power models of the appliances and then verify the rendering with the appliance recognition module. Figure 8 and 9 show the detected and actual power profiles for a typical day respectively.

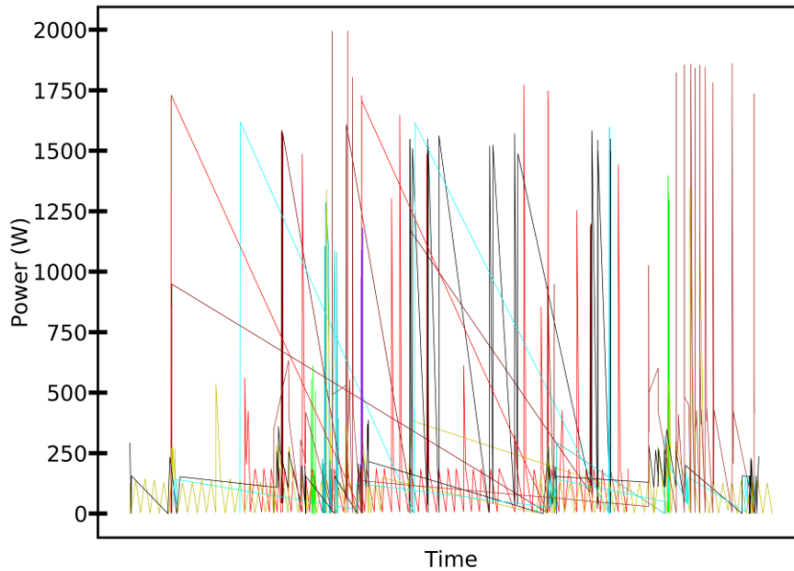


Figure 8: Extracted power profile of a typical day from House 1.

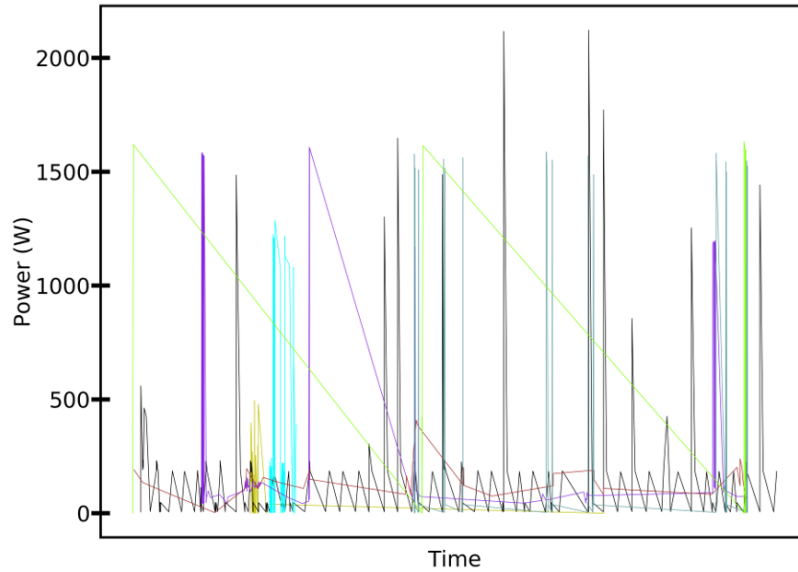


Figure 9: Actual power profile of a typical day from House 1.

After profile identification, the reconstitution of the total aggregate load is necessary to have an overview of the amount of energy that has been taken into account by the algorithm. A comparison of the two loads is made in Figure 10 where we can notice the great similarity between the values predicted and the real one. The final phase was to evaluate how good our energy was distributed between the various devices. The details per device are presented in Table 3. It should be noted that the high-consumption devices are effectively detected while the low-consumption device exhibits relative inefficiency. This is due to the large number of low (less than 100 W) amplitudes cycles detected. This creates a noticeable disproportion in the distribution of the characteristics after the clustering step (Figure 11). However, the misclassification of those devices has a relatively low impact on the final results since they are not widely represented and also did not record a long period of use.

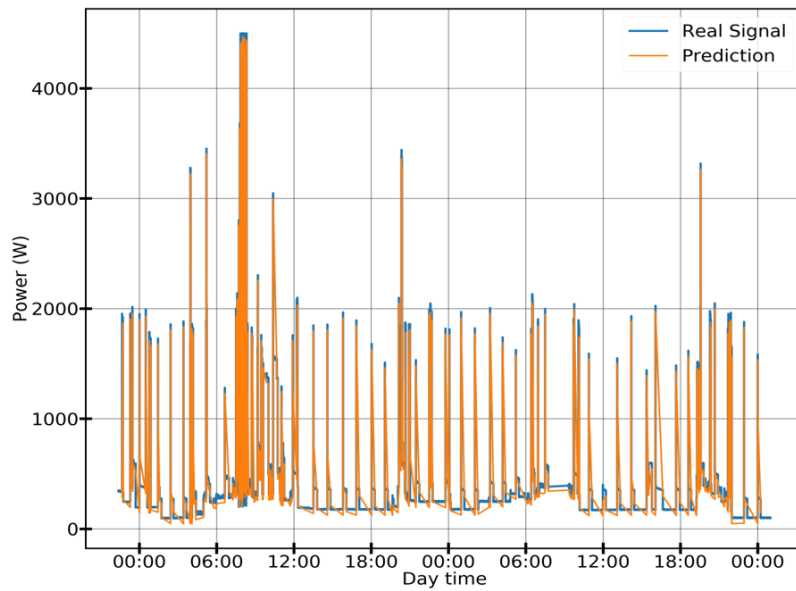


Figure 10: Two days actual and predicted overall load.

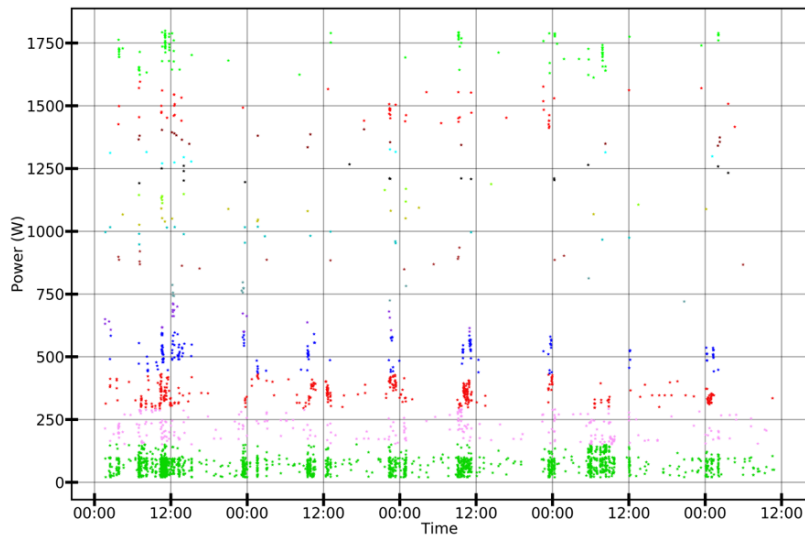


Figure 11: Distribution of extracted features from 5 subsequent days of House 6.

Table 3: Disaggregation Result Per Appliance

Appliance	Accuracy
Stove	89.74(%)
Refrigerator	99.03(%)
Lighting	82.64(%)
Washer dryer	91.14(%)
Air Conditioning	93.61(%)
Electric heater	97.43(%)
Dishwasher	95.56(%)
Microwave	98.41(%)
Overall	93.45(%)

The prowess of the appliance signal reconstitution module can be appreciated
 420 from the refrigerator load identification shown in Figure 12.

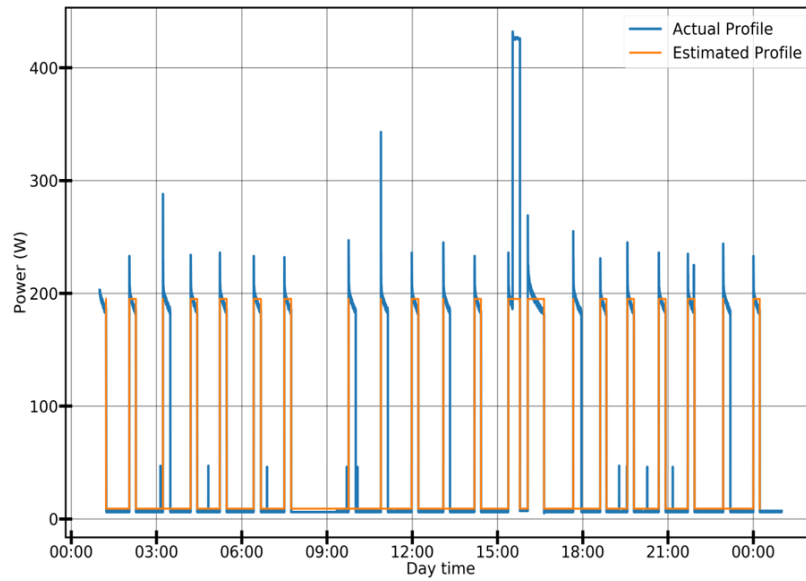


Figure 12: Reconstituted energy signal versus ground truth signal of a refrigerator

The experiment based on houses 1 and 6 were successfully carried out in
 324s duration, resulting in average 4s per day. This second experiment led to

a performance of 98.89% accuracy and TECA of 90.60% which is a satisfactory result.

425 We made use of the python programming language assisted by the NumPy, Pandas and Matplotlib libraries to perform the research experiments. The hardware side was covered by an Intel® Core™ i7-4720HQ computer running at 3.50GHz, 16GB RAM memory and a graphics accelerator GPU NVIDIA GEFORCE GTX 960M.

430 *5.3. Results and Discussion*

Our solution for energy disaggregation based on the principles originally defined by Hart [6] has proven to be effective due to its low complexity and ability to effectively separate the loads consumed by ON/OFF and Finite State Machine (FSM) appliances in a household. In the above experimental scenarios, 435 we have successfully evaluated all the steps of the solution using the publicly available experimental data from the REDD. The comparison of our results with the state of the art is provided in Table 4. As the approaches investigated and the underlying conditions on which these studies were based are not the same, a profound comparison cannot be made reliably. Nevertheless, on the 440 basis of the evaluation metrics presented, it appears that the solution proposed in this contribution takes up the challenge in terms of the general classification of performance.

Table 4: Performance Comparison Between The Proposed Solution And Others

Approach	Accuracy	F-measure	TECA
Proposed approach	98.89	96.04	90.60
AANNS [18]	98.7	95.4	–
PBN [10]	84.0	88.5	–
FHMM [11]	–	–	98.65
Viterbi Algorithm [37]	–	81.1	–

6. Conclusion and Future Work

In this contribution, we propose a novel event-based method for NILM based
445 on a non-parametric clustering algorithm and the active power as unique electrical feature. This research effort aims to produce a cost-effective solution that does not require human intervention during the disaggregation process and minimizes both computational and hardware complexity while preserving a high accuracy. In short, the study presents an approach for analyzing the
450 overall electrical load and generalizing the meaningful features contained in the signal for later identification of its aggregated components. The experiments and validation results of the proposed solution with data from the six houses of the REDD confirm the efficiency of the presented approach in estimating the consumption of ON/OFF and multi-states devices. Although our approach
455 performs effectively, it has some limitations regarding the ability to handle Continuously Variable Devices (CDV). In future research, the potential of inference algorithms for predicting appliance consumption may be an interesting direction to address the randomness of continuously variable devices.

7. Acknowledgement

460 This work has received funding from National Natural Science Foundation of China (No. 41911530242, 41975142), 5150 Spring Specialists (05492018012, 05762018039), Major Program of the National Social Science Fund of China (Grant No.17ZDA092), 333 High-Level Talent Cultivation Project of Jiangsu Province (BRA2018332), Royal Society of Edinburgh, UK and China Natural Science Foundation Council (RSE Reference: 62967.Liu.2018.2) under their
465 Joint International Projects funding scheme, basic Research Programs (Natural Science Foundation) of Jiangsu Province (BK20191398) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund.

470 **References**

- [1] X. Xu, X. Liu, Z. Xu, F. Dai, X. Zhang, L. Qi, Trust-oriented iot service placement for smart cities in edge computing, *IEEE Internet of Things Journal*.
- [2] B. Dudley, et al., Bp statistical review of world energy, BP Statistical Review, London, UK, accessed Aug 6 (2018) 2018.
475
- [3] M. A. Devlin, B. P. Hayes, Non-intrusive load monitoring and classification of activities of daily living using residential smart meter data, *IEEE Transactions on Consumer Electronics* 65 (3) (2019) 339–348.
- [4] A. Ridi, C. Gisler, J. Hennebert, A survey on intrusive load monitoring for appliance recognition, in: 2014 22nd international conference on pattern recognition, *IEEE*, 2014, pp. 3702–3707.
480
- [5] M. D. Burbano Acuña, Intrusive and non-intrusive load monitoring (a survey), *Latin American Journal of Computing Faculty of Systems Engineering Escuela Politécnica Nacional Quito-Ecuador* 2 (1).
- [6] G. W. Hart, Nonintrusive appliance load monitoring, *Proceedings of the IEEE* 80 (12) (1992) 1870–1891.
485
- [7] M. Kaselimi, E. Protopapadakis, A. Voulodimos, N. Doulamis, A. Doulamis, Multi-channel recurrent convolutional neural networks for energy disaggregation, *IEEE Access* 7 (2019) 81047–81056.
- [8] J. Kim, T.-T.-H. Le, H. Kim, Nonintrusive load monitoring based on advanced deep learning and novel signature, *Computational intelligence and neuroscience* 2017.
490
- [9] S. W. Park, L. B. Baker, P. D. Franzon, Appliance identification algorithm for a non-intrusive home energy monitor using cogent confabulation, *IEEE Transactions on Smart Grid* 10 (1) (2017) 714–721.
495

- [10] S. Welikala, C. Dinesh, M. P. B. Ekanayake, R. I. Godaliyadda, J. Ekanayake, Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting, *IEEE Transactions on Smart Grid* 10 (1) (2017) 448–461.
- 500 [11] M. A. Mengistu, A. A. Girmay, C. Camarda, A. Acquaviva, E. Patti, A cloud-based on-line disaggregation algorithm for home appliance loads, *IEEE Transactions on Smart Grid* 10 (3) (2018) 3430–3439.
- [12] J. Kelly, W. Knottenbelt, The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes, *Scientific data* 2 (1) (2015) 1–14.
- 505 [13] W. Kong, Z. Y. Dong, D. J. Hill, J. Ma, J. Zhao, F. Luo, A hierarchical hidden markov model framework for home appliance modeling, *IEEE Transactions on Smart Grid* 9 (4) (2016) 3079–3090.
- [14] D. Bajović, K. He, L. Stanković, D. Vukobratović, V. Stanković, Optimal detection and error exponents for hidden semi-markov models, *IEEE Journal of Selected Topics in Signal Processing* 12 (5) (2018) 1077–1092.
- 510 [15] A. Zoha, A. Gluhak, M. Nati, M. A. Imran, Low-power appliance monitoring using factorial hidden markov models, in: 2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, IEEE, 2013, pp. 527–532.
- 515 [16] Y. Xiao, Y. Hu, H. He, D. Zhou, Y. Zhao, W. Hu, Non-intrusive load identification method based on improved km algorithm, *IEEE Access* 7 (2019) 151368–151377.
- [17] T. Bernard, M. Verbunt, G. vom Bögel, T. Wellmann, Non-intrusive load monitoring (nilm): Unsupervised machine learning and feature fusion: Energy management for private and industrial applications, in: 2018 international conference on smart grid and clean energy technologies (ICSGCE), IEEE, 2018, pp. 174–180.
- 520

- [18] L. R. Morais, A. R. Castro, Competitive autoassociative neural networks for
525 electrical appliance identification for non-intrusive load monitoring, *IEEE
Access* 7 (2019) 111746–111755.
- [19] V. Singhal, J. Maggu, A. Majumdar, Simultaneous detection of multiple
appliances from smart-meter measurements via multi-label consistent deep
dictionary learning and deep transform learning, *IEEE Transactions on*
530 *Smart Grid* 10 (3) (2018) 2969–2978.
- [20] Y. Liu, X. Wang, W. You, Non-intrusive load monitoring by voltage–
current trajectory enabled transfer learning, *IEEE Transactions on Smart
Grid* 10 (5) (2018) 5609–5619.
- [21] C. Dinesh, S. Makonin, I. V. Bajić, Residential power forecasting using load
535 identification and graph spectral clustering, *IEEE Transactions on Circuits
and Systems II: Express Briefs* 66 (11) (2019) 1900–1904.
- [22] M. B. Figueiredo, A. De Almeida, B. Ribeiro, An experimental study on
electrical signature identification of non-intrusive load monitoring (nilm)
systems, in: *International Conference on Adaptive and Natural Computing*
540 *Algorithms*, Springer, 2011, pp. 31–40.
- [23] K. Osathanunkul, K. Osathanunkul, Different sampling rates on neural
nilm energy disaggregation, in: *2019 Joint International Conference on Dig-
ital Arts, Media and Technology with ECTI Northern Section Conference
on Electrical, Electronics, Computer and Telecommunications Engineering*
545 *(ECTI DAMT-NCON)*, IEEE, 2019, pp. 318–321.
- [24] L. Qi, X. Wang, X. Xu, W. Dou, S. Li, Privacy-aware cross-platform ser-
vice recommendation based on enhanced locality-sensitive hashing, *IEEE
Transactions on Network Science and Engineering*.
- [25] N. Henao, K. Agbossou, S. Kelouwani, Y. Dubé, M. Fournier, Approach
550 in nonintrusive type i load monitoring using subtractive clustering, *IEEE
Transactions on Smart Grid* 8 (2) (2015) 812–821.

- [26] X. Wang, L. T. Yang, L. Kuang, X. Liu, Q. Zhang, M. J. Deen, A tensor-based big-data-driven routing recommendation approach for heterogeneous networks, *IEEE Network* 33 (1) (2019) 64–69.
- 555 [27] X. Wang, L. T. Yang, Y. Wang, L. Ren, M. J. Deen, Adtt: A highly-efficient distributed tensor-train decomposition method for iiot big data, *IEEE Transactions on Industrial Informatics*.
- [28] X. Xu, C. He, Z. Xu, L. Qi, S. Wan, M. Z. A. Bhuiyan, Joint optimization of offloading utility and privacy for edge computing enabled iot, *IEEE Internet of Things Journal*.
- 560 [29] M. Huang, K. Zhang, Z. Zeng, T. Wang, Y. Liu, An auv-assisted data gathering scheme based on clustering and matrix completion for smart ocean, *IEEE Internet of Things Journal* (2020) 1–1.
- [30] X. Xu, X. Zhang, H. Gao, Y. Xue, L. Qi, W. Dou, Become: Blockchain-enabled computation offloading for iot in mobile edge computing, *IEEE Transactions on Industrial Informatics*.
- 565 [31] X. Wang, L. T. Yang, H. Li, M. Lin, J. Han, B. O. Apduhan, Nqa: A nested anti-collision algorithm for rfid systems, *ACM Transactions on Embedded Computing Systems (TECS)* 18 (4) (2019) 1–21.
- 570 [32] A. Filip, Blued: A fully labeled public dataset for event-based nonintrusive load monitoring research, in: *2nd Workshop on Data Mining Applications in Sustainability (SustKDD)*, 2011, p. 2012.
- [33] B. Jiang, G. Huang, T. Wang, J. Gui, X. Zhu, Trust based energy efficient data collection with unmanned aerial vehicle in edge network, *Transactions on Emerging Telecommunications Technologies* e3942.
- 575 [34] D. Egarter, V. P. Bhuvana, W. Elmenreich, Paldi: Online load disaggregation via particle filtering, *IEEE Transactions on Instrumentation and Measurement* 64 (2) (2014) 467–477.

- [35] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature
580 space analysis, *IEEE Transactions on pattern analysis and machine intel-*
ligence 24 (5) (2002) 603–619.
- [36] J. Z. Kolter, M. J. Johnson, Redd: A public data set for energy disaggrega-
tion research, in: *Workshop on data mining applications in sustainability*
(SIGKDD), San Diego, CA, Vol. 25, 2011, pp. 59–62.
- 585 [37] M. Zeifman, Disaggregation of home energy display data using probabilistic
approach, *IEEE Transactions on Consumer Electronics* 58 (1) (2012) 23–31.