



**HAL**  
open science

# Enhancing Energy Disaggregation with Attention-Based Neural Network

Nidhal Balti, Baptiste Vrigneau, Pascal Scalart

► **To cite this version:**

Nidhal Balti, Baptiste Vrigneau, Pascal Scalart. Enhancing Energy Disaggregation with Attention-Based Neural Network. SEB 2024, Sep 2024, Santa Cruz, France. <hal-04907448>

**HAL Id: hal-04907448**

**<https://hal.science/hal-04907448v1>**

Submitted on 23 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/387339384>

# Enhancing Energy Disaggregation with Attention-Based Neural Network

Conference Paper · September 2024

CITATIONS

0

READS

13

3 authors:



**Balti Nidhal**

Tunis El Manar University

5 PUBLICATIONS 2 CITATIONS

SEE PROFILE



**Baptiste Vrigneau**

University of Rennes

56 PUBLICATIONS 421 CITATIONS

SEE PROFILE



**Pascal Scalart**

University of Rennes

171 PUBLICATIONS 2,644 CITATIONS

SEE PROFILE

# Enhancing Energy Disaggregation with Attention-Based Neural Network

Nidhal Balti<sup>1</sup>, Baptiste Vrigneau<sup>1</sup>, and Pascal Scalart<sup>1</sup>

Univ of Rennes, IRISA, CNRS, Lannion, France,  
{nidhal.balti,baptiste.vrigneau, pascal.scalart}@irisa.fr

**Abstract.** Deep Neural Networks (DNNs) have been the subject of much research over the years, with a particular emphasis on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The purpose of this study is to evaluate the energy usage of single-source appliances, like smart meters, in the context of Non-Intrusive Load Monitoring (NILM). This paper introduces a novel workflow for Non-Intrusive Load Monitoring (NILM) by combining a regression subnetwork with a classification subnetwork. The architecture incorporates an RNN with attention mechanism in the regression and classification networks, drawing inspiration from successful applications in temporal attention. The experiments conducted on our own dataset called SmartSense shows that our proposed architecture outperforms the state-of-the-art under various conditions.

**Keywords:** Energy disaggregation, NILM, Environmental sensors data, RNN, CNN.

## 1 Introduction

Effective energy management has become more and more important in recent years due to growing environmental concerns about the effects of energy consumption and the increasing demand for sustainable solutions [1]. The increased consciousness in this area has led to significant progress in the field, most notably with the introduction of technologies for Intrusive Load Monitoring (ILM) and Non-Intrusive Load Monitoring (NILM).

Intrusive Load Monitoring involves the installation of specific hardware directly on individual appliances, such as sensors or meters, providing precise, device-specific data on energy consumption. This intrusive approach allows for accurate measurements, facilitating detailed insights into usage patterns and operational states. In contrast, Non-Intrusive Load Monitoring (NILM) requires only one smart meter installed on the external panel to disaggregate total energy consumption into appliance-level usage without additional sensors on each device. Leveraging signal processing and machine learning, NILM analyzes aggregate power consumption data from sources like smart meters, discerning unique electrical signatures to create detailed appliance-level profiles. While ILM demands a more involved setup, it offers unparalleled granularity in load monitoring. Both

ILM and NILM contribute to a smarter, more sustainable approach to energy consumption, empowering individuals and organizations to optimize usage, identify faults, and participate in efficient energy management practices within the context of smart grids and beyond.

In existing literature, the majority of works have centered mostly on electrical power data, including measurements like voltage, current, and active power. However, these studies often overlook the exploration of other critical data types, such as environmental variables. We propose that integrating wireless sensor data into the Non-Intrusive Load Monitoring (NILM) paradigm. By doing so, we unlock exciting opportunities for researchers to delve into the complex relationships between energy consumption and various environmental factors. Through seamless incorporation of data from wireless sensors—such as temperature, humidity, occupancy, and ambient light—we can achieve a more nuanced understanding of the correlations, dependencies, and causal relationships that significantly impact appliance-specific energy usage [2][3]. Additionally, our better understanding of the various factors that influence energy consumption helps to refine more comprehensive and sophisticated NILM models.

Our main contributions are three-fold:

1. We introduce a novel approach known as Semi-Intrusive Load Monitoring (SILM) which use environmental sensor data like CO<sub>2</sub>, humidity and temperature which provide a wealth of *a-priori* knowledge about an appliance state.
2. Accordingly, we propose a novel ON/OFF state NILM model that features a dual-RNN with attention architecture: one RNN (state-RNN) for capturing the appliance’s ON/OFF states, the other one (value-RNN) for predicting the power consumption for each state. A cross-entropy is used to state-RNN as a regularization item, which degrades the error in predicting the appliance’s power consumption.
3. Experimental results on our own real-world dataset called SmartSense show that our model generalizes well to unseen appliances where it achieves roughly 55% Mean Absolute Error (MAE) gain and 30% Root Mean Square Error (RMSE) gain over the state-of-the-art on the accuracy of energy disaggregation.

The paper is organized as follows. In Section 2, we briefly present an overview of the available NILM approaches in the literature. Section 3 presents our proposed method for energy disaggregation. Finally, we present evaluation methodology and we analyse the experiment results in section 4.

## 2 Related Work

Two decades ago, Hart proposed a non-intrusive method for disaggregating electrical measurements[4]. This method involved examining the overall load data to identify power consumption signatures for individual appliances. Unlike other approaches, Hart’s methodology did not require installing equipment inside the customer’s property. Instead, aggregated data for the building’s energy usage

could be obtained externally from the main electrical panel. The main objective of this process was to separate the whole-house energy data into its primary consumption components as shown in Fig 2. Non-Intrusive Load Monitoring (NILM) algorithms proved effective in achieving device-specific energy disaggregation using only aggregated data from a single measurement point. As a result, NILM remains an attractive approach for energy disaggregation.

A key area of interest in the field of Non-Intrusive Load Monitoring (NILM)

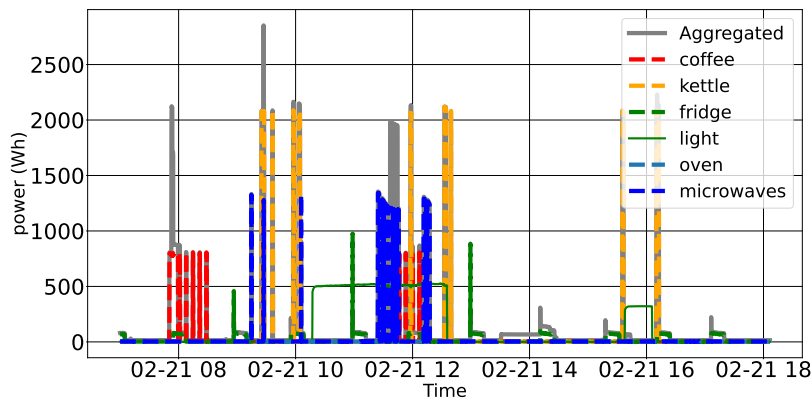


Fig. 1: Concept of typical NILM: decomposing of aggregated power consumption.

is the integration of environmental sensor data to improve the accuracy and contextual comprehension of energy use in residential and commercial settings. Environmental sensing and diverse information play a crucial role in addressing challenges encountered by current Non-Intrusive Load Monitoring (NILM) techniques [5]. However, incorporating these sources can introduce complexity. A more efficient approach involves integrating supplementary data with the overall power consumption of a building for load monitoring. Notably, the increasing number of intelligent sensors within buildings, serving various purposes, has made acquiring this information more feasible without additional setup. Consequently, there has been a recent emphasis on leveraging this method. As far as we are aware, no prior research has used environmental sensors in the NILM area.

Non-Intrusive Load Monitoring (NILM) has undergone significant advancements in recent years, driven by the exploration of signal processing, machine learning and deep learning. Signal processing approaches have enabled the extraction of appliance-specific features and patterns from aggregate power data [6], while machine learning techniques have leveraged supervised [7] and unsupervised [8], [9] learning algorithms to discern individual load signatures. Deep learning techniques, particularly deep neural networks, have emerged as powerful tools for

NILM owing to their capacity to automatically learn hierarchical representations from raw data. Convolutional Neural Networks (CNN) [10], Recurrent Neural Networks (RNN) [11], and their variants [12] have demonstrated superior performance in load disaggregation tasks. Researchers have proposed various architectures, such as sequence-to-point models [13], attention mechanisms [14], and generative adversarial networks (GANs) [15], to enhance the accuracy and generalizability of deep learning-based NILM systems.

RNNs with attention mechanisms [16] are an extremely powerful paradigm for sequence modeling and analysis. RNNs are designed to capture temporal dependencies in sequential data, making them ideal for tasks that require time-series information, such as natural language processing, speech recognition, and financial forecasting. The addition of attention mechanisms improves the capabilities of traditional RNNs by allowing the model to focus selectively on specific parts of the input sequence while assigning varying levels of importance to different elements. This dynamic attention mechanism allows the network to effectively evaluate the significance of each input, resulting in a more nuanced understanding of context and an improved ability to capture long-term dependencies. In essence, attention-based RNNs not only perform well in capturing temporal patterns, but also demonstrate an enhanced capacity to distinguish and prioritize relevant information within the input sequence, making them valuable tools across a wide range of applications.

Therefore, in this paper, we opted to use an RNN model with attention mechanism for energy disaggregation challenge, as they provide substantial advantages over training simpler neural networks at the edge. RNN networks excel in capturing long-term dependencies within sequential data, a critical capability for accurately modeling and disaggregating complex energy consumption patterns over time.

### 3 Proposed Method

#### 3.1 Problem formulation of energy disaggregation

The NILM approach aims to identify the states  $s_m(t)$  of an appliance  $m$  and estimate its power consumption  $y_m(t)$  from the total power consumption  $x(t)$  for  $m = 1, 2, \dots, M$ , where  $M$  indicates the number of appliances. We refer to this as a multi-task NILM problem, where given an observed aggregate power signal, unobserved states  $s(t)$  of electrical appliances are detected, and the corresponding power consumption  $y(t)$  estimated. The aggregate power consumption  $x(t)$  at time  $t$  can be expressed by Equation 1:

$$x(t) = \sum_{m=1}^M y_m(t) \cdot s_m(t) + \epsilon(t) \quad (1)$$

where  $\epsilon_t$  represents the contributions from appliances not accounted for and measurement's noise [17]. Specifically, the problem is formulated as follows.

Let  $X \in \mathbb{R}^T = \{x_1, \dots, x_T\}$  denote a set of input features derived from the aggregate power consumption of  $M$  appliances, and  $Y \in \mathbb{R}^{T \times M}$  indicates the

power signals of the associated appliances, where each appliance has  $k$  states denoted as  $s_m(t) = \{s_{m,1}(t), \dots, s_{m,k}(t)\}$  such that  $s_{m,k}(t) \in \{0, 1\}$ . The matrix  $S \in \mathbb{R}^{T \times M}$  indicates the associated multi-label states for the  $M$  appliances. Given  $\mathcal{D} = \{x(t), s(t) \mid t = 1, \dots, T\}$  datasets, the goal is to learn a multi-task model that predicts the state vector  $s(t) = \{s_m(t), \dots, s_M(T)\}$  and power signal vector  $y(t) = \{y_m(t), \dots, y_M(T)\}$  from the input aggregate power feature vector  $x(t)$ .

### 3.2 SILM workflow

This paper proposes a Semi-Intrusive Load Monitoring architecture (SILM) that uses power consumptions combined with environmental sensors data to detect individual power consumption of devices from aggregated power data as shown in Figure 2. In fact, the proposed architecture is composed mainly of 2 parts:

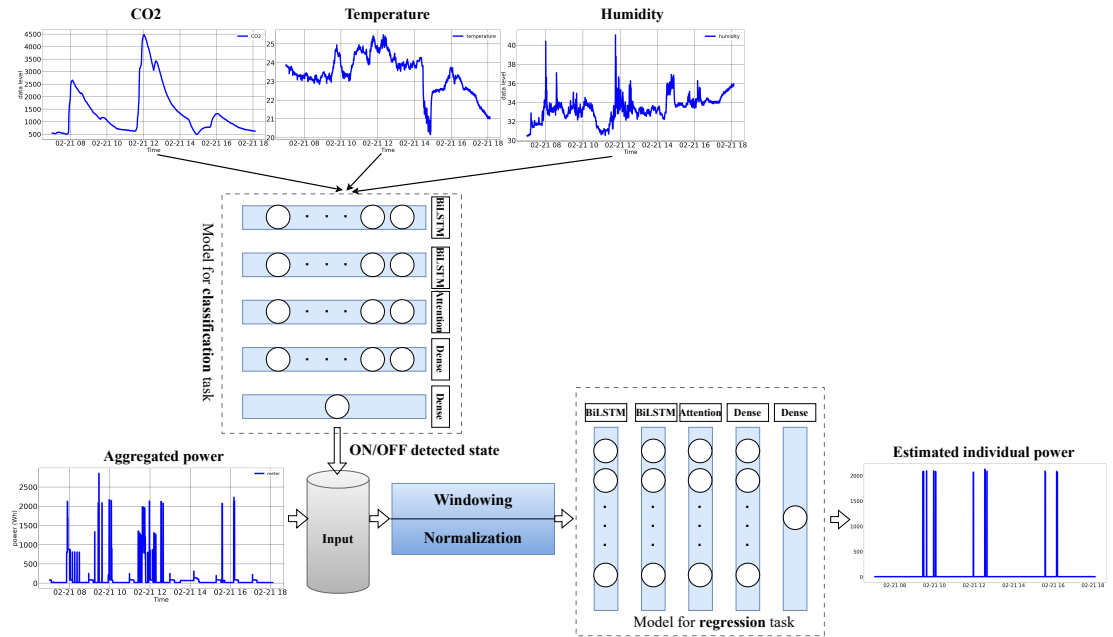


Fig. 2: Architecture of RNN with attention for energy disaggregation and ON/OFF states detection.

1. The first part consists of a classification task that intends to detect the On/Off state of appliances from environmental sensors data and then feed them to the input of the next step;
2. The second part is a regression task that aims to estimate the individual power consumptions from the aggregate power data and the detected states.

### 3.3 Detection of ON/OFF states

The majority of appliances have a minimum ON-power threshold, which defines the amount of energy required for operation. If an appliance’s energy consumption is less than this on-power threshold, it is considered to be in an off state; otherwise, it is considered to be on. For example, a kettle typically requires 2000 watts to be operational, whereas a coffee maker machine only needs 800 watts. In this paper, we address the problem of On/Off detection by using temperature, humidity, and CO2 levels data from environmental sensors. This involves estimating whether an appliance is currently in an On or Off state based on the recorded sensor data. The learning framework based on classification, as shown in Figure 2, uses a training method for a binary classifier to tackle the problem of on/off detection. To be more precise, an appliance’s energy readings are first binary-coded according to the appliance’s ON-power threshold. Then, using the binary-coded appliance readings as target values and the wireless sensors data as inputs, a binary classifier is trained. Lastly, the trained classifier is used to predict the appliance’s on/off state sequence when it is given a fresh set of environmental sensors data. We choose an RNN with attention mechanism as the binary classifier with an architecture as follows:

1. Input (sequence length  $L$  determined by the appliance duration)
2. BiLSTM (bidirectional LSTM with 128 units, and tanh activation function)
3. BiLSTM (bidirectional LSTM with 256 units, and tanh activation function)
4. Attention (single layer feed-forward neural network with 128 units, and tanh activation function)
5. Dense (fully connected layer with 128 units, and ReLU activation function)
6. Dense (fully connected layer with 1 unit, and sigmoid activation function)
7. Output (ON/OFF state)

To train an efficient binary classifier for a specific appliance in a sequence-to-point model, the final layer of the model consists of a fully connected layer followed by a sigmoid function, indicating the probability of the appliance being in the "On" state. Assuming the output and its corresponding target values are represented as  $\hat{y}$  and  $y$  respectively, the loss is computed using Binary Cross-Entropy (BCE) loss which is calculated using the formula:

$$BCE(y, \hat{y}) = -\frac{1}{T} \sum_{t=1}^T (y(t) \cdot \log(\hat{y}(t)) + (1 - y(t)) \cdot \log(1 - \hat{y}(t))) \quad (2)$$

where  $T$  represents the total number of samples within one recording,  $t$  signifies the index of a specific sample in the dataset ranging from 1 to  $T$ ,  $y(t)$  is the true label (ground truth) for the sample at a time  $t$ , and  $\hat{y}(t)$  represents the predicted probability that the sample at time  $t$  belongs to class ON or OFF.

### 3.4 Energy disaggregation

Energy disaggregation aims to estimate the energy usage of individual appliances based on the recordings of the mains power meter that measures the total energy

consumption and the detected states provided by the binary classifier. For this specific task of energy disaggregation, we trained another RNN with attention mechanism model which had the following architecture:

1. Input (sequence length  $L$  determined by the appliance duration)
2. BiLSTM (bidirectional LSTM with 128 units, and tanh activation function)
3. BiLSTM (bidirectional LSTM with 256 units, and tanh activation function)
4. Attention (single layer feed-forward neural network with 128 units, and tanh activation function)
5. Dense (fully connected layer with 128 units, and ReLU activation function)
6. Dense (fully connected layer with 1 unit, and linear activation function)
7. Output (power consumption estimation)

The inputs are sequences of aggregate energy recordings  $x(t)$  and detected states  $s(t)$  while the target values are sequences of appliance energy readings  $y$ . Assuming an output and the corresponding target value are denoted as  $\hat{y}$  and  $y$  respectively, the loss can then be calculated using Mean Squared Error (MSE):

$$MSE(y, \hat{y}) = \frac{1}{T} \sum_{t=1}^T (y(t) - \hat{y}(t))^2 \quad (3)$$

where  $T$  represents the total number of samples within one recording, with  $t$  serving as the index of a specific sample ranging from 1 to  $T$ . The MSE loss quantifies the average squared difference between the true and predicted values across all samples, providing a measure of how well the model is performing in a regression task.

## 4 Experiments

### 4.1 Experimental Setup

**Dataset and Performance Metrics** The dataset used in this article is called SmartSense developed in our lab for the purpose of this work. It comprises a comprehensive collection of data including energy consumption and environmental sensor readings from the IRISA lab in France. These readings were recorded at a low frequency, approximately every 10 seconds, over a period exceeding three years. Notably, the dataset involves a wide range of data, including both lab-level energy usage (aggregate recordings) and appliance-level energy usage (appliance recordings) for more than 10 different types of appliances, in addition to environmental sensors data such as humidity, CO2 level, temperature from each office. For the purposes of this article, our focus is primarily on the energy disaggregation of two specific appliance types: kettle and coffee maker as these are the appliances most used across our lab’s break room. For more information on the hardware, software, and details of the data collection process, please visit the project’s official website<sup>1</sup>. The dataset files are easily accessible via this storage space<sup>2</sup>. A summary of SmartSense features is shown in Table 1.

We quantitatively evaluate the performance of our proposed method with both

<sup>1</sup> <https://smartsense.inria.fr/>

<sup>2</sup> <https://files.inria.fr/smartsense/>

Table 1: SmartSense dataset features.

Measurement device	Data type	Sampling Rate	Duration	Appliances
EcoCompteur	global power data	7.5 Hz		coffee maker, kettle, fridge
ZigBee meter	indiv power data	100 mHz	3 years	oven, light
Sensors node	sensors data	1 Hz		PC, printer

regression and classification metrics. To assess the performance of the regression task, we adopt three standard metrics used in NILM, namely the mean absolute error (MAE), Root Mean Squared Error (RMSE) and Normalized Disaggregation Error (NDE) [18]. The MAE and RMSE quantify the error in predicted power at every time point, given by

$$MAE = \frac{1}{T} \sum_{t=1}^T |y(t) - \hat{y}(t)| \quad \text{and} \quad RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T |y(t) - \hat{y}(t)|^2} \quad (4)$$

where  $T$  represents the total number of samples within one recording,  $\hat{y}(t)$  denotes the predicted power at time  $t$ , and  $y(t)$  is the true power at time  $t$ . The NDE metric measures the normalized error of the squared difference between the prediction and the ground truth, defined as

$$NDE = \frac{\sum_{t=1}^T (y(t) - \hat{y}(t))^2}{\sum_{i=1}^T (y(t))^2} \quad (5)$$

In order to measure how accurately each appliance is running in ON/OFF states, we use classification metrics such as precision, recall and F1-score [19]. Due to the extreme imbalance in the dataset, we will consider only the recall metric. For instance, the kettle is only active for about 1% of the time. The F1 score is interpreted as the harmonic average of precision ( $P$ ) and recall ( $R$ ):

$$F1 = \frac{2 \cdot P \cdot R}{P + R}, \quad P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN} \quad (6)$$

where  $TP$ ,  $FP$ , and  $FN$  stand for true positive, false positive, and false negative, respectively. An appliance is considered "ON" when the active power is greater than a certain threshold and "OFF" when it is less than or equal to the same threshold. In our experiments, we employ a threshold of 20 Watts for labeling the disaggregated loads. Precision, recall, and F1-score return a value between 0 and 1, where a higher number corresponds to better classification performance.

**Data preprocessing** The SmartSense dataset has different sampling rates based on the collected data, as shown in table 1. As a result, resampling becomes necessary in the preparation of training data. To reduce the introduction

of missing values due to resampling, a resampling rate of 10 seconds was implemented. Furthermore, we addressed data gaps using forward-filling, assuming that the gaps were caused by data transmission issues. Finally, each appliance type, the aggregate, and the environmental sensors data were normalized by subtracting mean values and dividing by the corresponding standard deviations. The training phase is performed with a sliding window technique over the aggregated signal, using overlapped windows of length  $L$  with hop size equal to 1 sample. As stated in reference [20], it is critical to ensure that the window size for input and output pairs is large enough to cover the entire activation of an appliance. However, it should not be too large to accommodate contributions from other appliances.

## 4.2 Experimental Results

The dataset used in the following experiments covered a complete 30-day period to ensure that the neural network was adequately characterized and not over-trained. Subsequently, to mitigate over-fitting and validate the network’s performance robustly, we partitioned the dataset into distinct subsets for training and testing, each encompassing 15 days. This division allowed us to assess the network’s ability to generalize beyond the training data and to evaluate its performance on unseen test data effectively. Throughout the training process, we closely monitored both the training and validation loss curves to detect signs of over-fitting and to adjust model parameters as necessary. This dataset includes the overall power consumption as well as the particular power usage trends for each appliance. We are primarily concerned with breaking down the power usage of the most used appliances in the lab’ break room such as the kettle, and coffee maker.

To evaluate our approach, we first implement the basic NILM pipeline which

(a) Confusion Matrix of coffee maker states detection

True Class	Predicted Class		
	OFF	ON	Support
OFF	99726	32464	132190
ON	141	1229	1370
Precision Recall F1 Score			
OFF	1.00	0.90	0.86
ON	0.04	0.90	0.07

(b) Confusion Matrix of kettle states detection

True Class	Predicted Class		
	OFF	ON	Support
OFF	96221	36348	132569
ON	260	731	991
Precision Recall F1 Score			
OFF	1.00	0.73	0.84
ON	0.02	0.72	0.04

Fig. 3: Confusion matrices for state detection from CO2, humidity, and temperature data.

uses only the aggregate power data as input and a model based regression task.

Table 2: Evaluation of disaggregation performance using basic NILM and our proposed method (in bold)

Appliances	RMSE	MAE	NDE
Kettle	103.14 <b>69.82</b>	12.02 <b>5.2</b>	0.96 <b>0.65</b>
Coffee maker	88.9 <b>59.61</b>	15.07 <b>7.55</b>	1.11 <b>0.74</b>

We choose to implement the RNN model with attention available in the toolkit NILMTK[21] which is an adaptable software framework designed specifically for NILM research. The architecture of this model is the same as the model described in section 3.4.

Table 2 summarizes the outcomes of our experiments for both our approach and the basic NILM pipeline. The table includes results for three metrics that were calculated for each appliance separately. The values reported consist of the predicted value of the target appliance consumption and the estimated consumption. Tables 3a and 3b shown in figure 3 in present the results of the appliances ON/OFF states detection by measuring the F1-score, the recall and precision.

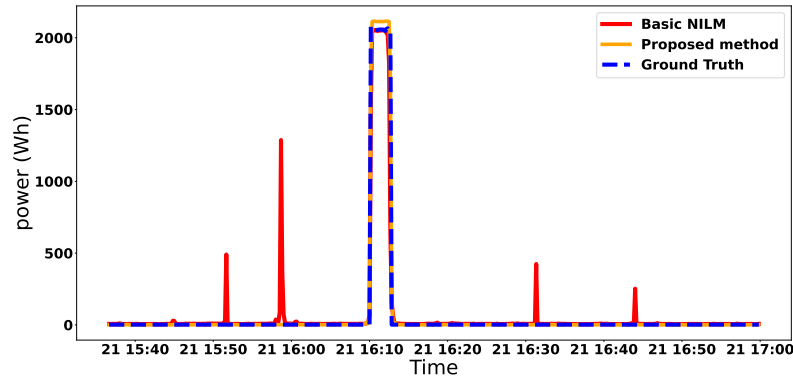


Fig. 4: Comparison of kettle power consumption predictions between our proposed method and the basic NILM approach.

According to table 2, our approach surpassed the conventional NILM disaggregation with an error decrease by 30% for RMSE and above 55% for MAE. These results demonstrates that incorporating sensors data about the environment can enhance NILM disaggregation. As shown in Figure 4, the basic NILM method produces false ON states detection when compared to the ground truth,

while our approach accurately identifies the correct ON states. These results demonstrate that incorporating an *a-priori* information on appliances states can enhance NILM disaggregation performance.

## 5 Conclusion

This paper presents a new NILM approach for disaggregation energy data from one smart meter and three environmental sensors (CO<sub>2</sub>, temperature and humidity) into individual appliance-level. This approach is based on a two-steps procedure: firstly, we detect the appliances ON/OFF states by training an RNN attention as a binary classifier. Secondly, these predicted states as well as the aggregated data are then fed to another RNN attention which acts as a disaggregator to estimate the individual power data.

Experimental evaluation on the SmarSense dataset shows good disaggregation performance for the coffee maker and the kettle appliances. Overall, the proposed method can accurately capture the dependencies between the appliance’s usage, efficiently detect states appliances and estimate their corresponding power consumption. Compared to existing DNNs approaches for NILM, the advantage of the proposed approach comes from its ability to provide both appliance states and power consumption values from environmental sensors and the total energy data.

Future work should explore the model evaluation with a larger number of appliances such as printers and laptops as well as other environmental sensors available in the SmartSense dataset like the audio and light sensors.

## References

1. International Energy Agency. *World Energy Outlook 2019*. OECD Publishing: Paris, France, 2019.
2. Abdullah Alsalemi, Yassine Himeur, Fayal Bensaali, Abbes Amira, Christos Sardinios, Iraklis Varlamis, and George Dimitrakopoulos. Achieving domestic energy efficiency using micro-moments and intelligent recommendations. *IEEE Access*, 8:15047–15055, 2020.
3. Weizheng Hu, Yonggang Wen, Kyle Guan, Guangyu Jin, and King Jet Tseng. itcm: Toward learning-based thermal comfort modeling via pervasive sensing for smart buildings. *IEEE Internet of Things Journal*, 5(5):4164–4177, 2018.
4. G.W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.
5. Ahmed Zoha. *Load Disaggregation and Monitoring in a Smart Office Space*. University of Surrey (United Kingdom), 2014.
6. Kanghang He, Lina Stankovic, Jing Liao, and Vladimir Stankovic. Non-intrusive load disaggregation using graph signal processing. *IEEE Transactions on Smart Grid*, 9(3):1739–1747, 2016.
7. Linda Farinaccio and Radu Zmeureanu. Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses. *Energy and Buildings*, 30(3):245–259, 1999.
8. Hugo Gonçalves, Adrian Ocneanu, Mario Bergés, and R Fan. Unsupervised disaggregation of appliances using aggregated consumption data. In *The 1st KDD work-*

- shop on data mining applications in sustainability (SustKDD)*. ACM San Diego, CA, USA, 2011.
9. J Zico Kolter and Tommi Jaakkola. Approximate inference in additive factorial hmms with application to energy disaggregation. In *Artificial intelligence and statistics*, pages 1472–1482. PMLR, 2012.
  10. David Murray, Lina Stankovic, Vladimir Stankovic, Srdjan Lulic, and Srdjan 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)*, pages 8330–8334. IEEE, 2019.
  11. Wan He and Ying Chai. An empirical study on energy disaggregation via deep learning. In *2016 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2016)*, pages 338–342. Atlantis Press, 2016.
  12. Maria Kaselimi, Nikolaos Doulamis, Athanasios Voulodimos, Eftychios Protopapadakis, and Anastasios Doulamis. Context aware energy disaggregation using adaptive bidirectional lstm models. *IEEE Transactions on Smart Grid*, 11(4):3054–3067, 2020.
  13. Chaoyun Zhang, Mingjun Zhong, Zongzuo Wang, Nigel Goddard, and Charles Sutton. Sequence-to-point learning with neural networks for non-intrusive load monitoring. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.
  14. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
  15. Maria Kaselimi, Nikolaos Doulamis, Athanasios Voulodimos, Anastasios Doulamis, and Eftychios Protopapadakis. Energan++: A generative adversarial gated recurrent network for robust energy disaggregation. *IEEE Open Journal of Signal Processing*, 2:1–16, 2020.
  16. Zhenrui Yue, Camilo Requena Witzig, Daniel Jorde, and Hans-Arno Jacobsen. Bert4nilm: A bidirectional transformer model for non-intrusive load monitoring. In *Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring*, pages 89–93, 2020.
  17. Christoph Klemenjak, Stephen Makonin, and Wilfried Elmenreich. Towards comparability in non-intrusive load monitoring: On data and performance evaluation. In *2020 IEEE power & energy society innovative smart grid technologies conference (ISGT)*, pages 1–5. IEEE, 2020.
  18. Lucas Pereira and Nuno Nunes. Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—a review. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, 8(6):e1265, 2018.
  19. László A Jeni, Jeffrey F Cohn, and Fernando De La Torre. Facing imbalanced data—recommendations for the use of performance metrics. In *2013 Humaine association conference on affective computing and intelligent interaction*, pages 245–251. IEEE, 2013.
  20. Jack Kelly and William 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*, pages 55–64, 2015.
  21. Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. Nilmtk: An open source toolkit for non-intrusive load monitoring. In *Proceedings of the 5th international conference on Future energy systems*, pages 265–276, 2014.