Gramian Angular Field and Recurrence Plots as Feature Engineering Techniques on Residential Appliances Labeling: A Comparative Analysis

J. S. Corrêa[1,*], D. L. Cavalca[2,†] and R. A. S. Fernandes[1,‡]
[1]Department of Electrical Engineering, Federal University of São Carlos, São Carlos, Brazil
[2]Graduate Program in Computer Science, Federal University of São Carlos, São Carlos, Brazil
Email: *juan.salin@estudante.ufscar.br, †diegocavalca@ufscar.br, ‡ricardo.asf@ufscar.br

2023 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT-LA) | 979-8-3503-3696-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISGT-LA56058.2023.10328293

***Abstract*—The demand for electricity has increased rapidly and, for this reason, there is a need to efficiently use it. In this way, the identification of residential appliances enables such use for consumers and is crucial for demand response programs. Due to the variety of appliances in homes and their dynamic behavior, the search for patterns that explain and allow the correct labeling of temporal windows becomes a challenging task, since a window may contain more than one appliance. In this sense, the present paper proposes the transformation of time-series into images, using Gramian angular field and recurrence plots. The dataset composed of images was submitted to the labeling process, considering the use of convolutional neural networks. A comparative analysis was performed using the UK-DALE dataset. The results demonstrated the effectiveness of the proposed feature engineering stage, since the labeling task reached F1-scores until 94%.**

***Index Terms*—Convolutional neural networks, Gramian angular field, nonintrusive load monitoring, recurrence plots.**

I. Introduction

One of the biggest world challenges is related to electricity, whose global demand increased about 16 times in the 20th century and continues to rise [1]. Furthermore, commercial and residential buildings are responsible for 40% of electricity demand [2] and at least 1/3 of global carbon dioxide emissions come from these buildings [3]. Thus, solutions to use electricity efficiently become necessary. In this way, the appliances/loads labeling contributes to the nonintrusive load monitoring (NILM) and, consequently, as mentioned by [4] and [5], enables the load scheduling in demand response programs.

In this sense, it is important to note that NILM research was initially proposed in the 1990s, gaining notoriety by Hart's paper [6], in which the author proposed the identification of appliances from sudden changes in energy consumption. The first proposed contributions made use of only real, reactive and/or apparent power as relevant features to be analyzed [6]–[8]. However, due to advances in the artificial intelligence research field, there was a rising use of machine learning techniques [9]–[12]. In addition, the NILM researches were divided by [13] into: (i) low level – methods that consider the current signature and, consequently, use meters with high sampling rates; and (ii) high level – methods that use current and/or apparent power in RMS (root mean square), i.e., employ meters with low sampling rates. In this sense, research using high-level data is more suited for real-life applications due to the development of smart residential meters [14], [15].

This work was supported by the São Paulo Research Foundation (FAPESP) [Grant Numbers 2022/00750-9 and 2023/00182-3] and The National Council for Scientific and Technological Development (CNPq) [Grant Number 406453/2021-7].

Unlike the approaches proposed to deal with the low level data, when considering RMS-based measurements, it is non-trivial to define a feature engineering stage. Therefore, such measurements are directly used as inputs to machine learning algorithms, i.e., without the feature engineering stage [16]–[20]. Additionally, deep learning techniques have gained attention in the NILM research area, such as the autoencoders [21] and the convolutional neural networks (CNN) [3], [22], [23].

In [21], the authors proposed the use of variational autoencoders (VAE) with instance-batch normalization networks (IBN-Net). It is important to emphasize that the authors used the UK-DALE dataset, where one house was used as the validation set and two others as the training set (only the houses #1, #2, and #5 were considered).

The authors of [22] proposed the usage of temporal bar graphs to allow pattern visualization of the appliances on the defined windows. In order to demonstrate the performance of the proposed approach, the following deep learning algorithms were used: Very Deep One Dimensional CNN (VDOCNN), Extreme Inception (Xception), and Concatenate-DenseNet121.

A multiscale neural network model based on dilated convolution was proposed by [23]. The convolutional dilated residual blocks avoid the problems coming from the increase of network layers, more specifically, gradient explosion and vanishing. They observed that such convolution can obtain more information from the background to assist on disaggregation and the multiscale allow to achieve good performance, mainly for appliances with complex characteristics. On the other hand, in [3], the causal dilated convolution was replaced for bidirectional dilated convolution (named as BitcnNILM), which was inspired on a temporal convolutional network that takes advantage of CNN's parallel computation and recurrent neural networks' time dependency.

Following the aforementioned context, it is noted that the usage of feature engineering stage in high level is a challenge and, therefore, represents the gap treated in this

paper. Moreover, the increase use of deep learning algorithms for the labeling of residential appliances is notable. Since these deep learning algorithms have been more adequate to analyze images as inputs and capable of achieving better results than the classical machine learning models, in this paper, we propose the transformation of RMS time series into images as a feature engineering stage. This way, the proposed approach aims to verify the potentials of recurrence plots (RPs) and Gramian angular fields (GAFs) on this non-trivial task when applied to high-level data. In order to compare their behaviors, the UK-DALE benchmark dataset was used and CNN-based models were trained and validated to identify residential appliances.

The remainder of this paper was organized as follows. Section II presents an overview of the methods used to transform time series into images. The proposed approach is described in Section III. Next, Section IV presents the results and discussions. Finally, Section V draws the conclusions.

II. Bidimensional Images

As RPs and GAFs have capability to characterize the periodic nature of time series and its recurrences, they could be considered to the feature engineering stage and are explained in more detail in the sequence.

A. Recurrence Plots

Eckmann, Kamphorst and Ruell [24] created a technique named as Recurrence Plot, which has the capability to measure the recurrence of a trajectory $\vec{x}_i \in \mathbb{R}^d$ in the state-space and can be expressed as:

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - \|\vec{x}_i - \vec{x}_j\|), \tag{1}$$

where $\vec{x}_i$ represents a measurement with $i, j = 1, \ldots, N$; ε is the threshold; $\Theta(\cdot)$ is the Heaviside function ($\Theta(x) = 0$, if $x < 0$, else, $\Theta(x) = 1$); and $\|\cdot\|$ is the norm. From this equation, a time series can be transformed into an image, such as illustrated in Fig. 1.

B. Gramian Angular Fields

In the GAF transformation process, the data are initially normalized. Then, given a time-series $X = \{x_1, x_2, x_3, ..., x_n\}$, which n is the number of samples, the normalization for the interval $[0, 1]$ can be performed as:

$$\tilde{x}_i^{[0,1]} = \frac{x_i - min(X)}{max(X) - min(X)}. \tag{2}$$

Next, each normalized time series element must be mapped to the polar coordinate system. For that, the following calculations are executed:

$$\phi_i = arcos(\tilde{x}_i), \tag{3}$$

$$r_i = \frac{t_i}{M}, \tag{4}$$

where ϕ_i and r_i represent the polar coordinate; t_i is the instant of time visualization and M is a constant factor that regularizes the extent of the polar coordinate system. Therefore, the temporal correlation between different time intervals can be explored by considering the trigonometric sum (named as Gamian Angular Summation Field – GASF) and difference (named as Gamian Angular Difference Field – GADF) between each polar coordinate system sample. In this sense, GASF and GADF can be defined, respectively, as:

$$GASF = \begin{pmatrix} cos(\phi_1 + \phi_1) & \cdots & cos(\phi_1 + \phi_n) \\ \vdots & \ddots & \vdots \\ cos(\phi_n + \phi_1) & \cdots & cos(\phi_n + \phi_n) \end{pmatrix}, \tag{5}$$

$$GADF = \begin{pmatrix} sin(\phi_1 - \phi_1) & \cdots & sin(\phi_1 - \phi_n) \\ \vdots & \ddots & \vdots \\ sin(\phi_n - \phi_1) & \cdots & sin(\phi_n - \phi_n) \end{pmatrix}. \tag{6}$$

Figs. 2 and 3 illustrate the GASFs and GADFs obtained for the UK-DALE dataset, respectively.

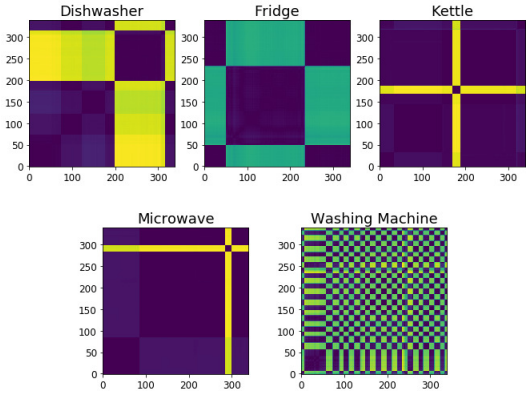

Fig. 1. RPs extracted from the UK-DALE dataset for each considered appliance.

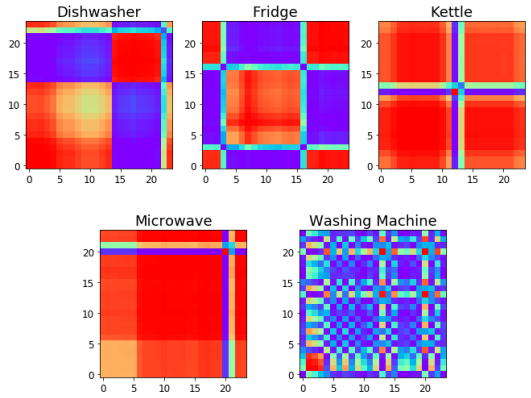

Fig. 2. GASFs extracted from the UK-DALE dataset for each considered appliance.

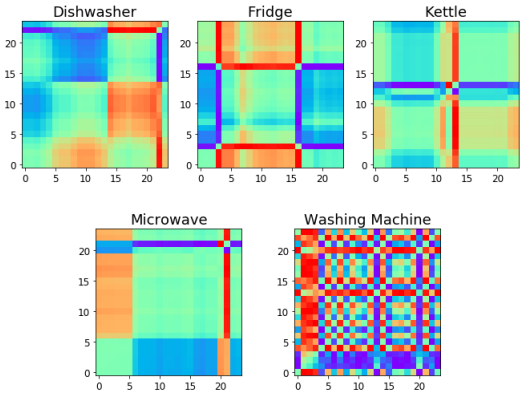

Fig. 3. GADFs extracted from the UK-DALE dataset for each considered appliance.

III. Proposed Approach

An overview of the proposed approach is shown in Fig. 4, where the main stages were defined as: data organization and cleaning; feature engineering; CNN-based predictive models; and performance evaluation metrics adopted to analyze the results and, consequently, compare RP and GAF (i.e., GASF and GADF).

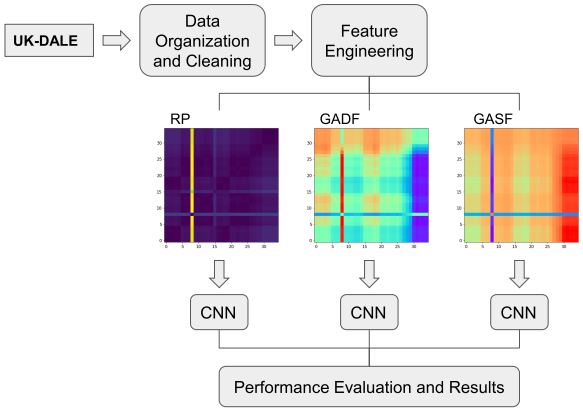

Fig. 4. Overview of the proposed approach.

A. Data organization and cleaning

During this stage, a survey of the most considered appliances and houses of UK-DALE dataset in the state-of-the-art papers was carried out. All the authors used dishwasher, kettle, microwave, fridge and washing machine. It is noted that those appliances have different operating modes and distinct energy consumption, allowing to verify the performance of predictive models in relation to two-state appliances (fridge, kettle, and microwave) and multi-state appliances (washing machine and dishwasher).

Based on these appliances, five experimental tests were performed in order to find out which one adapted best to the labeling task. It is worth mentioning that the missing values were replaced by zero in this stage.

1) Test #1: This test consists of using data from the five houses of UK-DALE. However, it is worth noting that not all appliances were available in all houses. Additionally, in house #4, it was not possible to label training data for microwave and washing machine. In this sense, the relationship between appliances and houses are presented next:

- dishwasher (DW) – houses #1, #2 and #5;
- fridge (FR) – houses #1, #2, #4 and #5;
- kettle (KE) – houses #1, #2, #3, #4 and #5;
- microwave (MW) – houses #1, #2 and #5;
- washing machine (WM) – houses #1, #2 and #5.

The start and end dates of the data that compose each residence are described in Table I.

TABLE I
START AND END DATES OF DATA USED IN THE TEST #1.

House	Start date and time	End date and time
#1	Jul 1st, 2013, 00:00:00	Jul 1st, 2014, 00:00:00
#2	Feb 17th, 2013, 15:39:19	Oct 10th, 2013, 06:15:58
#3	Feb 27th, 2013, 20:35:14	Apr 8th, 2013, 06:15:05
#4	Mar 9th, 2013, 14:40:07	Oct 1st, 2013, 06:15:14
#5	Jun 29th, 2014, 17:23:48	Nov 13th, 2014, 18:00:03

The other tests had their data chosen according to approaches that are part of the state-of-the-art, as shown in the sequence.

2) Tests #2, #3 and #4: As used by [22], these tests were based only on houses #1, #2 and #5. The start and end dates of each house considered can be seen in Table II.

TABLE II
START AND END DATES OF DATA USED IN THE TESTS #2,#3 AND #4.

House	Start date and time	End date and time
#1	Jan 1st, 2014, 00:00:00	Jan 12th, 2014, 00:00:00
#2	Mai 20th, 2013, 00:00:00	Jun 1st, 2013, 00:00:00
#5	Jun 29th, 2014, 00:00:00	Jul 11th, 2014, 00:00:00

3) Test #5: This test uses only the house #2, such as considered by [3]. The start date and end dates is presented in Table III.

TABLE III
START AND END DATES OF DATA USED IN THE TEST #5.

House	Start date and time	End date and time
#2	Feb 17th, 2013, 15:39:19	Oct 10th, 2013, 06:15:58

B. Feature engineering

After the data organization and cleaning, for test #1 the windowing was performed through sections of 2,040 seconds with steps of 2,040 seconds (equivalent to 340 samples). This strategy was assumed in accordance with [12], where the authors analyzed the impact of different fixed-time windowing on CNN's performance. On the other hand, the windowing for tests #2, #3 and #4 was carried out through sections of 60 seconds with steps of 60 seconds (equivalent to 10 samples), such as defined by [22]. Furthermore, for test #5, the

windowing was performed through sections of 599 seconds with steps of 599 seconds (equivalent to 100 samples), as presented in [3]. Thus, the objective was to analyze the behavior of feature engineering techniques when confronted with both a subset with more data (test #1) and also subsets already defined by the papers that correspond to the state of the art (tests #2 to #5).

In order to carry out the windowing, it was considered the use of thresholds, as shown in Table IV.

TABLE IV
THRESHOLDS DEFINED FOR ALL CONSIDERED APPLIANCES.

Appliances	Threshold		
	Test #1	Tests #2, #3 and #4	Test #5
Dishwasher	10	10	10
Fridge	50	50	50
Kettle	10	20	200
Microwave	10	200	200
Washing machine	20	10	20

In test #1, with the exception of the microwave and kettle, the thresholds were set as default by NILMTK [25]. For tests #2, #3 and #4, the thresholds were set according to [22] and, consequently, for test #5, they were set based on [3]. It is important to highlight that each appliance was labeled as "on" in the window if the average power of the window was greater than its threshold. From the extracted windows, the RP, GASF and GADF images were obtained by using a script developed in the Python programming language.

C. Predictive models

According to the literature, the images extracted for each test were divided into 80% and 20% for training and validation, respectively.

Since CNNs have shown success in the image's pattern recognition due to the multiple layers equipped with different processing units responsible for learning a complex hierarchy of features by capturing the data singularities, those were considered as the predictive models in this paper. This way, the images referring to the windows were used as inputs to the CNNs, while the output is the label of each window, i.e., the identification of appliances.

D. Performance evaluation metrics

The performance evaluation metrics used in this paper were defined from the analysis of the state-of-art related to the labeling of UK-DALE appliances, which are: accuracy and F1-score. Accuracy (acc) can be expressed as:

$$acc = \frac{(TP + TN)}{(TP + TN + FP + FN)}, \tag{7}$$

where TP, FP, TN and FN are, respectively, the true positives, false positives, true negatives and false negatives. On the other hand, the F1-score is given by:

$$F_1 = \frac{2P \times R}{P + R}, \tag{8}$$

where $P = \frac{TP}{TP+FP}$ and $R = \frac{TP}{TP+FN}$ represent precision and recall, respectively.

IV. RESULTS AND DISCUSSIONS

As previously mentioned, five experimental tests were carried out and the best model among the CNN hyperparameters was used for each appliance in its respective image (RP, GADF, or GASF). In Table V, it is possible to visualize the results obtained for each test on its respective validation dataset, considering each kind of extracted image.

TABLE V
RESULTS OBTAINED ON EXPERIMENTAL TESTS.

Test	Metric	DW	FR	KE	MW	WM	Avg.
#1 (RP)	acc	0.69	0.71	0.86	0.75	0.83	0.77
	F1-score	0.68	0.69	0.86	**0.77**	**0.83**	0.77
#1 (GADF)	acc	0.71	0.74	0.83	0.74	0.78	0.76
	F1-score	**0.68**	**0.73**	0.82	0.76	0.79	0.76
#1 (GASF)	acc	0.62	0.71	0.88	0.75	0.76	0.74
	F1-score	0.60	0.69	**0.88**	0.75	0.75	0.74
#2 (RP)	acc	0.59	0.54	0.76	0.87	0.81	0.71
	F1-score	**0.72**	0.58	0.83	0.88	0.82	0.77
#2 (GADF)	acc	0.56	0.55	0.78	0.87	0.81	0.71
	F1-score	0.66	0.60	**0.89**	**0.89**	0.80	0.77
#2 (GASF)	acc	0.56	0.54	0.75	0.85	0.82	0.70
	F1-score	0.69	**0.60**	0.86	0.88	**0.83**	0.77
#3 (RP)	acc	0.59	0.55	0.85	0.85	0.67	0.70
	F1-score	**0.63**	0.46	0.85	0.87	0.64	0.69
#3 (GADF)	acc	0.62	0.54	0.79	0.80	0.71	0.69
	F1-score	0.59	**0.57**	0.79	0.80	**0.70**	0.69
#3 (GASF)	acc	0.59	0.55	0.86	0.85	0.67	0.70
	F1-score	0.60	0.56	**0.86**	**0.88**	0.64	0.71
#4 (RP)	acc	0.59	0.54	0.76	0.87	0.81	0.71
	F1-score	**0.72**	0.58	0.83	0.88	0.82	0.77
#4 (GADF)	acc	0.56	0.55	0.78	0.87	0.81	0.71
	F1-score	0.66	0.60	**0.89**	**0.89**	0.80	0.77
#4 (GASF)	acc	0.56	0.54	0.75	0.85	0.82	0.70
	F1-score	0.69	**0.60**	0.86	0.88	**0.83**	0.77
#5 (RP)	acc	0.69	0.72	0.92	0.86	0.77	0.79
	F1-score	**0.70**	0.70	0.92	0.83	0.79	0.79
#5 (GADF)	acc	0.73	0.74	0.94	0.90	0.81	0.82
	F1-score	0.70	**0.75**	0.93	**0.88**	**0.84**	0.82
#5 (GASF)	acc	0.67	0.71	0.93	0.88	0.83	0.80
	F1-score	0.71	0.70	**0.94**	0.83	0.82	0.80

Analyzing the results presented in the Table V, it is noted that, in test #5, the average accuracies and F1-scores are slightly better than those obtained for the other tests. However, it was not possible to determine which of the techniques for transforming the time series into images has the best performance in terms of feature extraction, since each has its importance depending on the data set (tests # 1 to #5) and the type of appliance to be labeled.

Taking into account the F1 scores highlighted for all tests and appliances presented in Table V, it was observed that the best labeling architecture would be composed of: RP for dishwasher; GADF for fridge, microwave, and washing machine; and GASF for kettle.

In order to verify the real effectiveness of this combined architecture in relation to the exclusive use of each of the techniques for transforming time series into images, in Fig. 5 a comparison of average F1-scores is shown considering all

datasets used. Thus, a maximum difference of 1.2% between the combined architecture (RP, GASF and GADF) and the exclusive use of RPs was observed, demonstrating the robustness of using any of the feature extraction techniques.

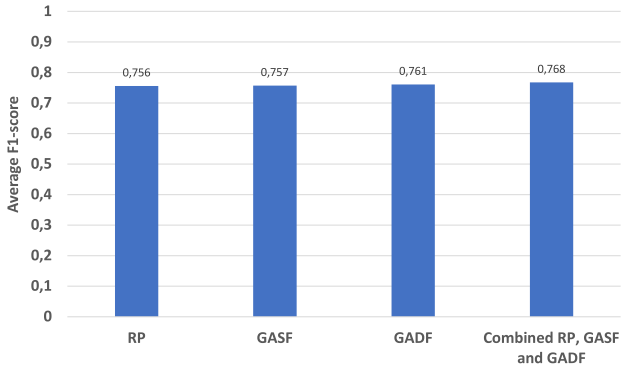

Fig. 5. Comparative analysis between the combined architecture and the exclusive use of RPs, GASFs and GADFs.

In general, the dishwasher and the fridge proved to be more difficult appliances to label, since their accuracies and F1-scores were between 60% and 68%. On the other hand, kettle and microwave were the appliances with the best hit rates, as their accuracies and F1-scores were in the range of 82% to 88%. In addition, the CNN's responsible by labeling the washing machine reached F1-scores and accuracies in the range of 77% and 79%, which can be considered as adequate in terms of the NILM research area.

V. Conclusions

In this paper, the proposed approach used RPs, GADFs, and GASFs as feature engineering techniques to transform time series into images. In order to perform a comparative analysis, five experimental tests were conducted in which the datasets were extracted considering different start/end dates and window threshold per appliance. The labeling task was done by CNNs trained and validated for each dataset.

The results were in line with the state-of-the-art, when considering the other approaches that employ the UK-DALE. In this way, the proposed approach can be considered as a plausible alternative of use, since CNNs have been widely applied to NILM and, therefore, the proposed feature engineering stage is remarkably suited to the principles of use of this kind of deep learning algorithm. In addition, mainly the use of RPs, were important to highlight the occurrence of recurrence in the time series.

References

[1] C. Nalmpantis and D. Vrakas, "Machine learning approaches for non-intrusive load monitoring: From qualitative to quantitative comparation," *Artificial Intelligence Review*, vol. 52, no. 1, p. 217–243, jun 2019.

[2] V. Singhal, J. Maggu, and 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 Smart Grid*, vol. 10, no. 3, pp. 2969–2978, 2019.

[3] Z. Jia, L. Yang, Z. Zhang, H. Liu, and F. Kong, "Sequence to point learning based on bidirectional dilated residual network for non-intrusive load monitoring," *International Journal of Electrical Power Energy Systems*, vol. 129, p. 106837, 2021.

[4] A. L. da Fonseca, K. M. Chvatal, and R. A. Fernandes, "Thermal comfort maintenance in demand response programs: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 141, p. 110847, 2021.

[5] W. L. Rodrigues Junior, F. A. Borges, R. d. A. Rabelo, J. J. Rodrigues, R. A. Fernandes, and I. N. da Silva, "A methodology for detection and classification of power quality disturbances using a real-time operating system in the context of home energy management systems," *International Journal of Energy Research*, vol. 45, no. 1, pp. 203–219, 2021.

[6] G. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[7] S. B. Leeb, S. R. Shaw, and J. L. Kirtley, "Transient event detection in spectral envelope estimates for nonintrusive load monitoring," *IEEE Transactions on Power Delivery*, vol. 10, no. 3, pp. 1200–1210, 1995.

[8] L. K. Norford and S. B. Leeb, "Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms," *Energy and Buildings*, vol. 24, no. 1, pp. 51–64, 1996.

[9] D. Srinivasan, W. Ng, and A. Liew, "Neural-network-based signature recognition for harmonic source identification," *IEEE Transactions on Power Delivery*, vol. 21, no. 1, pp. 398–405, 2006.

[10] R. A. S. Fernandes, I. N. da Silva, and M. Oleskovicz, "Load profile identification interface for consumer online monitoring purposes in smart grids," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 3, pp. 1507–1517, 2013.

[11] C. C. Yang, C. S. Soh, and V. V. Yap, "A systematic approach in appliance disaggregation using k-nearest neighbours and naive bayes classifiers for energy efficiency," *Energy Efficiency*, vol. 11, no. 1, pp. 239–259, 2018.

[12] D. L. Cavalca and R. A. Fernandes, "Deep transfer learning-based feature extraction: An approach to improve nonintrusive load monitoring," *IEEE Access*, vol. 9, pp. 139 328–139 335, 2021.

[13] J. Liang, S. K. Ng, G. Kendall, and J. W. Cheng, "Load signature study—part i: Basic concept, structure, and methodology," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 551–560, 2009.

[14] R. T. Caropreso, R. A. Fernandes, D. P. Osorio, and I. N. Silva, "An open-source framework for smart meters: Data communication and security traffic analysis," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1638–1647, 2019.

[15] R. T. Caropreso, R. A. Fernandes, and I. N. Silva, "Multi-node data exchange traffic analysis for a communication framework embedded on smart meters," *Electric Power Syst. Research*, vol. 205, p. 107734, 2022.

[16] Y. Zhou, Z. Shi, Z. Shi, Q. Gao, and L. Wu, "Disaggregating power consumption of commercial buildings based on the finite mixture model," *Applied Energy*, vol. 243, pp. 35–46, 2019.

[17] W. A. de Souza, F. D. Garcia, F. P. Marafão, L. C. P. Da Silva, and M. G. Simões, "Load disaggregation using microscopic power features and pattern recognition," *Energies*, vol. 12, no. 14, p. 2641, 2019.

[18] M. A. Mengistu, A. A. Girmay, C. Camarda, A. Acquaviva, and E. Patti, "A cloud-based on-line disaggregation algorithm for home appliance loads," *IEEE Trans. on Smart Grid*, vol. 10, no. 3, pp. 3430–3439, 2019.

[19] L. Pereira and N. Nunes, "An empirical exploration of performance metrics for event detection algorithms in non-intrusive load monitoring," *Sustainable Cities and Society*, vol. 62, p. 102399, 2020.

[20] D. Hua, F. Huang, L. Wang, and W. Chen, "Simultaneous disaggregation of multiple appliances based on non-intrusive load monitoring," *Electric Power Systems Research*, vol. 193, p. 106887, 2021.

[21] A. Langevin, M.-A. Carbonneau, M. Cheriet, and G. Gagnon, "Energy disaggregation using variational autoencoders," *Energy and Buildings*, vol. 254, p. 111623, 2022.

[22] H. Kim and S. Lim, "Temporal patternization of power signatures for appliance classification in nilm," *Energies*, vol. 14, no. 10, 2021.

[23] H. Zhou, J. Zhang, Y. Zhou, X. Guo, and Y. Ma, "A feature selection algorithm of decision tree based on feature weight," *Expert Systems with Applications*, vol. 164, p. 113842, 2021.

[24] J.-P. Eckmann, S. O. Kamphorst, and D. Ruell, "Recurrence plots of dynamical systems," *Europhys. Letters*, vol. 4, no. 9, pp. 973–977, 1987.

[25] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, "Nilmtk: An open source toolkit for non-intrusive load monitoring," *Proceedings of the 5th international conference on Future energy systems*, Jun 2014.