Performance of Scattering Transform Feature Extraction for Electrical Load Classification

Everton Luiz de Aguiar
Federal University of Technology
Curitiba - Brazil
Email: eaguiar@utfpr.edu.br

André Eugênio Lazzaretti
Federal University of Technology
Curitiba - Brazil
Email: lazzaretti@utfpr.edu.br

Daniel Rodrigues Pipa
Federal University of Technology
Curitiba - Brazil
Email: danielpipa@utfpr.edu.br

***Abstract*—The Scattering transform (ST) presents itself as an alternative approach to the classic methods that involve neural networks and deep learning techniques for the feature extraction and classification of signals and images. Among its main advantages, one can emphasize that the coefficients of the ST are determined analytically and do not need to be learned, as typically performed in Convolutional Neural Networks (CNNs). Additionally, ST has time-shifting and small time-warping invariance, which reduces the need for precise temporal localization (detection) for subsequent classification. This paper originally proposes the application of ST to extract features and classify Non-intrusive Load Monitoring (NILM) high-frequency signals. We validate the extraction strategy performance varying several parameters for ST calculation, such as signal length, number of examples, and sampling frequency. The results outperform other state-of-the-art feature extraction techniques, reaching up to 99.98% of accuracy and 99.51% of FScore for a publicly available dataset, demonstrating the feasibility and promising aspects of the ST for NILM problems.**

I. Introduction

Non-intrusive load monitoring, proposed in [1], comprises a set of computational signal processing solutions applied to residential electrical appliances signals. Typically, it is desired to identify (non-intrusively) the nature of each electrical load in a home (classification task) or to identify the respective energy consumption of each equipment from the aggregated signal (disaggregation task).

Fig. 1 shows the electrical appliances of a house. These devices are connected to the switchboard through independent electrical circuits. The smart-meter is connected to the switchboard, which interfaces the consumer unit with the electricity utility. The signal $\bar{X}$ in the Fig. 1 is called the aggregated signal.

The two main tasks of NILM are the classification and disaggregation. The classification, which is the main focus of this paper, consists of identifying the electric appliance or the set of electric appliances activated in a given instant of time. The steps for the NILM classification are (i) event detection, (ii) feature extraction, (iii) training of the classification model, and (iv) prediction [2].

As a consequence of the decrease in the cost of acquiring equipment and the increase in computational power, techniques for extracting features and classification using CNNs and deep learning have become more frequent in the literature. In the particular NILM field, they reached state-of-the-art results for disaggregation and classification [3].

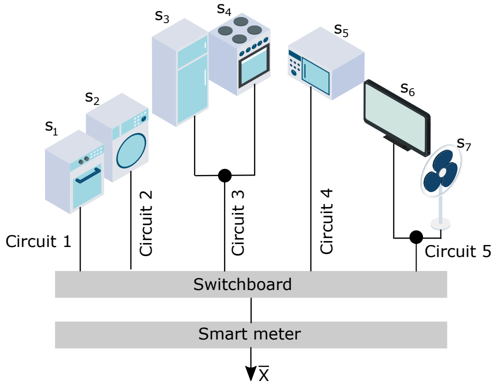


Fig. 1. Representation of the electrical appliances of a consumer unit.

The efficiency of extracting features for NILM using deep neural networks depends on the complexity of CNN and, therefore, directly influences the processing time. This means that better results are achieved at the expense of more time to learn the features of the signal and a larger database. These dependencies add to the overall complexity of the classification task.

The Scattering Transform, first proposed in [4], is a convolutional network composed of multi-scale filter banks that implement cascaded wavelets transforms. Such filters have analytically determined coefficients, i.e., they do not require the training process for convolutional layers. Each layer of the convolutional network gives rise to coefficients that can be used as features of the electrical signal for classification. This allows the feature extraction step to deal better than CNN with smaller datasets (with trained coefficients) and simplify the classification task. In addition, the overall complexity is reduced.

In this sense, we propose a framework for extracting features from NILM signals based on ST. The proposed framework contemplates classification task experiments with parametric variations of different parameters, such as signal length, number of examples by class, and sampling frequency. Such analysis is essential to validate the feature extraction strategy in different situations and define the most appropriate scenario for ST features. Additionally, we use a publicly available dataset (Controlled On/Off Loads Library - COOLL) with high-frequency signals for NILM proposes and the Ensemble method for classification.

This paper is divided as follows. Section II presents a literature review of Scattering Transform in classification tasks. Section III presents the proposed framework, and its subsections present each building block of the proposed framework. Section IV presents the experimental analysis under several scenarios and cases. Section V shows the comparisons of the proposed method with state-of-the-art approaches, and the Conclusions and Future Works are discussed in Section VI.

II. Related Works

The feature extraction method for NIlM classification is a fundamental step that directly influences the global performance of the classification task. Several recent works use neural networks and deep learning to extract characteristics and classify NILM loads [5]–[10]. The primary limitations found in those works are the need for a large set of data for training and testing and the demand for a learning stage for the coefficients of the neural network filters. In this context, the computational complexity is still an open challenge for future works [11].

Fields other than NILM widely use the wavelet transform to extract signal features for classification [12]–[19]. The time convolution used to implement the wavelet transform makes the feature extraction method dependent on the signal delay time. For NILM applications, this means that methods based on wavelet transform require an additional algorithm for detecting the turn-on instant of each load to be classified. That is an important limitation of wavelets to NILM.

The Scattering Transforms are convolutional networks composed of cascaded multi-scale wavelets transforms [4]. Filter banks implement the Scattering transform. These filter banks comprise a network of convolutional filters with analytically determined coefficients. This means that there is no training to determine the features of the signal when using ST as a features extractor for NILM. ST approach has the important property, as well as the deep convolutional networks, of building large-scale invariants which are stable to deformations [20]. Besides that, STs are time-shifting invariant [20]. Applying to NILM, this removes the need for an event detector algorithm.

Wiatowski et al. [21] presented a ST general mathematical approach that comprise and explain analytically the signal classification with deep convolutional neural networks. To date, the STs have been used to extract signal features of threat analysis [22], walk pattern [23], fingerprint classification [24], and scar detection in hospitalized patients [25].

We propose a feature extractor for NILM based on the Scattering Transform. We show that this type of feature extractor allows the classification results not to be substantially altered with the decrease in the number of training examples, the change of the sampling frequency, or the absence of single load connection event detection. We propose a testing framework to verify these properties, and we compare the evaluated classification results with baseline (state-of-the-art) methods.

III. Proposed Framework

Fig. 2 shows the proposed classification framework for NILM using Scattering Transform.

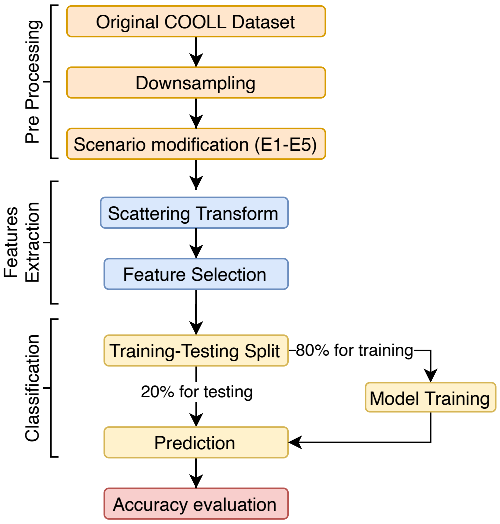


Fig. 2. Proposed framework for NILM Classification.

The following steps constitute the proposed framework: (i) Pre-processing, (ii) Feature Extraction, (iii) Classification, and (iv) Evaluation of Results. In the following subsections, we initially detail the dataset used, and later we will discuss each of the steps of the proposed framework (Fig. 2).

A. *COOLL Dataset*

Considering that the features extracted from high-frequency data are more discriminative than low-frequency data, we chose the Controlled On/Off Loads Library dataset (COOLL), proposed in [26]. The COOLL dataset has submetered instantaneous current records from 42 different electrical appliances at a sampling frequency of 100kHz. Each electric load has 20 samples of 6s duration, at 100kHz, totaling 840 samples, as presented in Table I.

B. *Pre-Processing*

The filter banks that implement the ST are multi-scale operators, and their implementation uses downsampling by two. For each downsampling by two, one of every two samples in the time domain is discarded, and the sampling frequency drops by half. For this reason, we changed the sampling frequency of the original input signal, coming from the COOLL dataset in the downsampling stage. The signal resulting from this step has a sampling frequency that is a power of 2 (8 192Hz), which makes the filter banks of the Scattering transform feasible. After downsampling, we changed the dataset based on five different scenarios:

TABLE I
COOLL Dataset.

Appliances	Num. of Appliances	Num. of Signals
Drill	6	120
Fan	2	40
Grinder	2	40
Hair dryer	4	80
Hedge trimmer	3	60
Lamp	4	80
Paint stripper	1	20
Planer	1	20
Router	1	20
Sander	3	60
Saw	8	160
Vacuum cleaner	7	140
Total	42	840

- **Scenario 1 (SC1):** Different number of cycles for the signal window, *with* turn-on events detection;
- **Scenario 2 (SC2):** Different number of cycles for the signal window, *without* turn-on events detection;
- **Scenario 3 (SC3):** Different number of examples per-class;
- **Scenario 4 (SC4):** Different sampling frequency;
- **Scenario 5 (SC5):** Whole signal-length, at 8 192Hz of sampling frequency.

C. Feature Extraction

We apply the Scattering Transform to extract features for NILM signals. So, we explain in this section the mathematical definitions needed to clarify the Scattering transform.

Consider a set of features determined by the representations of two signals f and g. Let $\Phi(f)$ and $\Phi(g)$ representations of the signals f and g signs, respectively. Then the Euclidean distance $d(f, g)$, defined by $d(f, g) = ||\Phi(f) - \Phi(g)||$ must be small for elements of the same class and large for elements of different classes [20]. The similarity measure between $\Phi(f)$ and $\Phi(g)$ depends on the inner product of the two representations. The central question of the classification is to define a good kernel[1], which allows for a reliable measurement of similarity [4].

Consider a signal $x(t)$, a translated signal $x_c = x(t - c)$ and a deformed (time-warped) sign $x_d = x(t - \tau(t))$ from the same type of electrical appliance. In real cases, there may be both translation and deformation of electrical signals of the same class. This occurs, for example, when the same load is switched on at different times in the same sampling window (translation) or when there is measurement noise (time-warping). The classifier should be invariant to translation and also to the small time-warping.

Let a wavelet $\Psi_\lambda(t)$ be defined by $\Psi_\lambda(t) = 2^{-jQ}\Psi(2^{-jQ}t)$, where $\lambda = 2^{-jQ}$, Q is the number of wavelets per octave and j is the scale factor. So the wavelet transform of $x(t)$ is

$$Wx(t) = \{x * \Phi(t), x * \Psi_\lambda(t)\}_\lambda. \tag{1}$$

The *wavelet* transform has the following advantages: (i) it is stable for small time deformations; (ii) it is well located both in time and in frequency, but it has the disadvantage of being time-shifting variant. This this happens because the Wavelet transform is calculated using convolutions [27]. To solve this problem, Mallat et al. [4] used the coefficients module, followed by the average in time:

$$\{|x * \Psi_\lambda| * \Phi(t)\}_\lambda, \tag{2}$$

being $\Phi(t)$ a low-pass filter that implements the average. The modulus and average operators guarantee the time-shifting invariance, but results in loss of information [4], [20].

Mallat et al. [4] proposed applying successive module and averaging operations to new layers of convolutions with wavelets starting from $\{|x * \Psi_\lambda| * \Phi(t)\}_\lambda$. This gives rise to the ST of $x(t)$, given by

$$s[p]x(t) = ||x * \Psi_{\lambda_1}| * \Psi_{\lambda_2}| \cdots * \Psi_{\lambda_m} | * \Phi(t), \tag{3}$$

depending on the order m and the path p. The total number of coefficients for the Scattering transform is $Q^m log_m N$.

We consider for the experiments $m = 2$, implemented by two filter banks. The first filter bank bank, corresponding to the first layer, has $Q = 8$. The second filter bank has $Q = 1$. The choice of the ST structure is based on [28].

Fig. 3 shows graphically the structure of the ST. The electrical signal x, passes through a first layer of convolutions with the wavelets $\Psi_{m,j} = \Psi_{1,j_1}$. The subscript j_1 is the frequency scale of the first filter bank. The convolution module is taken at each node of the tree in the first layer. The output of each node in the first layer is used to calculate the second level of the convolutional network.

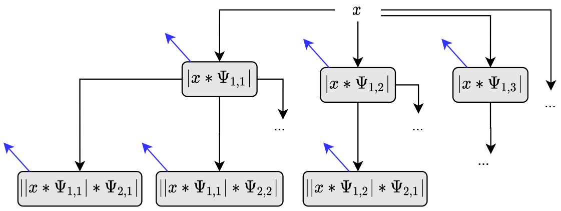

x
$|x * \Psi_{1,1}|$
$|x * \Psi_{1,2}|$
$|x * \Psi_{1,3}|$
...
...
...
...
$||x * \Psi_{1,1}| * \Psi_{2,1}|$
$||x * \Psi_{1,1}| * \Psi_{2,2}|$
$||x * \Psi_{1,2}| * \Psi_{2,1}|$

Fig. 3. Scattering Transform Structure.

At the second layer, the convolution of $|x * \Psi_{1,j_1}|$ with a second set of wavelets, $\Psi_{m,j} = \Psi_{2,j_2}$ results in a new set of coefficients. The index j_2 represents the second frequency scale of the transform ($m = 2$), implemented by the second filter bank. Each second layer node comes from convolution-modulus with Ψ_{2,j_2}.

The blue arrows in Fig. 3 represent the convolution operation with the low-pass filter Φ_T. The cutoff frequency of Φ_T is half the bandwidth on the j scale. We propose to use the results of those convolutions to construct features used for classification.

[1] It is an operator f that takes elements from set A to set B, which preserves the form (homeomorph). The kernel of this operator f is the inverse image of $0 \in B$.

1) Proposed feature selection technique: We determine the features matrix employing three strategies: (i) by taking the average of first order-path Scattering coefficients; (ii) by taking the energy of the first order-path Scattering Coefficients, and (iii) by taking all the first order-path Scattering Coefficients.

Let $S_{1,i} = |x * \Psi_{1,i}| * \Phi$ the first order-path coefficients of the i sub-band. Then, we compute the features by the averaging method ($A_f[i]$), for the i-th sub-band, as

$$A_f[i] = \frac{1}{n_f} \sum_{n_f} S_{1,i}, \tag{4}$$

where n_f is the number of coefficients at the i sub-band.

As a second strategy, we compute the energy of the first order coefficients, $E_f[i]$, as

$$E_f[i] = \sum_{n_f} S_{1,i}^2. \tag{5}$$

We use the library *Wavelet Scattering*, from Matlab® r2021, to implement the Scattering transform.

D. Classification

We perform the classification task considering the scenarios presented in subsection III-B and the strategies of feature selection presented in subsection III-C1. The training and test sets are separated from the feature matrix and the label vector. From the total number of instances, 80% were used to train the classifier and 20% for test.

We train the classification models using Ensemble Method (ENS). This method comprises a set of classifiers whose individual decisions are combined in some way – normally averaging – to classify new examples. In classification methods, ensembles are often much more accurate than the individual classifiers that make them up [29]. Ensemble classification combines a set of trained weak learner models. It can predict ensemble responses for new data by aggregating predictions from its weak learners. This method can use different algorithms for sequential learning (weaker learning models), such as *AdaBoostM1, AdaBoostM2 Bag, GentleBoost, LogitBoost, LPBoost, LSBoost, RobustBoost, RUSBoost, Subspace*, and *TotalBoost*.

After training the classification model, we perform the prediction with the test subset (for each scenario). Let n_c be the number of classes of the dataset. For COOLL, $n_c = 42$. The prediction, for each experiment, results in a $\mathbf{M}_{\mathbf{n_c \times n_c}}$ confusion matrix:

$$\mathbf{M} = \begin{bmatrix} a_{1,1} & \dots & a_{1,n_c} \\ \vdots & \ddots & \vdots \\ a_{n_c,1} & \dots & a_{n_c,n_c} \end{bmatrix}, \tag{6}$$

whose rows represent the predicted classes and the columns represent the actual classes. From the confusion matrix we calculated two performance metrics: FScore and Accuracy. The FScore, for each i-th class, is defined by

$$\text{FScore}_i = \frac{2 \times \text{Recall}_i \times \text{Precision}_i}{\text{Recall}_i + \text{Precision}_i}, \tag{7}$$

where

$$\text{Precision}_i = \frac{a_{i,i}}{\sum_{k=1}^{n_c} a_{i,k}}, \tag{8}$$

and

$$\text{Recall}_i = \frac{a_{i,i}}{\sum_{k=1}^{n_c} a_{k,i}}. \tag{9}$$

The accuracy for each class is

$$\text{Accuracy}_i = \frac{\text{TP}_i + \text{TN}_i}{\text{TP}_i + \text{TN}_i + \text{FN}_i + \text{FP}_i}, \tag{10}$$

were $\text{FN}_i = (\sum_{k=1}^{n_c} a_{k,i}) - a_{i,i}$, $\text{TN}_i = (\sum_{k=1}^{n_c} a_{k,k}) - a_{i,i}$, $\text{TP}_i = a_{i,i}$, and $\text{FP}_i = (\sum_{k=1}^{n_c} a_{i,k}) - a_{i,i}$.

From the FScore_i and Accuracy_i per class, we calculate the macro FScore and macro Accuracy by the expressions

$$\text{FScore}_{\text{macro}} = \frac{1}{n_c} \sum_{i=1}^{n_c} \text{FScore}_i, \tag{11}$$

and

$$\text{Accuracy}_{\text{macro}} = \frac{1}{n_c} \sum_{i=1}^{n_c} \text{Accuracy}_i. \tag{12}$$

For simplification, we refer to $\text{FScore}_{\text{macro}}$ and $\text{Accuracy}_{\text{macro}}$ as FScore and Accuracy in the next sections.

IV. Experimental Analysis

We show in this section the results obtained from the five scenarios derived from the COOLL dataset and described in subsection III-B. As a baseline comparison, we performed the experiments using both the proposed method (ST) and the Discrete Wavelet Transform (DWT) as feature extraction strategies. The investigations follow the structure of Fig. 1. The choice of DWT as a baseline is because both ST and DWT are based on wavelet filters [27].

We modify the input signal from the original dataset for all experiments to analyze the performance metrics for the classification process. The modifications are detailed in Table II

First, we present the setup of the Scattering Transform for the experiments. Then, we show the setup of the DWT baseline extraction method. At the end of this section, we present the classification results and discussions.

A. Scattering Transform (ST) Experimental Setup

For the feature extraction, we parameterized the ST as follows:

- **Sampling Frequency (Fs):** 8 192Hz for Scenarios SC1, SC2, SC3, and SC5. 4096Hz and 2048Hz for SC4;
- **Number of Layers (m):** 2 layers;
- **Number of Filter Banks:** 2;
- **Number of wavelets per-octave, or Quality Factor (Q):** 8 for first layer, and 1 for second layer;
- **Type of Wavelet Filters:** Complex Morlet.

TABLE II
SCENARIOS AND EXPERIMENTAL CASES DETAILS.

		Number of 60Hz Cycles						
Scenario	Case	N_{before}	N_{after}	Total time by example	Examples by class	Samples by example	Total of examples	Fs [Hz]
SC1	1	5	5	166.67ms	20	1 365	840	8 192
	2	10	10	333.33ms		2 731		
	3	20	20	666.67ms		5 461		
SC2	1	0	5	83.33ms	20	683	840	8 192
	2		10	166.67ms		1 365		
	3		50	833.33ms		6 827		
	4		100	1.67s		13 653		
		Number of Cycles:						
SC3	1	240		4s	10	32 768	420	8 192
	2				15		630	
	3				20		840	
SC4	1	240		4s	20	16384	840	4 096
	2			4s		8 192		2 048
SC5	1	360		6s	20	49 152	840	8 192

B. Discrete Wavelet Transform (DWT) Baseline Experimental Setup

For DWT, we use ten layers of detail signals. For each layer, we compute the energy of the wavelet coefficients. Then, we use the ten energies of each detail layer as features for the baseline classification model. We add to these features the energy of the approximation coefficients [27], totaling 11 features for DWT baseline. The algorithm we use for DWT implementation is based on [30]–[32], implemented by function *wavedec*, on Matlab® r2021.

C. Results and Discussions

We follow the structure of Fig. 1 and trained five classification models for each classifier in section II-D. We use different training-test sets for each one of those classification models (five-fold cross-validation). Then, we perform the prediction five times for the Ensemble classifier, one for each different training-test set. Finally, we compute FScore and Accuracy for each trained model and each class. With these metrics, we calculate the macro FScore and Macro Accuracy of Table III. To facilitate visualization, we also present the results of Table III in Fig. 4.

Fig. 4a shows that with the smaller cycle window (10 cycles), both DWT and ST have the same accuracy (99.84%) and FScore (96.48%). For 20 cycles window, both methods present the same accuracy (99.98%), but DWT shows slightly better FScore than ST (99.66% vs. 99.48%). On the other hand, with 40 cycles, the ST method shows marginally better accuracy than DWT (99.98% vs. 99.94%). In this case, ST presents a significantly better global FScore than DWT (99.51% vs. 98.55%).

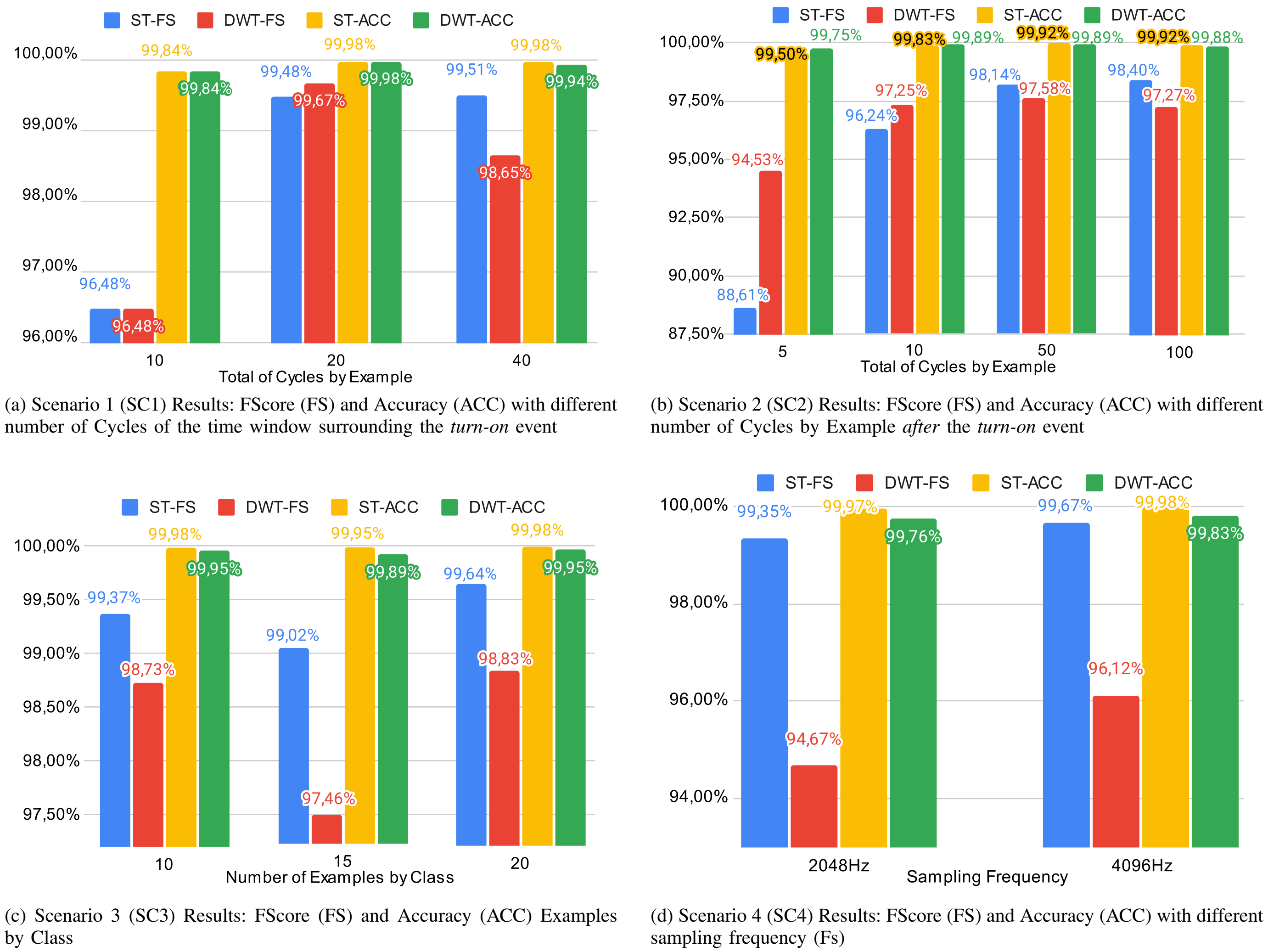
Similar to SC1 results, in scenario 2 (Fig. 4b), DWT presents slightly better FScores and Accuracy than ST for five cycles: FScore 94.53% vs. 88.61% and Accuracy 99.75% vs. 99.50%. For 10 cycles per sample: FScore 97.25% vs. 96.24% and Accuracy 99.89% vs. 99.83%. For the cases with 50 and 100 cycles per sample, ST outperforms DWT in both FScore and Accuracy. FScore 98.14% vs. 97.58% and Accuracy 99.92% vs. 99.89% for 50 cycles, and FScore 98.40% vs. 97.27% and Accuracy 99.92% vs. 99.88% for 100 cycles per sample.

TABLE III
MACRO ACCURACIES AND FSCORES FOR ALL SCENARIOS AND CASES.

			Accuracy		**FScore**	
Scn.	**Case**	**Description**	**ST**	**DWT**	**ST**	**DWT**
SC1	1	5 cycles	99.84%	99.84%	96.48%	96.48%
	2	10 cycles	99.98%	99.98%	99.48%	99.67%
	3	20 cycles	**99.98%**	99.94%	**99.51%**	98.65%
SC2	1	5 cycles	99.50%	99.75%	88.61%	94.53%
	2	10 cycles	99.83%	99.89%	96.24%	97.25%
	3	50 cycles	**99.92%**	99.89%	**98.14%**	97.58%
	4	100 cycles	**99.92%**	99.88%	**98.40%**	97.27%
SC3	1	10 examples	**99.98%**	99.95%	99.37%	98.73%
	2	15 examples	**99.95%**	99.89%	**99.02%**	97.46%
	3	20 examples	**99.98%**	99.95%	99.64%	98.83%
SC4	1	2048Hz	**99.97%**	99.76%	**99.35%**	94.67%
	2	4096Hz	**99.98%**	99.83%	**99.67%**	96.12%
SC5	1	6s samples	**99.95%**	99.91%	**98.86%**	98.02%

Fig. 4c shows that ST has better performance than DWT for all cases of Scenario 3. The macro FScore is 0.65% better with ten examples by class, 1.6% with 15 examples, and 0.82% with 20 samples per class. The improvements in accuracy are smaller than the improvements in FScore: 0.03% in 10 examples by class, 0.06% in 15 examples, and 0.03% in 20 examples.

ST presents better FScore and Accuracy than DWT for the 2 cases of Scenario 4, as shown in Fig. 4d. This indicates that ST deals better with sub-sampling conditions than DWT for COOLL dataset (considering Ensemble classifier). Fig. 4d also shows that both ST and DWT increase their performance metrics (FScore and Accuracy) when the sampling frequency



(a) Scenario 1 (SC1) Results: FScore (FS) and Accuracy (ACC) with different number of Cycles of the time window surrounding the *turn-on* event

(b) Scenario 2 (SC2) Results: FScore (FS) and Accuracy (ACC) with different number of Cycles by Example *after* the *turn-on* event

(c) Scenario 3 (SC3) Results: FScore (FS) and Accuracy (ACC) Examples by Class

(d) Scenario 4 (SC4) Results: FScore (FS) and Accuracy (ACC) with different sampling frequency (Fs)

Fig. 4. Macro Accuracy (ACC) and Macro FScore (FS) results for different experimental Scenarios using Scattering Transform (ST) and Discrete Wavelet Transform (DWT) feature Extractors.

rises. From 99.35% to 99.67% and from 99.97% to 99.98% for ST (FScore and Accuracy, respectively), and from 94.67% to 96.12% of FScore and from 99.76% to 99.83% of Accuracy for DWT.

V. Comparisons with State-of-the-Art Approaches

Other state-of-the-art works presented features extraction strategies applied to COOLL dataset to classification tasks. Table IV shows the comparison of the classification results metrics between those approaches and the proposed method. For all values of FScore and Accuracy in Table IV, the better scenario for each method is considered.

As shown in Table IV, the proposed method presents an improvement of Accuracy of 8.67% in relation to [33], 0.54% to [34], and 2.02% in relation to [35].

TABLE IV
Comparison between literature approaches and the proposed method.

Reference	Method	FScore [%]	Accuracy [%]
[33]	Traditional Wavelets	-	92.00%
[34]	Hybrid V-I Trajectory	-	99.44%
[35]	Prony	-	98.00%
Proposed	Scattering Transform	**99.51%**	**99.98%**

VI. Conclusions and Future Works

This paper proposed a classification framework using a Scattering Transform-based feature extraction method applied to NILM. Such framework includes important variations in the dataset properties: signal length, number of examples by class, and sampling frequency. The two main contributions of our work are: (i) apply the Scattering transform to improve NILM state-of-the-art classification results and (ii) validate these improvements under dataset properties variations.

We perform experiments with five different scenarios and with several cases by scenario. From SC1, the proposed method outperformed the DWT baseline for cases with more cycles by example (20 and 40 for SC1; 50 and 100 for SC2). The experiments with Scenario SC3 showed that the proposed method outperformed DWT when reducing the number of examples per class for all evaluated cases. We also conclude, with SC4 results, that Scattering Transform presented better classification metrics for downsampling conditions in the dataset. Finally, the proposed method resulted in better macro accuracy when compared to state-of-the-art methods of feature extraction for NILM in the literature, as shown in Table IV.

The need for high frequency data is an important limitation of our proposed framework. Currently high frequency data is not directly found in traditional smart-meters, which can limit the embedded implementation of our proposal. The implementation and analysis of our framework with low frequency data is a proposal for future work.

Also for future works, the authors intend to implement the proposed framework in aggregated data and perform a comparison between Scattering Transform and Convolutional Neural Networks (CNN) in NILM disaggregation and classification tasks.

Acknowledgments

The authors acknowledge the financial support from CNPq grant 311726/2018-6. The authors would like to thank UTFPR, CAPES, Fundação Araucária, FINEP and CNPq for additional funding. E. L. Aguiar thanks to the Brazilian Ministry of Education for the license to develop this research.

References

[1] G. W. Hart, "Nonintrusive Appliance Load Monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[2] M. H. Bollen and I. Gu, *Event Classification*, 2006, pp. 677–733.

[3] D. Yang, X. Gao, L. Kong, Y. Pang, and B. Zhou, "An Event-Driven Convolutional Neural Architecture for Non-Intrusive Load Monitoring of Residential Appliance," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 2, pp. 173–182, 2020.

[4] S. Mallat, "Group Invariant Scattering," *Communications on Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, 2012.

[5] P. Paulo, "Applications of Deep Learning Techniques on Nilm.Ph.D. Thesis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil," pp. 4–6, 2016.

[6] Z. Lan, B. Yin, T. Wang, and G. Zuo, "A non-intrusive load identification method based on convolution neural network," *2017 IEEE Conference on Energy Internet and Energy System Integration, EI2 2017 - Proceedings*, vol. 2018-Janua, pp. 1–5, 2017.

[7] P. Dash and K. Naik, "A Very Deep One Dimensional Convolutional Neural Network (VDOCNN) for Appliance Power Signature Classification," *2018 IEEE Electrical Power and Energy Conference, EPEC 2018*, pp. 1–6, 2018.

[8] M. Sun, F. M. Nakoty, Q. Liu, X. Liu, Y. Yang, and T. Shen, "Non-Intrusive Load Monitoring System Framework and Load Disaggregation Algorithms: A Survey," *International Conference on Advanced Mechatronic Systems, ICAMechS*, vol. 2019-August, pp. 284–288, 2019.

[9] M. Mobasher-Kashani, J. Li, and S. Luo, "Light-weight recurrent deep learning algorithm for non-intrusive load monitoring," *Proceedings of 2019 IEEE 2nd International Conference on Electronic Information and Communication Technology, ICEICT 2019*, pp. 572–575, 2019.

[10] J. Jiang, Q. Kong, M. Plumbley, and N. Gilbert, "Deep Learning Based Energy Disaggregation and On/Off Detection of Household Appliances," vol. 1, no. 1, pp. 1–19, 2019. [Online]. Available: http://arxiv.org/abs/1908.00941

[11] F. K. K. Kee, Y. S. Lim, J. Wong, and K. H. Chua, "Non-Intrusive Load Monitoring (NILM)-A Recent Review with Cloud Computing," *2019 IEEE 6th International Conference on Smart Instrumentation, Measurement and Application, ICSIMA 2019*, no. August, pp. 1–6, 2019.

[12] T. Randen and J. H. Husoy, "Filtering for texture classification: a comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 291–310, 1999.

[13] A. J. Gakare, K. R. Singh, and J. F. Peters, "Wavelet-based tolerance near set approach in classifying hand images: A review," in *2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave)*, 2016, pp. 1–4.

[14] P. Lamba and K. Rawal, "A survey of algorithms for feature extraction and feature classification methods," in *2019 International Conference on Automation, Computational and Technology Management (ICACTM)*, 2019, pp. 338–341.

[15] A. I. Poernama, I. Soesanti, and O. Wahyunggoro, "Feature extraction and feature selection methods in classification of brain mri images: A review," in *2019 International Biomedical Instrumentation and Technology Conference (IBITeC)*, vol. 1, 2019, pp. 58–63.

[16] G. Y. Tian, A. Sophian, D. Taylor, and J. Rudlin, "Wavelet-based pca defect classification and quantification for pulsed eddy current ndt," *IEE Proceedings - Science, Measurement and Technology*, vol. 152, no. 4, pp. 141–148, 2005.

[17] C. J. Mariya and K. A. Nyni, "Review on feature extraction methods in neuromuscular disease diagnosis," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 519–522.

[18] C. J. Long and S. Datta, "Wavelet based feature extraction for phoneme recognition," in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, vol. 1, 1996, pp. 264–267 vol.1.

[19] S. D. Kalbhor and V. K. Harpale, "The review of detection and classification of epilectic seizures using wavelet transform," in *2016 International Conference on Computing Communication Control and automation (ICCUBEA)*, 2016, pp. 1–5.

[20] J. Bruna, "Invariant Scattering Convolution Networks," vol. 35, no. 8, pp. 1872–1886, 2013.

[21] T. Wiatowski and H. Bolcskei, "A Mathematical Theory of Deep Convolutional Neural Networks for Feature Extraction," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1845–1866, 2018.

[22] K. J. Bryan, K. E. Smith, M. Solomon, D. A. Clauter, A. O. Smith, and A. M. Peter, "Deep wavelet scattering features for infrasonic threat identification," no. May, p. 43, 2018.

[23] S. Kitić, G. Puy, P. Pérez, and P. Gilberton, "Scattering features for multimodal gait recognition," *2017 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2017 - Proceedings*, vol. 2018-Janua, pp. 843–847, 2018.

[24] B. Soro and C. Lee, "A wavelet scattering feature extraction approach for deep neural network based indoor fingerprinting localization," *Sensors (Switzerland)*, vol. 19, no. 8, 2019.

[25] Y. Jiang, W. Chen, and Y. You, "Scattering transform-based features for the automatic seizure detection," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 77–89, 2020. [Online]. Available: https://doi.org/10.1016/j.bbe.2019.11.002

[26] T. Picon, M. Nait Meziane, P. Ravier, G. Lamarque, C. Novello, J.-C. Le Bunetel, and Y. Raingeaud, "COOLL: Controlled on/off loads library, a public dataset of high-sampled electrical signals for appliance identification," *arXiv preprint arXiv:1611.05803 [cs.OH]*, 2016.

[27] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms A Primer*.

[28] J. Andén, V. Lostanlen, and S. Mallat, "Joint Time-Frequency Scattering," *IEEE Transactions on Signal Processing*, vol. 67, no. 14, pp. 3704–3718, 2019.

[29] V. S. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory, and Methods*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1998.

[30] Y. Meyer, *Wavelets and Operators*, ser. Cambridge Studies in Advanced Mathematics, D. H. Salinger, Ed. Cambridge University Press, 1993, vol. 1.

[31] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *Fundamental Papers in Wavelet Theory*, vol. II, no. 7, pp. 494–513, 1989.

[32] I. Daubechies, "The Wavelet Transform, Time-Frequency Localization and Signal Analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[33] D. P. B. Renaux, C. R. E. Lima, F. Pöttker, E. Oroski, A. E. Lazzaretti, R. R. Linhares, A. R. Almeida, A. O. Coelho, and M. C. Hercules, "Non-intrusive load monitoring: an architecture and its evaluation for

power electronics loads," in *2018 IEEE International Power Electronics and Application Conference and Exposition (PEAC)*, 2018, pp. 1–6.

[34] 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)*, 2019, pp. 1–6.

[35] H. C. Ancelmo, F. L. Grando, B. M. Mulinari, C. H. da Costa, A. E. Lazzaretti, E. Oroski, D. P. B. Renaux, F. Pottker, C. R. E. Lima, and R. R. Linhares, "A transient and steady-state power signature feature extraction using different prony's methods," in *2019 20th International Conference on Intelligent System Application to Power Systems (ISAP)*, 2019, pp. 1–6.