Adaptive Weighted Recurrence Graphs for Appliance Recognition in Non-Intrusive Load Monitoring

Anthony Faustine, Lucas Pereira, and Christoph Klemenjak

***Abstract*—To this day, hyperparameter tuning remains a cumbersome task in Non-Intrusive Load Monitoring (NILM) research, as researchers and practitioners are forced to invest a considerable amount of time in this task. This paper proposes adaptive weighted recurrence graph blocks (AWRG) for appliance feature representation in event-based NILM. An AWRG block can be combined with traditional deep neural network architectures such as Convolutional Neural Networks for appliance recognition. Our approach transforms one cycle per activation current into an weighted recurrence graph and treats the associated hyper-parameters as learn-able parameters. We evaluate our technique on two energy datasets, the industrial dataset LILACD and the residential PLAID dataset. The outcome of our experiments shows that transforming current waveforms into weighted recurrence graphs provides a better feature representation and thus, improved classification results. It is concluded that our approach can guarantee uniqueness of appliance features, leading to enhanced generalisation abilities when compared to the widely researched V-I image features. Furthermore, we show that the initialisation parameters of the AWRG's have a significant impact on the performance and training convergence.**

***Index Terms*—Non-Intrusive Load Monitoring, Load Disaggregation, Appliance Recognition, Weighted Recurrence Graphs, Recurrence Plots, V-I Trajectories, Convolutional Neural Networks, Deep Neural Networks**

I. Introduction

In recent times, wide portions of the world has witnessed rapidly increasing energy use in buildings, residential and industrial, mainly due to rapid urbanisation, economic growth and population growth. In Africa for example, energy consumed in buildings is estimated at 56% of the total national electricity consumption [1] while in Europe, 25% of the final energy consumption is accountable to households [2]. This rise in energy use is of great concern to climate and sustainability challenges and therefore, calls for effective and efficient energy saving in buildings. One way to manage energy use in buildings is through monitoring the energy consumption of end-use appliances. Information on appliance-specific energy consumption can play an essential role in improving energy consumption awareness of households, which is likely to stimulate energy-saving behaviour [3]–[5]. It was shown that monitoring individual loads at consumer premises plays an essential role in the design of customised energy efficiency and energy demand management strategies [6]. The use of power sensing technologies, such as smart-plugs and smart meters, has gained popularity recently with the aim of making information on power consumption available to households [2], [7]. These sensing devices can monitor aggregate power consumption at different sample rates, thereby enabling the development of a remote energy management system at the customer premises. With the help of cost-effective technologies such as Non-intrusive Load Monitoring (NILM), aggregate power data can be further used to facilitate appliance-specific energy consumption in buildings.

A. Non-Intrusive Load Monitoring

NILM, also known as load disaggregation, is a computation technique that extracts appliance-specific energy consumption from aggregate consumption data, monitored at a single point by utilising measurement equipment such as smart meters [8], [9]. Unlike the traditional approach, which requires deploying sensing infrastructure for each appliance in a building, NILM techniques are cost-effective, non-intrusive, and can easily be integrated into buildings. In general, NILM algorithms fall in one of two categories: 1) event-based, and 2) event-less approaches [10]. The former approaches seek to disaggregate appliances by means of detecting and classifying their individual transitions in the aggregated signal. In contrast to that, the latter approaches attempt to match each sample of the aggregated signal to the consumption of one or more appliances. Deep Learning is seen as an enabler in both approaches. More precisely, in event-based NILM through Convolutional Neural Networks (CNNs), e.g., [11], [12], and Recurrent Neural Networks (RNNs) in event-less, e.g., [13], [14]. The contributions presented in this paper are categorised under event-based approaches.

An event-based NILM pipeline usually consists of the following steps: *event detection*, *feature representation*, *appliance recognition* and *energy estimation*. The task of the event detection step is to detect changes in the aggregate power signal due to one or more appliance being activated. After that, unique electrical features are extracted in the feature representation stage. The extracted features should be distinctive enough to allow identifying one appliance from another. According to [15], an ideal feature representation should be able to minimise the difference of instances in the same class and maximise the contrast of instances in different categories. However, finding

A. Faustine was with the Department of Information Technology, University of Ghent, Gent, B-9052, Belgium. e-mail: sambaiga@gmail.com

L. Pereira is with ITI, LARSyS, Técnico Lisboa, Portugal.

C. Klemenjak is with the Institute of Networked and Embedded Systems, University of Klagenfurt, Lakeside B10, 9020 Klagenfurt, Austria.

Manuscript submitted January, 2020.

an efficient set of features for appliance classification is still an open challenge in NILM [16]. Once the feature representation has been extracted, it can be used for identifying the appliance category and its status in the appliance recognition stage. Appliance classification, also known as load identification, uses machine learning techniques to analyse the pattern of the feature representation, classify them into the respective appliance category.

Several studies have investigated the use of features-representation extracted from Voltage-Current (V-I) waveforms for appliance recognition in NILM e.g, [11], [12], [15], [15], [17]–[25]. The use of V-I based features for appliance classification was first introduced in [17], where eight features were hand-engineered from one cycle of steady-state voltage and current (e.g., are and self-interceptions). The basic assumption for using of V-I features was that after normalisation, independently of the brand and model, the shapes for the same types of appliances would be similar. Hence, providing a universal method for feature extraction, that could be used by classic machine learning algorithms such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), and k Nearest Neighbours (kNN). As for this writing, the original set of eight features evolved to a set of 22 features, including steady-state and transient features [20].

Other studies have shown that transforming the V-I trajectory into a visual representation is computationally efficient and improves classification performance [21], [22]. But more importantly, it provides a convenient way of leveraging advanced machine learning techniques such as CNNs to extract detailed features from the original V-I waveform that cannot be engineered by hand [11], [12], [23].

B. (Weighted) Recurrence Graphs

Despite the efforts, the performance of most V-I based approaches is still unsatisfactory as V-Is are not distinctive enough to recognise appliances that fall into the same category, e.g., purely resistive loads. This is because the V-Is have the same shape independently of the current magnitude, substantially reducing their discerning ability. Furthermore, most of the existing works [21]–[23] have been evaluated in noise-free sub-metered data, which is very likely to result in over-optimistic classification results [26]. Against this background, this paper presents an appliance recognition approach that uses the current waveform (CW) generated from the aggregated measurements. As with V-Is, the CW is converted into an image-like representation, in this case using a Recurrent graph (RG), also known as recurrence plot. RG analyse signal dynamics in phase space to reveal repeating and non-linear patterns [27] and have been used extensively for feature representation in time series classification problems [27]–[29]. Unlike the V-I image-representation, the RG feature representation uses a distance-similarity matrix to represent and visualise structural patterns in the CW. As a consequence, RG feature representation depends also on the magnitude of the current signal. The use of RG for the characterisation of appliances features was first proposed in [30]. However, similarly to other RG methods for classification, the proposed approach uses a compressed distance-similarity matrix representing all recurrences in the form of a binary pattern.

Binarizing the RG through thresholding can lead to information loss and therefore, degrade classification performance. Hence, to avoid information loss and improve classification performance, a Weighted Recurrence Graph (WRG) that gives a few more values instead of the binary output is introduced in [31]. However, both the RG and proposed WRG have hyper-parameters that influence classification performance and need to be determined somehow. To the best of our knowledge, there are no appropriate methods for selecting these RG hyper-parameters. Existing approaches found in literature are rather heuristic or use hyper-parameter search. Instead, the proposed approach handles the (W)RG hyper-parameters as learnable parameters, much like normal neural network weights. To this end, we introduce the Adaptive Weighted Recurrence Graph (AWRG) method, where hyper-parameters are adjusted by computing the gradient with respect to the other neural network parameters.

The AWRG technique is evaluated on two datasets, namely the Plug-Load Appliance Identification Dataset (PLAID) [32], and the Laboratory-measured Industrial Load of Appliance Characteristics dataset (LILACD) [33]. The PLAID dataset contains aggregated voltage and current measurements of residential appliances obtained at 30kHz, whereas the LILACD dataset contains three-phase aggregated measurements sampled at 50kHz in an industrial environment. To classify the generated image-like feature representations, we apply CNNs, which have been applied successfully for image classification tasks [11], [12], [34].

C. Research Contributions and Paper Organisation

The main contributions of this paper are threefold:

1) We present the AWRG method for feature representation in NILM. The proposed AWRG block is included in the learning pipeline as part of the end-to-end feature learning with deep learning networks. Ultimately, this allows fine-tuning of the WRG and consequently, improves the discerning power of the features and classification performance.
2) We conduct an extensive evaluation on two datasets that were collected in residential and industrial environments. In contrast to other approaches that use sub-metered data, we test our method on aggregated power measurements, which is said to be more realistic. Furthermore, we contrast the multi-dimension three-phase system in industrial settings and the single-phase system in residential buildings, which has received little attention in literature.
3) We benchmark the proposed AWRG feature representation with V-Is, which are its most direct competitor. Ultimately, we show that the proposed method significantly improves the classification performance in both, the single and three-phase case.

For the sake of reproducibility, we release an implementation of our contributions in an online code repository[1].

[1]https://github.com/sambaiga/AWRGNILM

The remainder of this paper is organised as follows: Section II introduces the methods utilised in this paper. In Section III, we describe the experimental design. Section IV presents the outcome of our evaluations. Section V summarises the contributions of this paper and provides an outlook on future work.

II. Methods

The goal of appliance classification in NILM is to identify active appliances $k \in \{i = 1, 2, \dots M\}$ from the aggregate signal x_t where M indicates the number of appliances. This can be seen as a multi-class problem. The aggregate current signal x_t at any time t is assumed to be $x_t = \sum_k^M y_t^{(k)} \cdot \mathbf{s}_t^{(k)} + \sigma_t$ where $y_t^{(k)}$ is the contribution of appliance k, $\mathbf{s}_t^{(k)} \in \{0, 1\}$ is one hot encoding for appliance labels, and σ_t represents both any contribution from appliances not accounted for and measurement noise. Suppose we extract x_t in brief windows of time containing only a single activation event, then the electrical feature $\hat{X}_t$ that causes an event can be obtained by considering the difference of x_t before and after an event. Therefore, given $D = \{\hat{\mathbf{x}}_t, \mathbf{S}_t | t = 1, \dots, T\}$ set of feature representation and their corresponding one-hot encoding of appliances labels, the goal is to learn a multi-class classifier that predicts the appliance class $\mathbf{s_t}$ from an input feature vector $\hat{x}_t$. This section introduces the proposed AWRG by first describing how activation current and voltage are obtained from the aggregate measurements for single and three-phase data. Finally, it presents the deep neural network and training procedure used to build the classification model.

A. Feature Extraction from Aggregate Measurements

We consider appliance features extracted in brief windows of time, containing only one event feature derived from aggregate power measurements, as it allows us to distinguish appliances based on their start-up events. We define an activation current i and voltage v to be a one-cycle steady-state signal extracted from aggregate current waveform in a brief time after state transition. To obtain the activation current from aggregate measurements, we measure N_s complete cycles of current and voltage before $\{\mathbf{i}_k^{(b)}, \mathbf{v}_k^{(b)}\}$ and after $\{\mathbf{i}_k^{(a)}, \mathbf{v}_k^{(a)}\}$ the transition, where $k \in \{1, 2, 3\}$ for three phase industrial data and $k = 1$ for single-phase residential data set.

In this work, the current and voltage before and after the transition are respectively $N_s = 10$ and $N_s = 2$. For PLAID $N_s = 2$ was chosen based on [11], for LILACD dataset $N_s = 10$ was chosen empirically after we had analysed the data. The N_s cycles correspond to steady-state behaviour and are equivalent to $T_s \times N_s$ samples where $T_s = \frac{f_s}{f}$, f_s is sampling frequency and f is the mains frequency.

The extracted cycles are aligned at zero-crossing of the voltage and thereafter, one-cycle activation current before $i_k^{(b)}$ and after $i_k^{(a)}$ event is extracted. Since the current waveform has an excellent additive property, the activation current can be extracted from the aggregate signal by subtraction [15]. The activation current i is then calculated as follows: $i_k = i_k^{(a)} - i_k^{(b)}$ and $v_k = v_k^{(a)}$ if the event is caused by activation of appliance, and $i_k = i_k^b - i_k^{(a)}$ if the event is caused by de-activation of appliance as illustrated in Fig. 1 [11], [31].

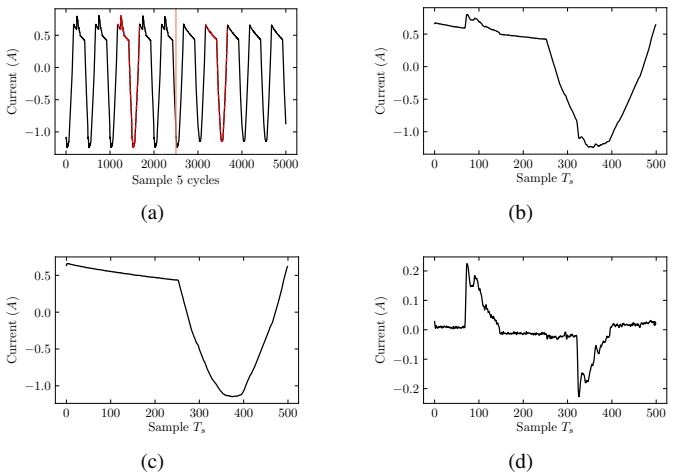

Fig. 1. Extraction of activation current i and voltage v as the result of CFL being switched on a) Aggregated current signal b) voltage before an event c) voltage after an event d) Current before i^b e) Current after i^a f) Activation current i obtained by taking the difference between i^a - i^b

B. Adaptive Weighted Recurrence Graph

The RG feature representation uses a distance similarity matrix $D_{w \times w}$ to represent and visualise structural patterns in the signal. The distance similarity matrix provides a relationship metric between each element in the time series [35]. The RG is one of the techniques used to encode time series data into visual representation [27]–[29].It uses a compressed distance similarity matrix to represent and visualise structural patterns in the signal [31]. The main idea is to reveal in which points trajectories return to a previous state. It is usually formulated as

$$R_{i,j} = \begin{cases} 1 & \text{if} \quad d_{i,j} \geq \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where ϵ is the recurrence threshold, $d_{i,j}$ is the distance similarity function such as Euclidean norm between data point x_i and x_j. This matrix contains both, patterns and information, that are not always very easy to see and interpret [36]. To improve classification performance and avoid information loss as the result of binarisation, we apply WGR because it gives a few more values instead of an binary output [31], defined as

$$R_{i,j} = \begin{cases} \delta & \text{if} \quad d_{i,j} > \delta \cdot \epsilon \\ n & \text{if} \quad n \cdot \epsilon \leq d_{i,j} \leq (n+1) \cdot \epsilon \end{cases} \tag{2}$$

where $0 \leq n \leq \delta$, $\delta \geq 1$ and $d_{i,j} = ||x_i - x_j||_p$. This improvement comes at the cost of additional hyper-parameters σ, to be selected. Unlike [31], we treat recurrence hyper-parameters $\theta = \{\epsilon, \delta\}$ as learn-able parameters like normal neural network weight and introduce the Adaptive Weighted Recurrence Graph (AWRG) block as depicted in Fig. 2.

The Piecewise Aggregate Approximation (PAA) technique is used to reduce the dimension of the decomposed signal $\mathbf{i}_a$ and $\mathbf{i}_f$ from T_s to a predefined size w where w is the

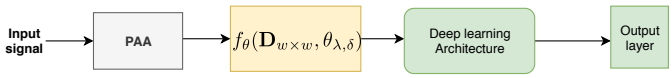

Fig. 2. Proposed recurrence block architecture for generating recurrence graph.The deep learning architecture box can be replaced with any architecture such as CNN.

embedding size. PAA is a dimension reduction method for high-dimensional time series signals [37]. This is a crucial pre-processing step as it reduces the high-dimensionality of the extracted activation current feature with minimal information loss. As regards the PAA block, the embedding size w is the hyper-parameter which needs to be selected carefully. Empirically, it was found that the choice of w does not significantly influences the classification performance. However, large values of w impact the learning speed as discussed in our experimental results. The adaptive weighted adjacency matrix sub-block generates recurrence graph $\mathbf{G}$ with weighted adjacency matrix $\mathbf{R}_{w\times w}$ by first computing the distance similarity matrix $\mathbf{D}_{w\times w}$ such that $d_{i,j} = \sqrt[p]{|x_i - x_j|^p}$. The recurrence matrix $\mathbf{R}_{w\times w}$ is then obtained as follows:

$$\mathbf{R}_{w\times w} = f_\theta(\mathbf{D}_{w\times w})) \tag{3}$$

where

$$f_\theta(D_{i,j}) = \begin{cases} \delta & \text{if} \quad \tau > \delta \\ \tau & \text{otherwise} \end{cases} \text{ and } \tau = \lfloor \frac{D_{i,j}}{\epsilon} \rfloor \tag{4}$$

and $\lfloor . \rfloor$ is the floor function. To avoid the vanishing gradient problem on ϵ, we apply parametrisation $\lambda = \frac{1}{\epsilon}$. Since the value of $\delta \geq 1$, and $0 \leq \frac{1}{\lambda} \leq 1$, it is necessary to adopt appropriate initialisation values for these learn-able parameters for optimal performance and faster convergence. In this paper, we initialise the value of $\delta = 10$ and that of $\lambda = 10$. We experimented with different initialisation parameters and found that these value achieve faster convergence and higher classification performance.

C. Neural Networks and Training Procedure

This works considers CNNs, a specific kind of neural networks for processing visual data. CNN leverages local connectivity and equivariant representations that make it useful for computer vision tasks. The CNN network used in this work consists of three-stage CNN layer each with $16, 32$ and 64 feature maps, 5×5 filter size, 2×2 stride size and ReLU activation function as depicted in Fig. 3. Two fully connected linear layers follow the last layer with a hidden size of 1024 and K, respectively, where the number of appliances available determines the number of classes. The final predicted class is obtained by applying softmax activation function. Since the problem at hand is multi-class classification, the objective function used is the Cross-Entropy Loss defined as $\mathcal{L}_\theta(y, p) = -\sum_{i=1}^{M} y_i \cdot \log p_i$. The CNN network is trained using mini-batch stochastic gradient descent (SDG) with a momentum of 0.9, a learning rate of 1^{-3}, and a batch size of 16. To avoid over-fitting, early stopping with patience is used where the training model is terminated once the validation performance does not change after 20 iterations.

III. Experimental Design

A. Datasets

The proposed method is tested on the aggregated PLAID [32] and the LILACD [33] datasets. The former contains single-phase aggregated voltage and current measurements of 12 different domestic electrical appliances. The dataset is sampled at 30 kHz and contains 1314 measurements obtained when multiple appliances were active simultaneously. The LILACD dataset contains three-phase aggregated and sub-metered current and voltage measurements sampled at 50 kHz for 16 different appliance types (industrial and home appliances). In this study, aggregated measurement data that contain measurements of more than one concurrently running appliances is used. Since the LILACD dataset is multi-dimensional, a separate activation current for each current phase is an input feature. We specifically consider both residential and industrial datasets because there is limited research about using NILM for industrial environment [38]. Compared to residential buildings, the industrial environment offers unique characteristics for existing NILM techniques. Industrial settings often consist of three-phases machines equipped with a variety of heavy equipment such as large motors [33], yet it can also comprise single-phase appliances that are common in residential buildings.

Throughout the paper, we use the following abbreviations: Compact fluorescent lamp (CFL), Bulb (ILB), Kettle (KT), Airconditiner(AC), SolderingIron (SLD), CoffeeMaker (CM), Hair-dryer (Dryer), Raclette (RC), and FridgeDefroster (FRZ), 1-phase-async-motor (1P-Motor), and Drilling-machine (DRL) for single phase appliances. The three-phase industrial appliances includes Dumper-machine (3P-DPM), 2x=Freq-conv-squirrel-3-2x (3P-FCS), Squirrel-3-async (3P-SQL), and Squirrel-3-async-2x(3P-SQL-2x). The distribution of appliances on the PLAID and LILAC aggregated measurements is depicted on Fig. 4.

The activation and deactivation's (events) in PLAID and LILACD dataset are labelled by default, making it easy to calculate the activation current signal of the appliance causing the event. However, after analysing the LILACD data, it was found that some events were not correctly labelled. As a result, we corrected the mislabelled events and created a sub-aggregated LILACD dataset with 1324 correct activation and deactivation's labels.

B. Performance Metrics

To quantitatively evaluate the classification performance, we use macro-averaged F_1 score, zero-loss score (ZL) and the Matthews correlation coefficient (MCC). Also, we utilise the confusion matrix to show correct predictions (on the diagonal) and provide a clear view on which appliances are confused with each other. The F_1 (%) score is defined as $F_{macro} = 100 \cdot \frac{1}{M} \sum_{i=1}^{M} F_1^{(i)}$ where M is the number of appliances and F_1 is the harmonic mean of precision and recall. The zero-loss gives the number of miss-classifications with the best performance being 0 and is defined as $ZL = \sum_{i=1}^{M} I(y_i \neq \hat{y}_i)$.

The Matthews correlation coefficient, MCC, provides a balanced performance measure of the quality of classification

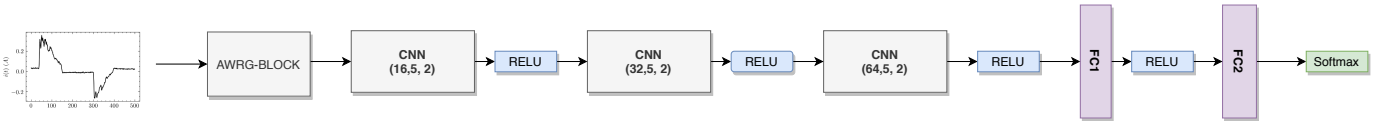

Fig. 3. The CNN network used in this work consists of AWRG block, CNN layers, and the FC's layers. The AWRG block receives CW of size T_s and produce image-representation of size $w \times w$ that is used as input to CNN.

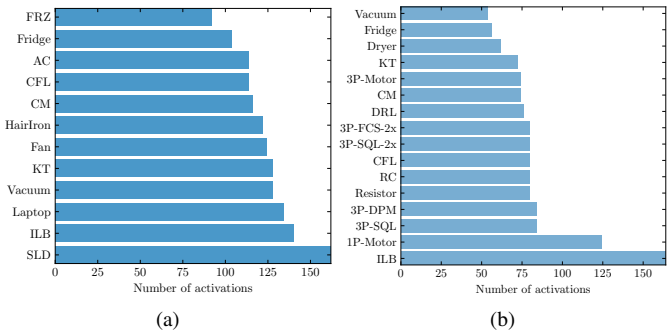

Fig. 4. Appliances distribution on the two datasets a) Aggregated single-phase PLAID dataset . The SLD has large number of activation because it had to two start-up events b) Aggregated three-phase LILACD data-set with three-phase loads.

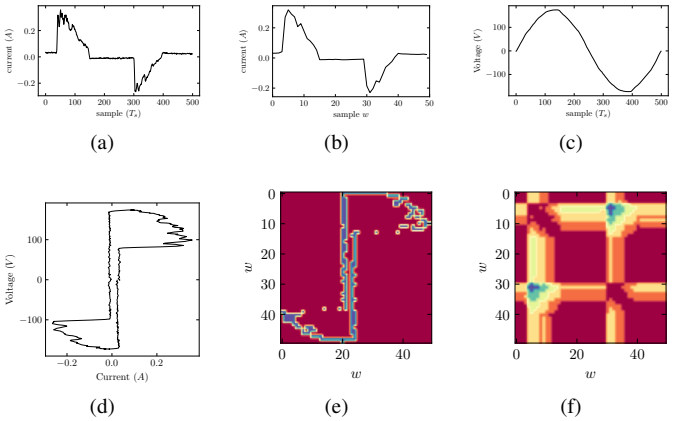

Fig. 5. Generation of the image-like representation from activation current and voltage. a) The CFL activation current b) activation current for CFL appliance after PAA d) activation voltage for CFL e) The V-I waveform d) The V-I image representation generated from activation current and voltage f) The WRG image representation made from CFL current after PAA.

algorithms [39]. Given a confusion matrix C, for M different classes, the MCC can be defined as

$$MCC = \frac{c \times s - \sum_i^M p_i \times t_i}{\sqrt{(s^2 - \sum_i^M p_i^2) \times (s^2 - \sum_i^M t_i^2)}} \tag{5}$$

where $t_i = \sum_k^M C_{ki}$, $p_i = \sum_k^M C_{ik}$, $c = \sum_k^M C_{kk}$, and $s = \sum_i^M \sum_j^M C_{ij}$. The maximum MCC score is +1 and the minimum value can be between -1 and 0. A score of +1 represents perfect prediction performance.

C. Experiment Description

To evaluate our method, we first investigate how embedding size w and different initialisation parameters influence classification performance. We investigate this by altering the emending size and the initialisation method and compare the obtained performance. To achieve more reliable results, the model is trained on 75% of the data for 100 iterations and tested on the remaining 25%. To ensure an equal distribution of classes in the train and validation sets, the first spilt from a stratified 4-fold cross-validation with random shuffle is used.

In the second experiment, we establish a baseline in which the V-I binary image is used as feature representation. The baseline is compared with feature representation produced by the proposed recurrence block. This experiment setup helps us answering an important question on whether AWRG feature learning with deep learning is sufficient for appliance classification in aggregated power measurements or not.

The binary V-I image with size $w \times w$, is generated from the activation current i and voltage v. The activation current and voltage are first resized into corresponding scale d_i and d_v respectively where $d_c = \max(|\min(i)|, \max(i))$ and $d_v = \max(|\min(v)|, \max(v))$ and transformed into $w \times w$ scale. The scaled current and voltage are then converted into $w \times w$ image by meshing the V-I trajectory and assigned a binary value that denotes whether it is traversed by trajectory as described in [23]. Fig. 5 illustrates the generation of V-I and WRG image-like representation from activation current and voltage. The V-I and the proposed AWRG feature representation are mainly based on single-dimensional time series data. Therefore, for a multidimensional three-phase activation current a separate AWRG and V-I for each current phase is used as a channel to form the respective image-like representations.

To obtain performance results, the learning architecture is trained using stratified 10-fold cross-validation with random shuffle. Using this approach, the model is trained on 90% of the data for 300 iterations, and tested in the remaining 10%. This process is repeated 10-times, where each fold is created by sampling without replacement [11].

IV. Results and Discussion

We first investigate how the parameters of the proposed AWRG influence classification performance by studying the impact of parameter initialisation on training convergence and model performance.

We experimented with different initialisation for δ and that of λ as shown in Table I. We see that when λ is initialised with a value greater than 10 the model achieves fast convergence and higher classification performance. We also observed that when the parameters are initialised with zero, the gradient fails to flow backwards and impedes the learning progress of a neural network. The initialisation of the

TABLE I
PARAMETER INITIALISATION RESULTS SUMMARY FOR AGGREGATED SINGLE-PHASE PLAID AND THREE-PHASE LILACD DATASET

Initilization	MCC score	
	LILAC	PLAID
$\lambda = 0, \quad \delta = 0$	0.00	0.00
$\lambda = 1, \quad \delta = 1$	0.97	0.752
$\lambda = 10, \quad \delta = 5$	0.948	0.961
$\lambda = 10, \quad \delta = 10$	0.951	0.971
$\lambda = 10, \quad \delta = 20$	0.945	0.964
$\lambda = 10, \quad \delta = 50$	0.916	0.968
$\lambda = 20, \quad \delta = 10$	0.948	0.974
$\lambda = 30, \quad \delta = 10$	0.945	0.971
$\lambda = 40, \quad \delta = 10$	0.952	0.965
$\lambda = 50, \quad \delta = 10$	0.941	0.962

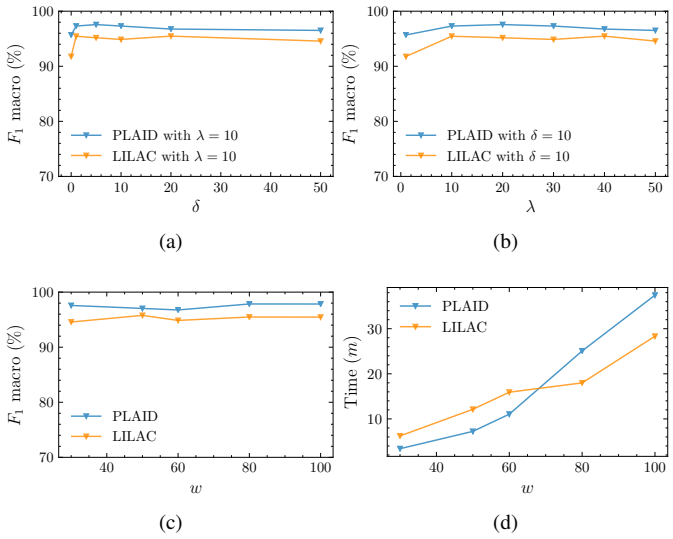

Fig. 6. Performance score for different parameters of the proposed AWRG on PLAID and LILACD datasets. a) The F_1-macro score for $\lambda = 10$ and varying δ. b) The F_1-macro score $\delta = 10$ and varying λ c) The F_1-macro score for different values of w d) The training time for different values of w

parameter with value $\delta = 10$ and that of $\lambda = 10$ seems to offer stable learning progress as depicted in Figs. 6a and 6c. We also investigate the influence of embedding size w on classification performance. From Fig. 6c, we see that the higher value of w does not significantly improve classification performance. However, large values of w impact the learning speed as shown in Fig. 6d.

The results of the comparison between the V-I image and the proposed AWRG feature representation for the PLAID and LILACD datasets are presented in Table II.

TABLE II
RESULTS SUMMARY FOR AGGREGATED SINGLE-PHASE PLAID AND THREE-PHASE LILACD DATASET WHEN V-I IMAGE AND AWRG IS USED.

Method	Dataset	F1	MCC	ZL
V-I	PLAID	91.67 ± 2.32	0.91 ± 0.03	8.32 ± 2.35
V-I	LILACD	85.36 ± 2.04	0.85 ± 0.02	13.82 ± 1.99
AWRG	PLAID	97.77 ± 1.05	0.98 ± 0.01	2.23 ± 1.05
AWRG	LILACD	98.33 ± 0.74	0.98 ± 0.01	1.66 ± 0.74

Considering PLAID, we observe that the recurrence block improves the performance from 0.91 to 0.98 MCC while reducing the miss-classification errors from 8.32% to 2.23% For LILACD, we also see that compared to the V-I representation, the use of AWRG improves the performance from 0.85% to 0.98% MCC while reducing the classification errors from 13.82% to 1.66%. To have further insights on the per-appliance performance and which appliances are miss-classified, we show per-appliance F_1 macro score and the confusion matrices for the baseline and the proposed AWRG in the two datasets. in Fig. 7 and Fig. 8.

From Fig. 7, we observe that when AWRG is used, the CNN achieve F_1 macro score above 95% for all appliances in the PLAID and LILACD dataset. We also see that compared to the V-I image, the proposed AWRG attains a higher score for each appliance. The V-I achieves low performance ($\leq$ 80%) for resistive appliances such as RC, KT and CM in the LILACD dataset and CM and ILB in the PLAID dataset.

Investigating the confusion matrix in Fig. 8 for PLAID, we observe a large number of confusions between KT and CM in the V-I representation. The reason for this confusion is attributed to the fact that both appliances consist of a resistive heating element and therefore nearly the same current waveform. We also observe that when the AWRG block is used, the classifier can effectively distinguish these two appliances as depicted in Fig. 8b. This implies that the proposed AWRG yields more robust feature representations for appliance recognition in residential environment. Yet, we still observe a few confusions between CFL and Laptop, Fridge and FRZ, the AC with Fan, and SLD with Laptop.

Fig. 9 shows the confusion matrices for the baseline and the proposed AWRG for LILACD. As observed from Fig. 9a, the V-I representation create many confusions for single-phase appliances like RC, KT, CM, Dryer and ILB. This is mainly due to the similarities in the current levels of these appliances that are also purely resistive. We also observe a few confusions for three-phase machine between 3P-SQL and 3P-Mort, 3P-SQL and 3P-DPM. However, as it can be observed in Fig. 9b, the AWRG feature representation significantly reduces these errors in single-phase and yield no mistakes between three-phases appliances. This implies that the proposed AWRG is distinctive enough to recognize multi-dimension three-phase appliances in industrial setting.

Overall, these results suggest that the feature representation learned from V-I images is not sufficient for appliance classification in both residential and industrial settings. In contrast, applying the proposed AWRG feature representation significantly improves the classification performance, which is a reflection of its superior capability of extracting unique appliance features.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose the AWRG method with learnable parameters for event-based NILM. We present the outcome of empirical investigations on how different parameters of the proposed AWRG influence model training and classification performance. We observe that the initialisation values for the

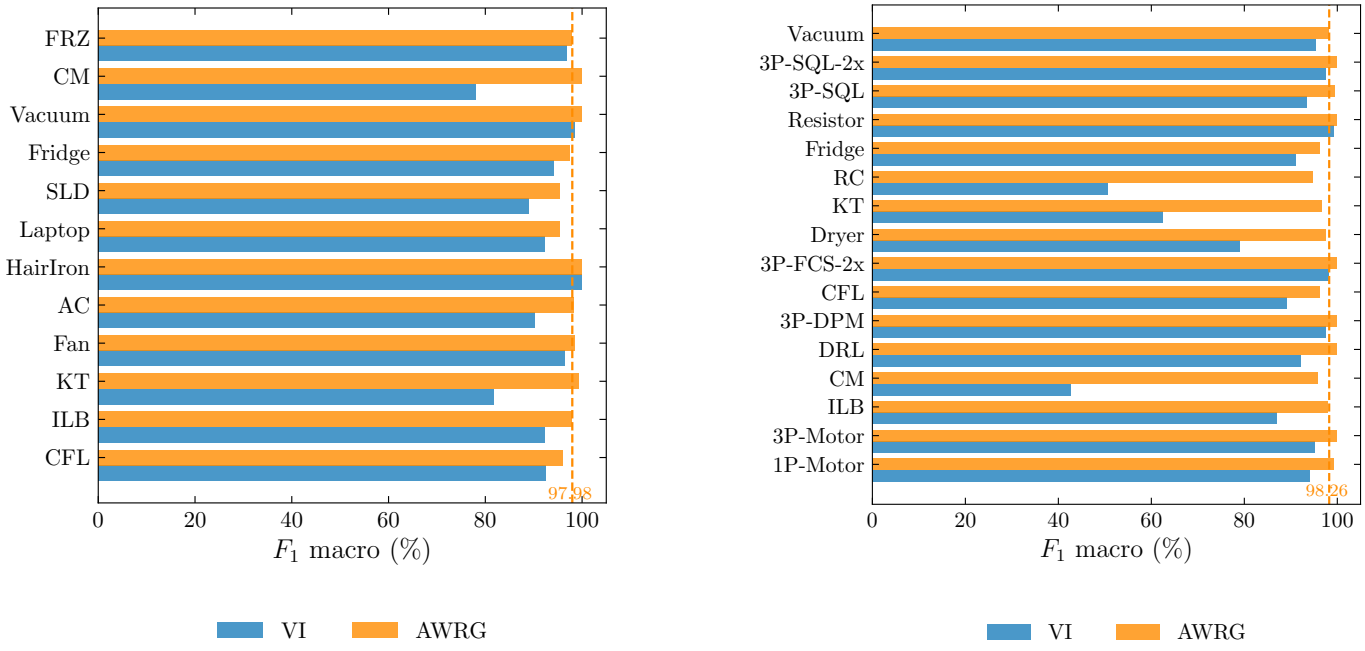

Fig. 7. The F_1 macro (%) score per appliance between V-I image and AWRG for a) PLAID dataset and b) LILACD dataset

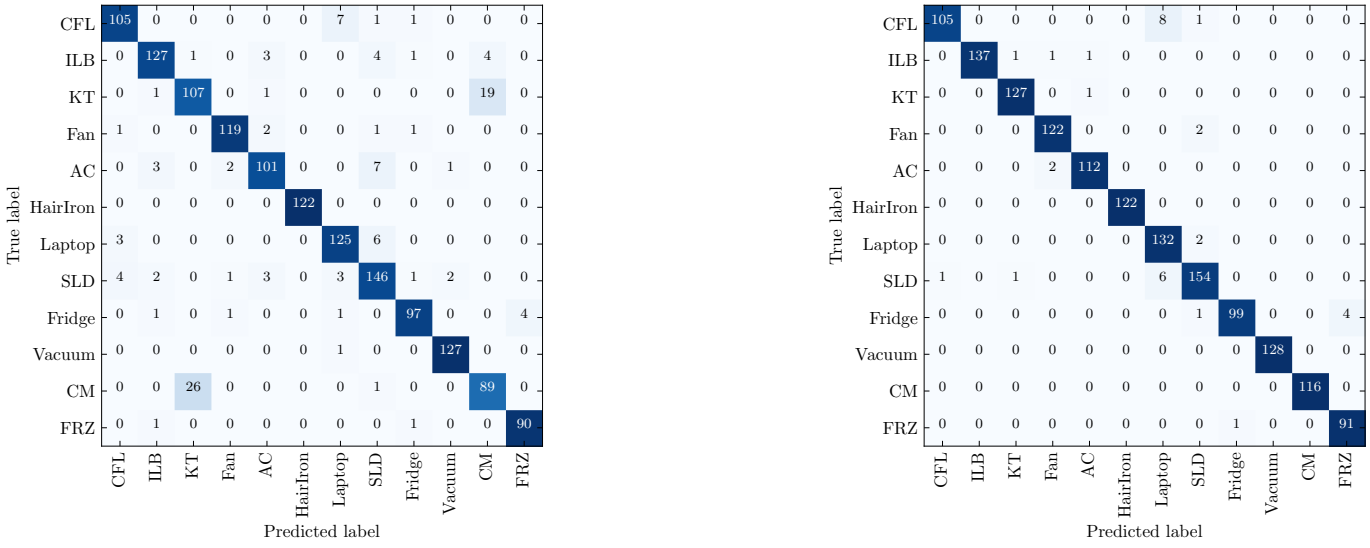

Fig. 8. Confusion matrix for the CNN network with V-I and AWRG on PLAID dataset. a) V-I baseline b) AWRG.

learn-able parameters have a decisive influence on the optimal performance and convergence.

We further evaluate the proposed approach in two public real-world energy datasets, one for residential and another for industrial settings. The outcomes indicate that the feature representation learned from AWRG is consistently superior to those from V-I images, and most importantly are able to discern between appliances of the same type, e.g., purely resistive loads.

At this point it is important to stress that the presented evaluation assumes that the appliance state transitions (i.e., activations and deactivations) are known in advance. In practice, this information must be provided by an event detection algorithm (e.g., [40]–[42]). Therefore, future work should investigate how to integrate the proposed approach in an event-based pipeline to understand how variations in the position of the detected events will affect the extraction of the activation/deactivation signals, as well as the feature representation process through the AWRG.

Furthermore, despite its superior performance when compared to the V-Is alternative, it still presents some miss-classifications that must be further studied. While more la-

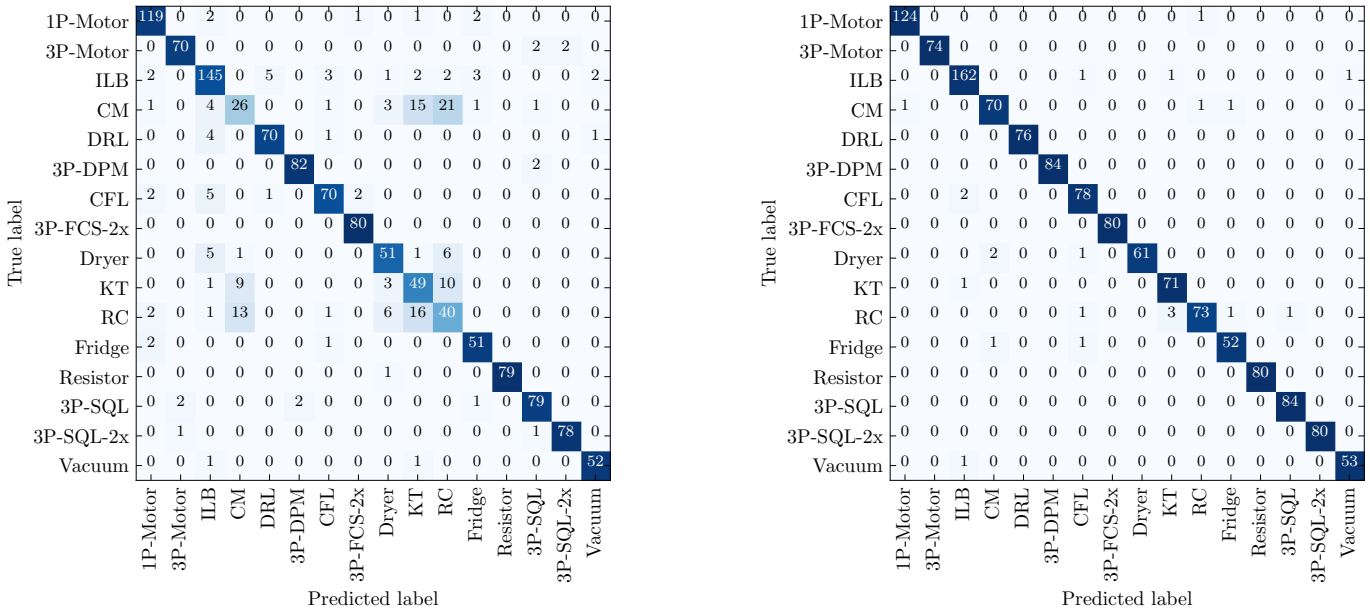

Fig. 9. Confusion matrix for the CNN network with V-I and AWRG on LILACD dataset. a) V-I baseline b) AWRG.

belled data can potentially improve the performance, labelled high-frequency datasets are scarce. Thus, a potential future work directions is to leverage the potential of unlabelled data by using pseudo-labels and semi-supervised learning [43].

ACKNOWLEDGMENT

Lucas Pereira has received funding from the Portuguese Foundation for Science and Technology (FCT) under grants CEECIND/01179/2017 and UIDB/50009/2020.

REFERENCES

[1] V. Kitio, "Promoting Energy efficiency in Buildings in East Africa," in *UNEP-SBCI Fall Symposium on Sustainable Buildings*, Paris, France, 2013.

[2] D. Geelen, R. Mugge, S. Silvester, and A. Bulters, "The use of apps to promote energy saving: a study of smart meter–related feedback in the netherlands," *Energy Efficiency*, vol. 12, no. 6, pp. 1635–1660, Aug 2019.

[3] N. Batra, A. Singh, and K. Whitehouse, "If You Measure It, Can You Improve It? Exploring The Value of Energy Disaggregation," in *In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments - BuildSys '15*, 2015, pp. 191–200.

[4] A. Huss, "Hybrid Model Approach to Appliance Load Disaggregation," Ph.D. dissertation, KTH Royal Institute of Technology, 2015.

[5] C. Klemenjak, S. Jost, and W. Elmenreich, "Yomopie: A user-oriented energy monitor to enhance energy efficiency in households," in *2018 IEEE Conference on Technologies for Sustainability (SusTech)*. IEEE, 2018, pp. 1–7.

[6] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. Rizzoli, "A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring," *Applied Energy*, vol. 185, pp. 331 – 344, 2017.

[7] A. Reyes Lua, "Location-aware Energy Disaggregation in Smart Homes," Master's Thesis, Delft University of Technology, 2015.

[8] 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. [Online]. Available: http://arxiv.org/abs/1404.0284

[9] N. Batra, A. Singh, and K. Whitehouse, "Gemello : Creating a Detailed Energy Breakdown from just the Monthly Electricity Bill," in *In Proceeding of the 22 ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2016.

[10] L. Pereira and N. Nunes, "Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—A review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 6, p. e1265, Nov. 2018.

[11] L. De Baets, T. Dhaene, D. Deschrijver, C. Develder, and M. Berges, "Vi-based appliance classification using aggregated power consumption data," in *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*, June 2018, pp. 179–186.

[12] D. Baptista, S. Mostafa, L. Pereira, L. Sousa, and D. F. Morgado, "Implementation strategy of convolution neural networks on field programmable gate arrays for appliance classification using the voltage and current (v-i) trajectory," *Energies*, vol. 11, p. 2460, 09/2016 2018.

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

[14] E. Gomes and L. Pereira, "Pb-nilm: Pinball guided deep non-intrusive load monitoring," *IEEE Access*, vol. 8, pp. 48 386–48 398, 2020.

[15] Z. Zheng, H. Chen, and X. Luo, "A supervised event-based non-intrusive load monitoring for non-linear appliances," *Sustainability*, vol. 10, no. 4, 2018.

[16] Y. Liu, X. Wang, and W. You, "Non-intrusive load monitoring by voltage–current trajectory enabled transfer learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5609–5619, Sep. 2019.

[17] H. Y. Lam, G. S. K. Fung, and W. K. Lee, "A novel method to construct taxonomy electrical appliances based on load signaturesof," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 653–660, May 2007.

[18] T. Hassan, F. Javed, and N. Arshad, "An Empirical Investigation of V-I Trajectory Based Load Signatures for Non-Intrusive Load Monitoring," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 870–878, Mar. 2014. [Online]. Available: https://ieeexplore.ieee.org/document/6575197

[19] L. Du, D. He, R. G. Harley, and T. G. Habetler, "Electric Load Classification by Binary Voltage–Current Trajectory Mapping," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 358–365, Jan. 2016, 00021.

[20] B. M. Mulinari, D. P. de Campos, C. H. da Costa, H. C. Ancelmo, A. E. Lazzaretti, E. Oroski, C. R. E. Lima, D. P. B. Renaux, F. Pottker, and R. R. Linhares, "A New Set of Steady-State and Transient Features for

Power Signature Analysis Based on V-I Trajectory," in *2019 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*, Sep. 2019, pp. 1–6, iSSN: 2643-8798.

[21] A. L. Wang, B. X. Chen, C. G. Wang, and D. Hua, "Non-intrusive load monitoring algorithm based on features of v–i trajectory," *Electric Power Systems Research*, vol. 157, pp. 134 – 144, 2018.

[22] J. Gao, E. C. Kara, S. Giri, and M. Bergés, "A feasibility study of automated plug-load identification from high-frequency measurements," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2015, pp. 220–224.

[23] L. De Baets, J. Ruyssinck, C. Develder, T. Dhaene, and D. Deschrijver, "Appliance classification using vi trajectories and convolutional neural networks," *ENERGY AND BUILDINGS*, vol. 158, pp. 32–36, 2018. [Online]. Available: http://dx.doi.org/10.1016/j.enbuild.2017.09.087

[24] M. Kahl, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen, "A comprehensive feature study for appliance recognition on high frequency energy data," in *Proceedings of the Eighth International Conference on Future Energy Systems*. ACM, May 2017. [Online]. Available: https://doi.org/10.1145/3077839.3077845

[25] N. Sadeghianpourhamami, J. Ruyssinck, D. Deschrijver, T. Dhaene, and C. Develder, "Comprehensive feature selection for appliance classification in nilm," *Energy and Buildings*, vol. 151, pp. 98 – 106, 2017.

[26] C. Klemenjak, S. Makonin, and W. 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)*. IEEE, 2020, pp. 1–5.

[27] E. Garcia-Ceja, M. Z. Uddin, and J. Torresen, "Classification of recurrence plots' distance matrices with a convolutional neural network for activity recognition," *Procedia Computer Science*, vol. 130, pp. 157 – 163, 2018, the 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018) / The 8th International Conference on Sustainable Energy Information Technology (SEIT-2018) / Affiliated Workshops.

[28] N. Hatami, Y. Gavet, and J. Debayle, "Classification of time-series images using deep convolutional neural networks," *CoRR*, vol. abs/1710.00886, 2017. [Online]. Available: http://arxiv.org/abs/1710.00886

[29] Y. Tsai, J. H. Chen, and C. Wang, "Encoding candlesticks as images for patterns classification using convolutional neural networks," *CoRR*, vol. abs/1901.05237, 2019. [Online]. Available: http://arxiv.org/abs/1901.05237

[30] F. Popescu, F. Enache, I. Vizitiu, and P. Ciotîrnae, "Recurrence plot analysis for characterization of appliance load signature," in *2014 10th International Conference on Communications (COMM)*, May 2014, pp. 1–4.

[31] A. Faustine and L. Pereira, "Improved appliance classification in non-intrusive load monitoring using weighted recurrence graph and convolutional neural networks," *Energies*, vol. 13, no. 13, p. 3374, Jul. 2020. [Online]. Available: https://doi.org/10.3390/en13133374

[32] R. Medico, L. De Baets, J. Gao, S. Giri, E. Kara, T. Dhaene, C. Develder, M. Bergés, and D. Deschrijver, "A voltage and current measurement dataset for plug load appliance identification in households," *Scientific Data*, vol. 7, no. 1, p. 49, Feb. 2020. [Online]. Available: https://doi.org/10.1038/s41597-020-0389-7

[33] M. Kahl, V. Krause, R. Hackenberg, A. U. Haq, A. Horn, H.-A. Jacobsen, T. Kriechbaumer, M. Petzenhauser, M. Shamonin, and A. Udalzow, "Measurement system and dataset for in-depth analysis of appliance energy consumption in industrial environment," *tm - Technisches Messen*, vol. 86, no. 1, pp. 1–13, Jan. 2019.

[34] L. De Baets, "Machine learning for non-intrusive load monitoring," Ph.D. dissertation, Ghent University, 2018.

[35] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, "Euclidean distance matrices: Essential theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 12–30, 2015.

[36] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5, pp. 237 – 329, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0370157306004066

[37] E. J. Keogh and M. J. Pazzani, "Scaling up dynamic time warping for datamining applications," in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '00. New York, NY, USA: ACM, 2000, pp. 285–289.

[38] S. Henriet, U. Şimşekli, B. Fuentes, and G. Richard, "A generative model for non-intrusive load monitoring in commercial buildings," *Energy and Buildings*, vol. 177, pp. 268 – 278, 2018.

[39] L. Pereira and N. Nunes, "A comparison of performance metrics for event classification in Non-Intrusive Load Monitoring," in *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Oct. 2017, pp. 159–164.

[40] L. Pereira, "Developing and evaluating a probabilistic event detector for non-intrusive load monitoring," in *2017 Sustainable Internet and ICT for Sustainability (SustainIT)*. Funchal, Portugal: IEEE, Dec. 2017, pp. 1–10.

[41] L. De Baets, J. Ruyssinck, C. Develder, T. Dhaene, and D. Deschrijver, "On the Bayesian optimization and robustness of event detection methods in NILM," *Energy and Buildings*, vol. 145, pp. 57–66, Jun. 2017.

[42] S. Houidi, F. Auger, H. B. A. Sethom, D. Fourer, and L. Miègeville, "Multivariate event detection methods for non-intrusive load monitoring in smart homes and residential buildings," *Energy and Buildings*, vol. 208, p. 109624, Feb. 2020.

[43] D.-H. Lee, "Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks," in *ICML 2013 Workshop : Challenges in Representation Learning (WREPL)*, Atlanta, GA, USA, 2013.

Anthony Faustine received the B.sc. Degree in Electronics Science and Communication from the University of Dar es Salaam, Tanzania, and the M.sc. Degree in Telecommunications Engineering from the University of Dodoma in 2010. He is currently pursuing a Ph.D. Degree in Computer Science Engineering at Ghent University. From 2017 to 2019, he was a machine learning researcher at IDLab, imec of the University of Ghent. Since 2020, he is with the Center for innovation and applied research in AI, Machine Learning and Data Analytics (CeADAR), University of College of Dublin. His research interests lie in the intersections between Signal processing, Machine-Learning, and Computational Sustainability. He effectively research methods for novel approaches to problems and develop prototypes to assess the approach's viability with a significant focus on smart-grid and earth observation. He is currently developing robust methods for load disaggregation.

Lucas Pereira received his Ph.D. in Computer Science from the University of Madeira, Portugal, in 2016. Since then, he is at ITI/LARSyS, where he leads the Further Energy and Environment research Laboratory (FEELab). Since 2019 he is a research fellow at Técnico Lisboa. Lucas's research interests lie in the intersections between Data Science, Machine-Learning, and Human-Computer Interaction. He works towards bridging the gap between laboratory and real-world applicability of ICT for sustainable development, with a significant focus on smart-grids and built environments.

Christoph Klemenjak received his Bachelors and Masters degree in information and communications engineering from the University of Klagenfurt, Austria. He currently pursues the PhD degree in information technology at same university. Since 2017, he is with the Institute of Networked and Embedded Systems, University of Klagenfurt, working as a research and teaching assistant in the group of Professor Wilfried Elmenreich. His research activities focus on the analysis of load disaggregation techniques, the design of smart metering equipment, and the conception of next-generation energy management systems.