

Classification of load waveform distortion signature based on novelty detection for electric railway systems[☆]

Rafael S. Salles^{a,*}, Sarah K. Rönnberg^a, Andrea Mariscotti^b

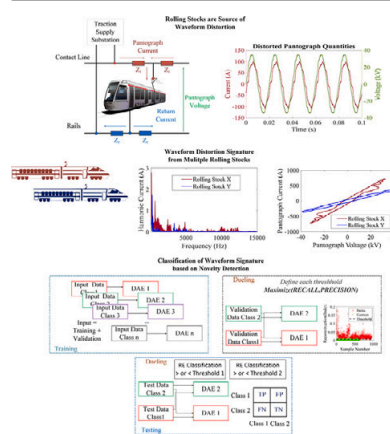
^a Electric Power Engineering Group, Luleå University of Technology, Skellefteå, 93187, Sweden

^b Department of Electrical, Electronics and Telecommunication Engineering and Naval Architecture (DITEN), University of Genoa, Genoa, 16145, Italy

HIGHLIGHTS

- A load classification method is proposed based on novelty detection applicable to distortion signatures of connected loads.
- A deep autoencoder (DAE) for each class learns its own features and confronts those of the other classes' DAE, having optimized recall and precision performance during the training phase.
- Accuracy is very high (often >99 %) for both the target problem (rolling stock in electrified railways) and a reference benchmark dataset with home appliances.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Load signature
Non-intrusive load monitoring
Deep learning
Power quality
Harmonics
Guideway transportation power systems

ABSTRACT

Electric railway systems (ERS) are characterized by several particularities regarding the return current circuits, moving loads, multiple sources of waveform distortion, and extensive deployment of static converters from various manufacturers, topologies, and solutions. This work presents a methodology for application of load monitoring to classify rolling stock (RS) waveform distortion signatures. The proposed methodology combines the benefits and advantages of unsupervised deep learning and reconstruction error performance classification for performing non-intrusive load monitoring (NILM) in ERS. It consists of adapting autoencoder-based novelty detection for load classification problems. The method is applied to pantograph measurements from four rolling stock items using two types of data input (harmonic spectra up to 15 kHz and VI diagram images), which are compared in binary classifications of the same kind of railway electrification. The methodology shows suitable classification performance with high accuracy, scoring an average of 98.81 % for spectrum input and 97.77 % for VI diagram input. It has also been validated with a NILM dataset (LIT) for multi-class applications showing 99.13 % for spectrum input and 94.28 % for VI diagram input. The proposed method has suitable computational times and scalability, allowing application to a wide range of NILM and classification problems using distortion signatures.

[☆] This work is funded by the Swedish Transport Administration. This research is performed in the scope of the project “23IND06 Met4EVCS Metrology for electric vehicle charging systems”. The project has received funding from the European Partnership on Metrology, co-financed by the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

* Corresponding author.

Email addresses: rafael.de.souza.salles@ltu.se (R.S. Salles), sarah.ronnberg@ltu.se (S.K. Rönnberg), andrea.mariscotti@unige.it (A. Mariscotti).

<https://doi.org/10.1016/j.segan.2025.102010>

Received 13 January 2025; Received in revised form 16 September 2025; Accepted 14 October 2025

Available online 17 October 2025

2352-4677/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In electric railway systems (ERS), static power converters onboard rolling stock (RS) are a significant source of waveform distortion [1,2]. ERS are characterized by several particularities regarding the return current circuit, moving loads, multiple sources of distortion (harmonics, interharmonics, and supraharmonics), and extensive deployment of static converters with a variety of manufacturers and topologies [3]. In this perspective, monitoring of the loads is useful to support operation and investigate problems. Monitoring is often achieved using distortion and power quality (PQ) parameters, including as a necessity load disaggregation and identification.

Non-intrusive load monitoring (NILM) methods can in fact support grid management, protection, diagnostics, PQ and condition assessment. NILM disaggregates load profiles at a common global point of measurement (e.g., point of common coupling) [4,5], providing information supporting energy management, protection, PQ analysis, etc. Different electrical quantities and indices can be considered [6]: active/reactive power, phase angle, VI diagrams, frequency deviation, harmonic and supraharmonic distortion, crest factor, and even electromagnetic radiated emissions.

High-frequency datasets and applications are crucial for NILM because they enable event-driven, real-world applicable frameworks that can accurately disaggregate multi-state appliances while handling noise, computational limits, and practical challenges [7]. Some works have analyzed industrial and domestic appliances. Current waveform distortion segmentation and feature extraction can support the analysis of appliance signatures in different operating conditions considering time and frequency domain characteristics [8,9]. Parameters like harmonic components (both amplitude and phase angle) and admittance spectra have shown usefulness in different algorithms for the load disaggregation task [10–12]. The frequency-domain sparseness can also be used for the decomposition of current waveforms and identification based on characteristic filters without feature extraction [13]. Another phenomenon inherent in the power electronics signature and non-linearity is the unintended radiated emission, exploiting spectral characteristics for load discrimination [14,15]. In [16], a combined approach involving digital twin and NILM was utilized, and the framework employs high-frequency current measurements to extract odd harmonic components, specifically the third and fifth, which are represented in scatter plots and later classified using a convolutional neural network (CNN). Using the PLAID dataset, the work evaluated the methodology across three scenarios: single appliance operation, combined appliance operation, and event detection (appliance activation and deactivation). The results demonstrate high classification accuracy and rapid processing times, with the entire identification process completed in under one second.

Lissajous (or VI) diagrams are conversely a time-domain representation of the waveshapes [17]. In [18], the assessment of load signatures was introduced using a VI diagrams and hierarchical clustering to classify household appliances, providing a taxonomy comparison with traditional metrics. Such a framework was further expanded, showing improved precision and robustness in a multi-class classification problem [5]. The similarity of VI diagrams among loads was also explored using binary mapping [8]. Based on time-domain power theory [19], two new load signatures based on non-active orthogonal current components were considered. In [20], a mathematical feature extraction was presented, which describes the transient and steady-state load signature using a 2-D Fourier series, showing suitable accuracy. Traditionally, harmonic active power terms have been found to be more informative despite some ambiguity in exceptional cases [4]. Additionally, focusing on ERSs and RS, different operating points of one RS type were characterized using unique VI diagrams and orthogonal harmonic active and reactive power spectra [4].

In [21], the proposed application introduced a novel current harmonic distortion image feature (based on percent distortion of the third and fifth order harmonics over 60 current cycles) extracted from

high-sampling-rate aggregated current, which enables robust appliance classification. It proposed a time-driven CNN framework using the proposed image input to classify both single and combined appliance operations within 1-second windows, achieving real-time performance with acceptable computational complexity. Also, a new public high-frequency dataset that captures real-world aggregated loads plus behind-the-meter distributed generation is proposed in [22], enabling the study of how photovoltaic generation affects load identification. It extends NILM methods based on Inception Time or CNN by performing both DG identification and load disaggregation in the presence of DG, achieving high accuracy in both distributed generation detection and load disaggregation. In the same direction, NILM has been performed and explored within different methods under complex background conditions, using models such as CNN or denoising autoencoder with adaptive neural tree [23–25]. Such works highlighted the need to consider complex background features and their impact on the performance of the NILM algorithm.

Other methods were explored, applying CNN architectures to high-frequency datasets for NILM classification using a variety of feature extraction techniques: color-encoding of VI diagrams [26], generative adversarial networks [27], wavelet scattering transform [28,29], frequency invariant transformation of periodic signals [30], adaptive weighted recurrence graphs [31], etc. Other examples also explored interesting representations of the high-frequency and waveform distortion, considering transformations in the input data. The paper in [32] proposed a NILM method that uses inrush current patterns, transformed into images via kurtograms, and performs classification of them with VGGNet-based CNNs, achieving high accuracy for appliance identification. Gramian Angular Fields were also utilized together with benchmark CNN models for NILM application in household load disaggregation [33,34]. Additionally, two-stage CNNs were used, where the first stage extracts learnable features and the second provides the classification [34].

Some other works did not incorporate learning techniques, exploring time-domain and waveform distortion characteristics, to deploy signal processing techniques such as dynamic time warping and profiling discrete harmonic cycle time [35,36], most of the works in the literature explore signal processing techniques combined with machine learning, pattern matching, or source separation for NILM performance [6]. Unsupervised deep learning techniques, such as deep autoencoder (DAE), are not designed for classification tasks; they are instead more suitable for feature extraction, pattern recognition, and event or anomaly detection. Nevertheless, autoencoder-based anomaly, novelty or rare-event detection can be exploited for classification methodology [37].

Besides the high number of previous literature contributions, NILM has been developed mainly for residential loads, investigating signature parameters related to the non-ideal behavior of small-power single-phase loads. In contrast, few studies on residential and light industries have explored distortion patterns using high-frequency measurements. Medium and large power scenarios in industrial systems are rarely considered where the significant challenges are [38,39]: large dynamics of loads, e.g., tooling, pumping, or, as the in the present case, operating conditions of rail vehicles such as transitioning stages and braking with reverse power flow; the diversity of industrial loads and the need for high-resolution data to achieve accurate disaggregation; and reliability of measurement devices in industrial environments. NILM using high-frequency data can support detection and provide more significant inferences into industrial load behavior, potentially improving accuracy and performance of online NILM systems in the future [40].

This paper aims to perform load classification for RS distortion signatures using current spectra and VI diagrams as data input. After training multiple DAE networks, one for each class, the reconstruction error (RE) performance determines the classification with the cross-class prediction between the networks. In other words, each network is tested with its assigned data and all other datasets from other networks. The output from

a network aimed at class 1 will present a poor performance index when predicting data input from class 2 (“1” and “2” are exemplifications). High accuracy classification can be achieved for loads with distinguishable signatures. The framework is constituted by data input preparation, DAE network training, definition of thresholds from validation data, and testing. This proposed process for load signature classification is based on novelty detection. Novelty detection is the task of classifying test data that have different characteristics, or differ in the investigated parameters, from the model training data [41]. The model learns the pattern of a normal process; anything that does not follow that pattern is classified as novelty. It is analogous to anomaly detection problems, but instead of anomalies or novelties, in the proposed method we compare two different patterns from different classes, evaluating binary classification performance. In case of multi-class problems, a binary classification pattern is followed between the class of interest and all other classes, then cascading the classification for each class.

The work applies the “novelty detection methodology” to four types of locomotives using pantograph measurements from different ERSs: two from 16.7 Hz 15 kV systems (Germany and Switzerland) and two from 50 Hz 2×25 kV systems (France and Italy). The results describe classifying a pair of rolling stock items from the same system. The load classification framework provides a practical and innovative application for railways for load monitoring using waveform distortion parameters. Load classification and monitoring can support the identification of highly consuming loads and the prioritization of energy-saving policies, including pricing and billing of electric energy. The proposed method based on features of the absorbed current allows distinguishing loads from the substation as a support to, or in competition with, onboard energy metering, as in the EN 50463-2 [42].

Additionally, the methodology is applied to a benchmark dataset, the LIT-dataset [43], to validate the framework for other types of measurements and systems in load monitoring applications.

The main contributions of this paper are listed below.

1. Proposing a methodology for load signature classification in terms of waveform distortion parameters that combines the benefits of unsupervised deep learning and novelty detection.
2. Providing a load signature classification method that demonstrates suitable performance and overcomes drawbacks for data with the following characteristics: unbalanced input data, high dimensionality, representations based on 1-D or 2-D, and high variety.
3. Defining a framework for load disaggregation based on waveform distortion signatures from rolling stocks for better assessment of PQ in ERS considering its complexity. Such framework has potential for application of NILM in other kinds of industrial power systems, having been validated also with a publicly available benchmark NILM dataset with a multi-class problem.

The rest of the work is structured as follows. Section 2 provides a brief discussion on waveform distortion signatures of RS. Section 3 provides a description of the methodology. Section 5 reports and describes the results and Section 6 presents a discussion regarding advantages and disadvantages of the method. Lastly, Section 7 concludes the work.

2. Waveform distortion signature of rolling stock

The rolling stock (RS) drive uses alternating current (AC) back-to-back conversion, based on diode or thyristor-based rectifiers for older trains and interleaved 4QCs for more modern trains [44]. Such static power converters cause current waveform distortion and consequential disturbances up to tens of kHz, leading to problems when interacting with ERS supply grid and other critical systems [45]. The use of pulse-width modulation (PWM) implies a significant amount of sideband harmonics located around the multiples of the carrier frequency. Fig. 1 shows a generic illustration of how RS current causes voltage distortion when interacting with the ERS supply system impedance.

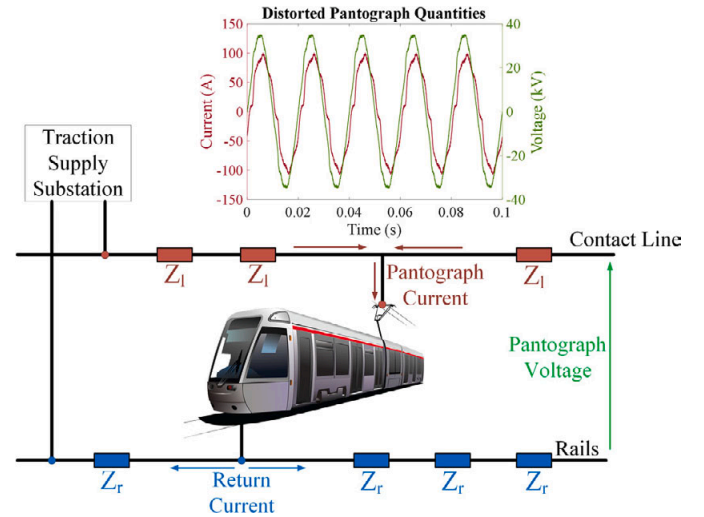


Fig. 1. Generic illustration of how an RS distorts the current and voltage in ERS.

The ERS power supply is constituted mainly by the traction supply substation connected to the moving loads through the catenary circuit, from which the RS collects the current that flows back through the rails and return conductors. For AC systems, the most common electrification solution is 50 Hz 1×25 or 2×25 kV systems and 16.7 Hz 15 kV systems: the former ERS systems are connected directly to the public transmission grid through transformers, whereas the latter are decoupled through traction converter stations, using frequency converters. For urban transportation and some railway systems, DC systems are a popular solution, especially for small or medium distances. Despite their peculiarities the proposed method is applicable also to DC systems, limited for the moment to the use of spectrum data.

RS current distortion characteristics, or signature, will depend on the type of onboard power converter, its switching frequency, topology, rated power, operating points, and any additional auxiliary converters. This means that one type of load can present a variety of emissions due to the operation and interactions with the power supply grid and that each RS will have a signature dependent on all such parameters, with an increasing complexity considering several tracks and trains at same time.

A methodology for disaggregation of these moving loads would be helpful in many situations, such as multiple trains on the same track, multiple trains on different tracks, but connected to the same traction power substation (TPS), etc. The location of a monitoring system can be at the TPS, on each track feeder, at some victim circuit nearby or onboard RS. Measurement-based analysis that explores data analytics, artificial intelligence, and pattern identification can help overcome the complexity of disturbance management in ERS, providing alternatives to standardized PQ methods that struggle with the observed variability [46].

3. Proposed methodology

The proposed methodology applies autoencoder-based novelty detection to tackle load classification challenges. Reconstruction-based novelty detection is widely applied, as it effectively identifies deviations from expected behavior in various applications [41]. These methods involve modeling only the data of interest for reconstruction (normal condition). When exposed to test data (different from the training data), the RE, calculated as the distance between the input and the reconstructed output, acts as a novelty score [41]. A higher RE typically indicates the presence of anomalous or novel data.

In this work, DAE networks are individually trained for each class to learn and reconstruct the input data's main features. By using RE indices

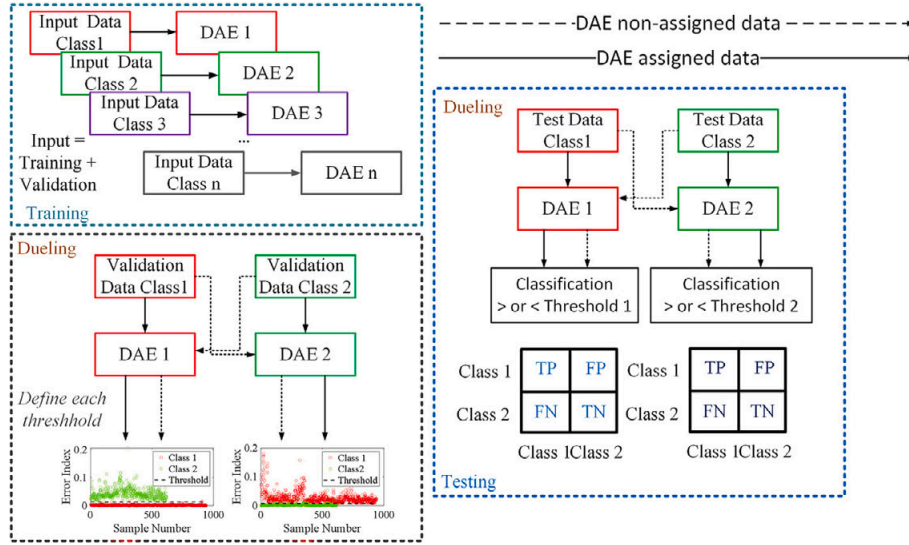


Fig. 2. Proposed framework for load waveform distortion signature classification.

and defining thresholds, the classification is accomplished when a DAE (assigned to a specific class) tries to predict data from other classes. A poor performance index indicates that the input is not to be assigned to that class, and the contrary is also true. In the case of multi-class problems, this process should be cascaded, and each binary classification should be carried out between the assigned class and all the other classes (a validation case is reported later in Section 5.3). Fig. 2 shows the proposed methodology schemes, and the steps are described below.

1. Split the input data into training (60 %), validation (20 %), and test (20 %) sets.
2. Train one DAE network for each class separately with a corresponding RE metric. Training and validation data are used in this step.
3. Validation data are used again to determine the threshold for each DAE by trading off the resulting values of recall and precision metrics. Taking two classes “1” and “2” (as in Fig. 2), carrying out the classification based on the metric being below or above the threshold, the threshold is adjusted to maximize recall and precision at same time at once. The threshold is defined as percentage of the maximum RE of the DAE assigned class: metrics used in this work are mean squared error (MSE) and cosine similarity (CS). This approach is the well-established principle for decisions on the threshold in binary classifications, where a trade-off between precision and recall performance will rule. According to [47], lowering the threshold generally increases recall and penalizes precision, and the contrary is true. Often, the threshold choice will depend on the cost of having false positives or negatives in a specific application [47]. In this work, optimizing the threshold using validation data ensures an effective balance between sensitivity and specificity associated with each DAE’s performance.
4. Also in the case of the test phase, dueling between the DAE networks takes place using the remaining test data aiming at verifying the classification performance with data not participating in the training process. The results are expressed by means of balanced accuracy.

3.1. Input data

The input data consists of measured pantograph current and voltage. The two input types are current harmonic spectrum up to 15 kHz (covering the harmonic range and part of the supraharmmonic range) and 2-D VI

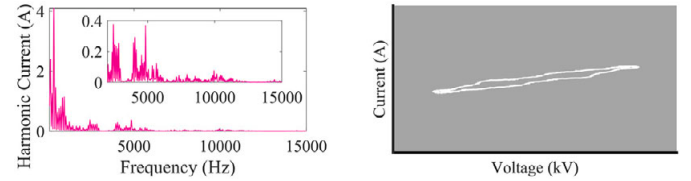


Fig. 3. Example of data input for spectrum (left) and VI diagram (right).

diagram representations. The former was obtained using a fast Fourier transform (FFT) of measured 5-cycle current waveforms. The latter consists of binary image representation of VI diagrams (240×300 pixels), which are the plots of current against voltage waveforms again with a duration of 5 cycles. The image is then flattened to 1-D by reshaping from 240×300 to 72,000 pixels. Fig. 3 shows an example of the input data.

The input data for each network are also normalized according to (1), through the process of z-normalization. The z-normalization guarantees an equal contribution of each data vector $\mathbf{u} \in U_n$, and the normalized matrix is expressed as $\bar{\mathbf{u}}_n$.

$$\bar{\mathbf{u}} = \frac{\mathbf{u} - \mu}{\sigma} \quad (1)$$

where $\bar{\mathbf{u}} \in \bar{U}_n$ is each vector normalized from each original data vector $\mathbf{u} \in U_n$; μ and σ are, respectively, the sample mean value and standard deviation of each original data vector $\mathbf{u} \in U_n$.

Both input types are representative of the distortion signature from RS power converters, as current harmonic spectrum and time-domain admittance. The original data sampled at 50 kS/s pertain to four RS (two for a 16.7 Hz 15 kV ERS and two for a 50 Hz 2×25 kV ERS) and are publicly available [48]. Each short record is tagged for relevant operating conditions and speed. Associated noise and measurement uncertainties are already implicit in the collected data used in this work since the pantograph quantities are actual measured data that refer to real-world problems. In [48], a complete description of the datasets for railway cases is provided, and subsequent references have quantified uncertainty and noise. Uncertainty is on the order of 1 %, and noise is both coherent (e.g., waveform distortion emission from other trains propagating along the line) and incoherent noise (thermal noise in sensors and electronics, pickup by the catenary of extraneous signals, such as radio

Table 1
Parameters and details about the different input data.

Input Class	Rolling Stock	System	Sample Vector Size (n)	Number of Samples (N)
Swiss	Re460 loco 5.6 MW	16.7 Hz 15 kV	Spectra 899 VI diag. 72000	Spectra 4698 VI diag. 4429
German	ICE EMU 9.6 MW	16.7 Hz 15 kV	Spectra 899 VI diag. 72000	Spectra 3083 VI diag. 3083
French	TGV Daisy EMU 9.8 MW	50 Hz 2×25 kV	Spectra 299 VI diag. 72000	Spectra 4217 VI diag. 4141
Italian	ETR 500 EMU 8.8 MW	50 Hz 2×25 kV	Spectra 299 VI diag. 72000	Spectra 3060 VI diag. 3054

transmissions, etc.). Background noise was not quantified at the time of the tests, as the catenary is always supplied at 15 or 25 kV, and the rolling stock was measured in operation. In contrast, the noise of sensors and electronics was quantified as a contribution to the uncertainty, and this evaluation is available in [49].

Input data characteristics are synthesized in Table 1. Only operating states of acceleration and braking are considered; some data with null

values were excluded during the preliminary screening. The dataset is slightly unbalanced with two classes having more than 4000 records (Swiss and French systems) and two with about 3000 records (German and Italian systems). The variety of data for each class depends on the mix of operating conditions.

Looking at the input data in more detail, one can visualize the emission characteristics of those different rolling stocks, as well as the complexities and issues associated with the railway aspects, such as moving loads, highly fluctuating demand, and the impact of operation on waveform distortion parameters, as discussed in Section 2. Fig. 4 shows a time-varying representation of the rolling stock emissions using spectrograms, highlighting that each train can have signature variability that will impose challenges to load detection techniques. Additionally the total distortion, by means of root-square of sum of squares of synchronous current components up to 15 kHz, distributions for each rolling stock used in this work is shown in Fig. 5. The variety of distortion signatures is related to the following aspects regarding railway features.

- It is expected that the rolling stock operating in the railway section generates an irregular and complex current flow, with high distortion and power fluctuations, including increased variance driven by vehicle operating modes, path, and driving patterns.
- Low-order and high order harmonics have different mechanisms for variation, while the first ones are probable with wider structured distribution, the second ones exhibit a narrower amplitude range but with randomness in behavior [50].

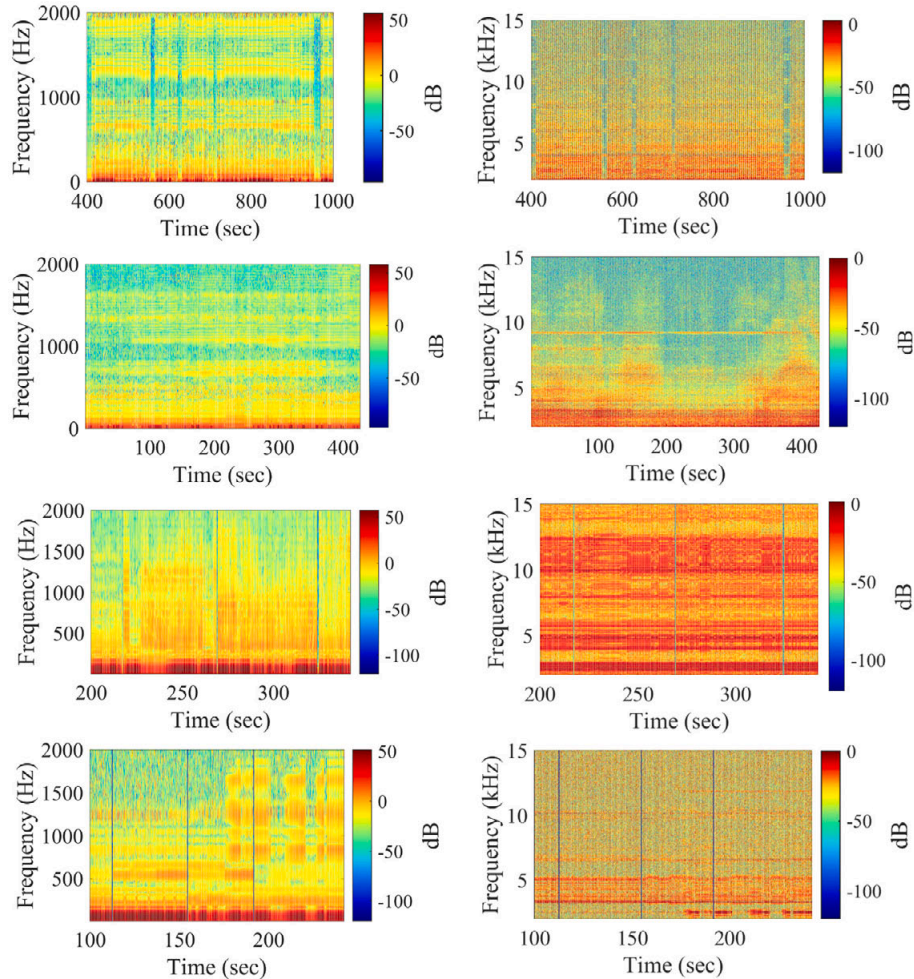


Fig. 4. Spectrograms characterizing the waveform distortion emission (current) of each rolling stock in the harmonic range (left) and supraharmonic range (right): Swiss (first row); German (second row); French (third row); Italian (fourth row).

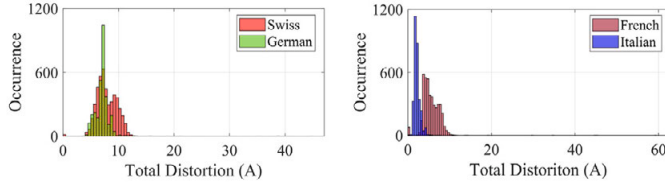


Fig. 5. Emission total distortion histograms considering the whole measurement set for both pairs of rolling stocks investigated in this work.

- The rolling stock moves along the system, altering factors such as background voltage, system impedance, and the interactions between distortion sources and the network. Since the total emission is not only dependent on the primary emission (base current injected by the vehicle), but also on the secondary emissions (current component portion due to impedance interactions between source and network [51]).

3.2. Deep autoencoder

The DAE is an unsupervised learning algorithm that aims to perform dimensionality reduction with the encoder and reconstruction of the signal with the decoder keeping the objective of minimizing the RE during training. In other words, the algorithm reconstructs the signal, aiming for an output “equal” to the input. During this training process, the model learns the main features of the signal, and the level of lower dimension, between encoder and decoder, contains the codification, or map, of features that describe those characteristics [52,53].

The encoder consists of several CNN layers as linear filters, parameterized by a weight matrix and a bias. At the same time, the decoder is a reverse linear filter (aiming at reconstruction) and the activation functions are non-linear, mapping the coded features into the range (0,1) [52]. The encoding step can be described by (2),

$$z = f(\{W, b\}; \mathbf{u}) = \sigma(W\mathbf{u} + b) \quad (2)$$

where the input feature vector $\mathbf{u}$ is represented by the coded feature vector $\mathbf{z}$, which propagates through the hidden layers. W represents the weight matrix, b the biases and σ is the activation function, which in the following is implemented using a Rectified Linear Unit (ReLU).

The decoding step is described by (3), where $\mathbf{z}$ is mapped back to the higher-dimensional reconstruction $\mathbf{y}$:

$$\mathbf{y} = f(\{W', b'\}; \mathbf{z}) = \delta(W'\mathbf{z} + b') \quad (3)$$

Biases and weights are updated and optimized to bring $\mathbf{y}$ as close to the input $\mathbf{u}$ as possible. In this work, the optimization process is performed using the mini-batch Adam optimization algorithm [54].

Error metrics are used to compare input and output during training, usually MSE, as in this work when working with spectra. For VI diagrams as input, however, the CS is the best option (4). CS assesses the similarity between two vectors by calculating the cosine of the angle between them: similar vectors will result in values nearly equal to unity, while dissimilar ones will be closer to 0 [55]. Binary representation of VI diagrams is a sparse, structured image that often resembles edge-like patterns. Such images and data contain information about shape and directionality, not transitions between tones, brightness, or continuous coloring. This means that the data pattern is more important than absolute pixel intensities, making metrics like MSE (which penalize absolute pixel-wise differences) fall short. For this reason, metrics like CS are recommended, as they compare orientation and pattern in highly-dimensional spaces, preserve structural features, and lead to better generalization [56].

Table 2

DAE network parameters and training for RS data.

Parameter	Spectrum input	VI Diagram input
Input	299 (50 Hz)/899 (16.7 Hz)	72,000
Encoder (neurons)	$64 \times 32 \times 16$	$512 \times 256 \times 128 \times 80$
Decoder (neurons)	32×64	$128 \times 256 \times 512$
Output	299 (50 Hz)/899 (16.7 Hz)	72,000
Neuron activation	ReLU	ReLU
Output activation	Linear	Sigmoid
Loss function	MSE	CS
Epochs	750	150
Learning rate	0.001	0.001

$$CS = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (4)$$

where A_i and B_i are the i -th components of vectors $\mathbf{A}$ and $\mathbf{B}$, respectively.

Table 2 details the parameters and architecture used for each type of input. Considering the studied dataset, the architecture and hyper-parameters were empirically chosen to minimize training losses and can be a good starting point for other similar datasets.

3.3. Classification metrics

Classification metrics are used for two steps of the methodology, defining the threshold, and testing the performance of the classifiers. The trade-off between recall and precision is used to define the threshold. The first one measures the rate of correct predictions compared to all the actual positive examples in a dataset, and the second measures the accuracy in correctly predicting positive examples within a specific dataset [57]. Both metrics are described by (5),

$$\text{Recall} = \frac{TP}{TP + FN} \quad \text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

where TP are the true positive predictions, FN are the false negative predictions, and FP are the false positive predictions.

The balanced accuracy (BA) index is used for evaluating the performance of the DAE networks against the test dataset (6). It can be described as the rate of correct predictions to the total samples of the dataset [57], applicable particularly to unbalanced datasets.

$$\text{BA} (\%) = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \times 100 \quad (6)$$

The immediacy of the classification quality is illustrated using confusion matrices, where each cell counts the number of correct or incorrect predictions per class.

4. Methodology applied to the LIT dataset

The LIT dataset is a NILM benchmark that is publicly available, and was developed by the Laboratory of Innovation and Technology in Embedded Systems [58]. The synthetic subset (LIT-SYN) was collected in the laboratory, where actual loads were connected in a controlled environment with known on and off times. This dataset contains voltage and current waveforms sampled at 15.360 Hz with a mains fundamental frequency of 60 Hz. The appliances are divided into 26 classes, distributed among microwaves, lamps, other lighting devices, monitors, cell-phone chargers, fans, resistors, drills, heaters, and hair dryers.

For this work, only the steady state waveforms were used in the training and classification processes, with the objective of identifying the waveform distortion signature of each class and confirming the results obtained on the RS dataset; for this the dataset with a single load switched on at a time was used, and 6-cycle waveforms were extracted resulting in a total of 58,241 waveforms. Each class has 1344 data samples for training, 448 for validation and 488 for testing. The classes correspond to the electrical appliances that are shown in Table 3.

Table 3
List of the electrical appliances in the LIT dataset.

Class	Load	Brand	Power
1	Microwave Oven Standby	Consul	4.5 W
2	Led Lamp	Taschibra	6 W
3	CRT Monitor	Sony	10 W
4	Led Pannel	Citiaqua	13 W
5	Fume Extractor	Toyo	23 W
6	Led Monitor	AOC	26 W
7	Asus Phone Charger	Asus	38 W
8	Soldering Station	Weller	40 W
9	Motorola Phone Charger	Motorola	50 W
10	Lenovo Laptop	Lenovo	70 W
11	Fan	Mondial	80 W
12	Resistor	Ohmtec	80 W
13	Vaio Laptop	Sony	90 W
14	Incandescent Lamp	Osram	100 W
15	Drill Speed 1	Bosch	165 W
16	Drill Speed 2	Bosch	350 W
17	Oil Heater Power 1	Pelonis	520 W
18	Oil Heater Power 2	Pelonis	750 W
19	Microwave Oven On	Consul	950 W
20	Nilko Air Heater	Nilko	1120 W
21	Hair Dryer Eleganza - Fan 1	GAMA Italy	365 W
22	Hair Dryer Eleganza - Fan 2	GAMA Italy	500 W
23	Hair Dryer Super 4.0 - Heater 1	Super 4.0	600 W
24	Hair Dryer Super 4.0 - Heater 2	Super 4.0	885 W
25	Hair Dryer Parlux - Fan 1	Parlux	660 W
26	Hair Dryer Parlux - Fan 2	Parlux	1120 W

Table 4
DAE network parameters and training for RS data.

Parameter	Spectrum input	VI Diagram input
Input	127 (60 Hz)	19,200
Encoder (neurons)	$64 \times 32 \times 16$	$128 \times 64 \times 32$
Decoder (neurons)	32×64	64×128
Output	127 (60 Hz)	19,200
Neuron activation	ReLU	ReLU
Output activation	Linear	Sigmoid
Loss function	MSE	CS
Epochs	750	50
Learning rate	0.001	0.001

Harmonic spectra up to 7620 Hz and VI diagrams were extracted; for the latter the binary images were scaled to 120 x 160 to reduce the training time. Table 4 shows the DAE network parameters of the LIT-SYN dataset.

5. Results

The methodology was applied to the described input data to compare performance considering RS from the same type of railway electrification. The results show separate classification performances for Swiss and German RS, as well as for French and Italian RS.

5.1. Swiss vs. German rolling stock

5.1.1. Swiss vs. German case: spectrum

Looking into the results from Swiss locomotive vs. German locomotive for spectrum input, the defined thresholds are 90 % (0.01407) and 15 % (0.00806) of the maximum RE for Swiss and German DAE networks, respectively. Fig. 6 shows the recall and precision trade-off plots for this case. Fig. 7 shows the validation data separation according to the spectrum input's threshold and RE.

The thresholds can classify without excessive data separation, where the DAE network assigned to one class performs worse for other classes. The pantograph quantities and operational behavior from two different RS can be similar under the same strict standardization for parameters such as power factors, maximum current, and voltage amplitudes, but the spectral signatures spot the different features related to differences

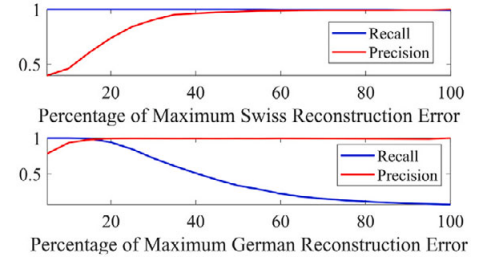


Fig. 6. Trade-off between recall and precision of Swiss (top) and German (bottom) DAE networks for spectrum input case.

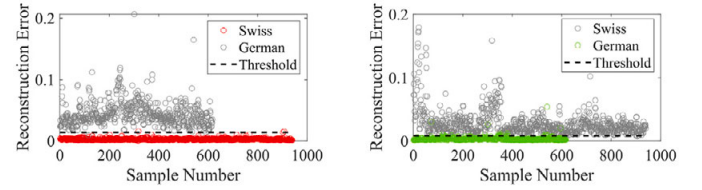


Fig. 7. RE of the validation dataset considering threshold separation of Swiss (left) and German (right) DAE networks for spectrum input case.

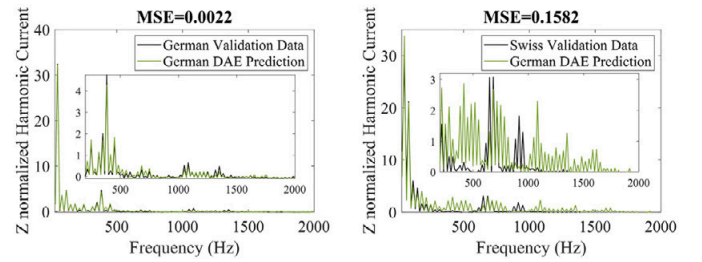


Fig. 8. Predictions from German DAE network for an example of German data input (left), an example of Swiss data input with RE far from the threshold (right).

True Class	D	615	2
	CH	5	935
		D	CH
		Predicted Class	

True Class	D	605	12
	CH	10	876
		D	CH
		Predicted Class	

Fig. 9. Confusion matrices for spectrum input performance of the Swiss (left) and German (right) classifiers.

in switching frequencies, interleaving, modulation, etc. in the onboard converters.

Fig. 8 shows the comparison between the prediction of the German DAE network for validation data from either class. It is possible to observe a high-fidelity reconstruction with its own data, whereas using Swiss data causes the appearance of “ghost frequencies” and discrepancies which are the reason for the effective separation of the two.

Performance evaluation is carried out with the spectrum data showing good performance in terms of accuracy for both DAE networks. The BA for the Swiss classifier was 99.57 %, with recall and precision of 0.9979 and 0.9947, respectively. For the German classifier BA was 98.17 %, with recall and precision of 0.9926 and 0.9811, respectively. Fig. 9 shows the results using confusion matrices.

5.1.2. Swiss vs. German case: VI diagrams

Looking into the results for Swiss vs. German VI diagrams, we see suitability of the methodology for classification, with thresholds of 50 %

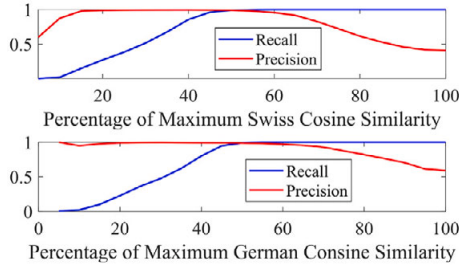


Fig. 10. Trade-off between recall and precision of Swiss (top) and German (bottom) DAE networks for the VI input case.

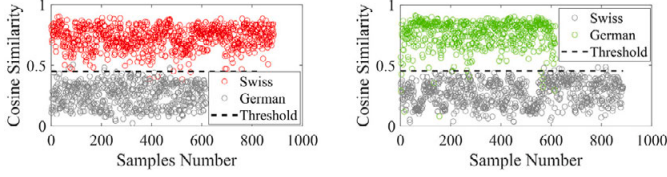


Fig. 11. CS of validation dataset considering threshold separation of Swiss (left) and German (right) DAE networks for the VI input case.

True Class	D	615	2
	CH	7	879
	Predicted Class	D	CH

True Class	D	605	12
	CH	10	876
	Predicted Class	D	CH

Fig. 12. Confusion matrices for VI input performance of the Swiss (left) and German (right) classifiers.

(0.4510) and 55 % (0.4552) of maximum CS for Swiss and German DAE networks, respectively. Fig. 10 shows the recall and precision trade-off plots for the VI diagrams with 16.7Hz fundamental. Fig. 11 shows the validation data separation according to the threshold and RE for the VI diagram input, and one can observe the precise separation based on CS.

The classification performance verified with the VI diagrams test data is also very good. The BA for the Swiss classifier is 99.44 %, with recall and precision of 0.9977 and 0.9921, respectively. For the German classifier BA is 98.46 %, with recall and precision of 0.9926 and 0.9811, respectively. Fig. 12 shows the results by means of confusion matrices.

The classification method showed high accuracy and suitability for load monitoring. It is also important to highlight that the two RSs, Re460 and ICE EMU, have power ratings that are significantly different, as well as constructive characteristics, converter arrangements, and topologies. Those factors also impact the waveform distortion signatures and the ease of separation. The French and Italian RS are much more similar and we expect a somewhat lower performance.

5.2. French vs. Italian rolling stock

The methodology is also applied to RS of a 50 Hz ERS. TGV Dasye EMU from France and ETR500 EMU from Italy were assessed using waveform distortion parameters. They are electric multiple unit (EMU) types with similar power ratings and similar architectures, representing a significant challenge. We will see that, even with lower performance compared to the previous case, the methodology has provided a suitable classification for such loads.

5.2.1. French vs. Italian case: spectrum

For spectrum as input, the defined thresholds are 10 % (0.0656) and 55 % (0.0665) of maximum RE for the French and Italian DAE networks, respectively. Fig. 13 shows the recall and precision trade-off plots for

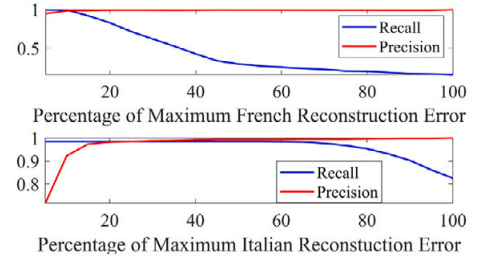


Fig. 13. Trade-off between recall and precision of French (top) and Italian (bottom) DAE networks for the spectrum input case.

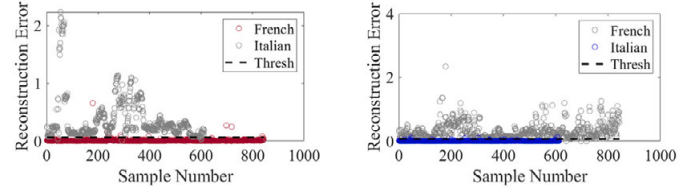


Fig. 14. RE of the validation dataset considering threshold separation of French (left) and Italian (right) DAE networks for the spectrum input case.

True Class	F	833	10
	I	5	607
	Predicted Class	F	I

True Class	F	785	43
	I	31	580
	Predicted Class	F	I

Fig. 15. Confusion matrices for spectrum input performance of the French (left) and Italian (right) classifiers.

this case. Fig. 14 shows the validation data separation according to the threshold and RE.

The classification test results show that the load identification is performed accurately for the French vs. Italian case. The BA for the French DAE network classifier was 99.00 %, with recall and precision of 0.9918 and 0.9838, respectively. In comparison, the Italian classifier BA was 98.52 %, with recall and precision of 0.9790 and 0.9822, respectively. Fig. 15 shows the results for the spectrum-based classification using confusion matrices.

5.2.2. French vs. Italian case: VI diagrams

Considering the VI diagrams as input, the overall performance was worse but still accurate. The thresholds were defined for each network: 50 % (0.4819) of maximum French RE based on CS for the French DAE network and 55 % (0.5425) of maximum Italian CS for the Italian DAE network. Fig. 16 shows the trade-off comparisons for both networks in this case.

Fig. 17 shows the validation data separation using the VI diagrams for the French vs. the Italian case. The French DAE network performed worst in validation for VI diagrams, with the recall being 0.9378 and the precision 0.9037.

Fig. 18 shows the results of the test classification using confusion matrix for the French vs. Italian case. The French DAE network classifier has a test BA of 94.87 %, with a recall of 0.9493 and precision of 0.9310, for the VI diagram case. At the same time, the Italian DAE network classifier has a test BA of 98.31 %, with a recall of 0.9788 and a precision of 0.9820.

5.3. Performance with LIT dataset and comparisons

While the threshold of each DAE is determined for the RS binary problems by trading off between recall and precision on the two classes

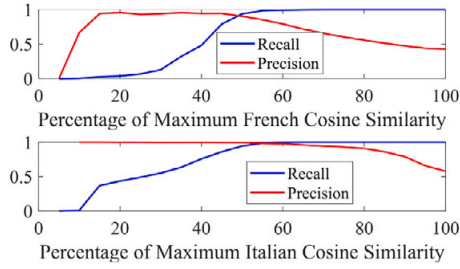


Fig. 16. Trade-off between recall and precision of French (top) and Italian (bottom) DAE networks for VI input case.

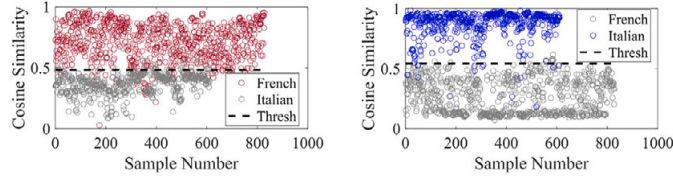


Fig. 17. CS of validation dataset considering threshold separation of French (left) and Italian (right) DAE networks for the VI input case.

True Class	French		True Class	Italian	
	F	I		F	I
Predicted Class	825	18	Predicted Class	815	13
	5	607		11	600

Fig. 18. Confusion matrices for VI input performance of the French (left) and Italian (right) classifiers.

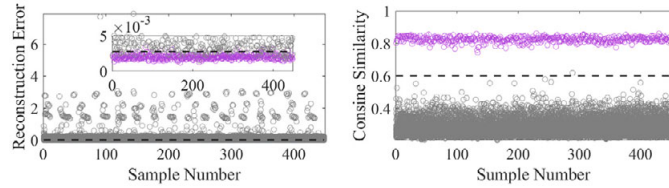


Fig. 19. RE (spectrum) and CS (VI diagram) validation of class 1 (black) against all other classes (gray) for the class 1 DAE. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of validation data, for the LIT-SYM dataset each trade-off is between the assigned class and all others (maintaining a binary classification perspective). Fig. 19 shows the RE for the validation of class 1 against all others. The methodology is effective and can be extensively applied to datasets with an arbitrary number of classes.

Fig. 20 shows the accuracy for each class-assigned classification based on DAE RE. It is notable that, with exception of classes 3 and 15, the spectral input performed better than the VI diagram inputs.

The similarity between several classes within the LIT-SYN dataset (check Table 3) can adversely affect and reduce performance of techniques based on novelty detection and for this reason the LIT case represents a challenging benchmark: this is evident when using VI diagrams. In particular, for loads with dominant resistive behavior, the active current constitutes the major portion of the load current due to the high power factor, which causes the VI diagrams of such appliances to be very close in shape [19].

For a fair benchmarking of the proposed novelty-based method, a comparison with the classification performance of previous literature contributions on the same LIT dataset was carried out, as shown in Table 5.

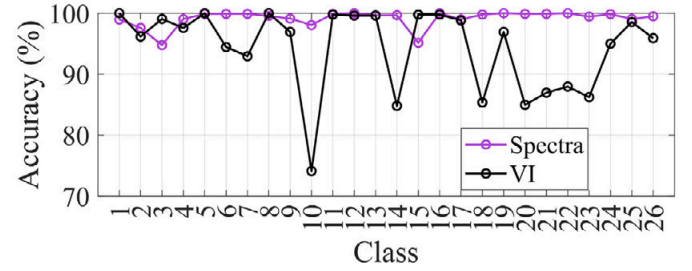


Fig. 20. Accuracy for each assigned DAE for each class (class “i” vs. all other classes).

Table 5
Accuracy comparison with literature.

Reference	Dataset	Method	Accuracy
[59]	LIT-SYN	Multiagent	96.00 %
[27]	LIT-SYN	Generative Adversarial Network CNN	95.00 %
[28]	LIT-SYN	Wavelet Scattering CNN	99.90 %
[20]	LIT-SYN	2-D Fourier Series Classifiers	96.80 %
This work	LIT-SYN	DAE-RE Spectra	99.13 %
This work	LIT-SYN	DAE-RE VI	94.28 %

6. Discussion

6.1. Overall discussion, advantages, and limitations

The results confirm the contributions of the proposed methodology. All BA values are higher than 98 %, except for the French DAE network classifier with VI diagram input, which is nevertheless above 90 %.

The high accuracy of the method is quite important since the input data has high-dimensional characteristics; while traditional classification techniques are dependent on signal processing or machine learning to extract key features, the DAE deep learning approach allows automatic feature extraction saving one onerous step.

An overall architecture for each type of input can present a challenge since the performance of the DAE network will also depend on the variety of observations available for each class. However, the study cases showed that the proposed arrangements of the models fit the application to both types of data. Other methodologies of classification based on machine learning train models with all classes together, while in the proposed method there is one DAE network for each class. This approach is also advantageous with unbalanced data: because of the independent training and prediction of the networks, the imbalance between classes does not impact the performance of the final classifier. Furthermore, the strategy does not require new training using all class data; if one new class should be added, only one new DAE network needs to be included with suitable definition of the thresholds between the new data and the DAE networks for all previous data, which is less computationally intensive.

In an ERS distortion perspective, the methodology performs the load identification using signatures of the distortion caused by the onboard power electronics, such as spectral components for switching frequency, low-order harmonics, etc. Considering VI diagrams, other parameters such as power factor, interharmonic products, maximum operating current, and time-domain admittance variation are included and assessed as features. The method shows potential for application in railway systems, providing an alternative for RS load monitoring.

Similar load distortion signatures surely hinder novelty-based detection methods. Loads with quite similar distortion features will make it harder to converge on a trade-off between recall and precision, leading to lower accuracy. In general, the presence of static power converters

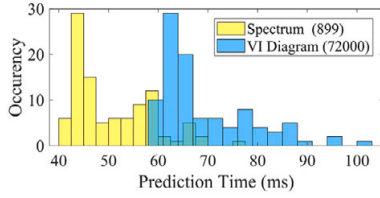


Fig. 21. Prediction time for spectrum and VI diagram input, considering the largest dimension used in this work (899 for the spectrum and 72,000 for the VI diagram).

Table 6

Comparison of computational complexity between NILM models (adapted from Ref. [25]).

Model	FLOPs (M)
AlexNet [26]	710.16
VGG16 [32]	15,466.26
DenseNet121-Concatenate-CNN [23]	2689.45
Xception-Concatenate-CNN [23]	4258.82
Generative Graph method [34]	1.61
Denoising autoencoder-adaptative neural tree [25]	7.88
Denoising autoencoder-adaptative neural tree (lightweight) [25]	1.86
This work - Spectrum Input (50 Hz)	0.09
This work - Spectrum Input (16.7 Hz)	0.24
This work - VI Diagram input (72,000)	148.15
This work - Spectrum Input (60 Hz)	0.04
This work - VI Diagram input (19,200)	9.87

allows some diversity in the distortion behavior, even with similar architectures and parameters. Looking at the LIT dataset accuracy results, loads with almost unity power factor have unavoidably very similar VI diagrams reducing classification performance. Classification based on spectrum, conversely, shows better performance and should in general be preferable. As stated, this problem is not likely to occur for non-linear ERS loads due to distinct and distinguishable characteristics of the onboard power conversion system.

Fig. 21 reports the prediction times considering the largest spectrum and VI diagram input data. The histograms were built with hundreds of DAE prediction runs: although some variability is unavoidable (in particular due to concomitant running programs and operating system activity), it is sensible to say that VI diagram calculations are 50 % slower than spectrum ones. Calculations were performed on Python 3.11, and the hardware is as follows: x64-based PC, 11th Gen Intel(R) Core (TM) i7-1165G7 @ 280 GHz, 2803 MHz, 4 Core(s). In addition to those quantified inference times, floating-point operations (FLOPa), which measure the computational complexity of the models, can be informed and compared to remarkable performance in the literature. Table 6 summarizes the comparisons regarding the computational complexity of the models in the literature compared with the proposed one, highlighting the potential of the model for implementation with suitable computational cost.

6.2. Explanations and visualizations of good classifications and misclassifications

To explore explanations for correct classifications and misclassifications, especially with attention to FP and FN instances of the binary classification in the rolling stock cases, one can examine the distribution of operating and distortion conditions regarding the load identification results. Additionally, this approach provides a good understanding of the variety of operating scenarios involving both vehicle state and the signature of waveform distortion and how the proposed methodology can capture and deliver suitable performance for a significant portion of those conditions. To exemplify and explain those aspects, the French versus Italian rolling stock spectrum case is explored.

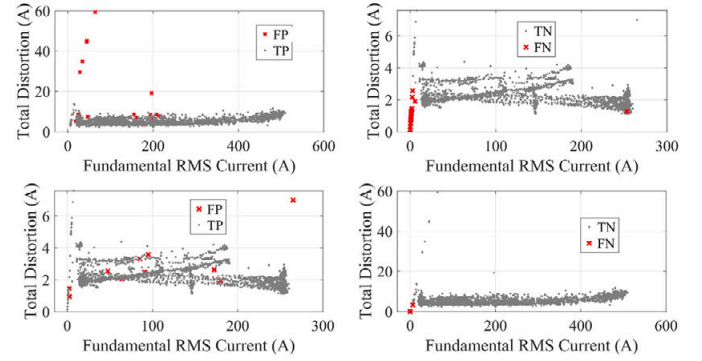


Fig. 22. Scatterplots between fundamental RMS current and total distortion, with get TP, FP, TN, and FN predictions of both French DAE (top) and Italian DAE (bottom).

Fig. 22 illustrates a scatterplot between fundamental RMS current and total distortion, with TP, FP, TN, and FN predictions of both French DAE and Italian DAE (considering the whole measured data for each class). From these illustrations, it is possible to observe the success of the methodology due to the coverage of correct load identification across a variety of scenarios. FP and FN in novelty detection applications are often associated with borderline reconstruction performance, where predicted instances are close to the threshold, which can impose a degree of confusion similar to that of any other classification method. As more distinguished load signatures are, fewer borderline predictions will occur, addressing the problem discussed in the previous subsection. Additionally, examining the scatter plots reveals abnormal conditions associated with high distortion and low to mid current flow in the pantograph, which also explains FP and can be linked to transient conditions. FN, on the other hand, is highly associated with very low levels of current concentrated in that region for both French and Italian DAEs. This phenomenon can be explained by the neutral sections in 50 Hz railway supply systems due to phase separation. A neutral section is deployed in electrified railways to prevent phase-to-phase short circuits when a train transitions between railway sections supplied by alternating phases from the public high-voltage network. Such transitory conditions impose an additional challenges to NILM methodologies.

Additionally, Fig. 23 illustrates explanations for the reduced accuracy of French DAE, which are related to voltage supply conditions causing anomalies in the data. These conditions are associated with an increase in high-frequency content in the voltage supply. In Fig. 23, a summation of VI diagrams for different CS scores of the test dataset. Such conditions lacked sufficient examples to develop strong learned features compared to normal conditions. Those samples, when mixed under normal conditions, also reduce the overall performance of the DAE in making threshold decisions with validation data. These aspects also highlight the challenges of applying NILM in an electromagnetic environment like railway systems.

Such approaches can support users and technical reports that utilize this work methodology in providing explanations for classification decisions in future implementations. Future work should also explore anomaly detection to remove abnormal conditions that do not characterize the waveform distortion load signature of the load under normal conditions.

6.3. VI diagram input flattening and comparison with traditional unsupervised machine learning

The impact of flattening 2-D input image data into a 1-D vector representation on pixel adjacency and spatial structure is well understood within the research community, consequently affecting the learning process and feature extraction of machine learning algorithms. Flattening,

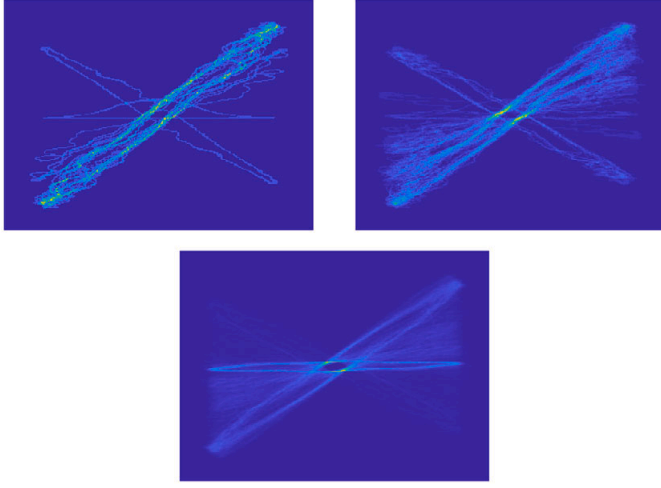


Fig. 23. Summation of VI diagram images of the French DAE test dataset dependent on the CS performance: below 0.4 (top left), between 0.4 and 0.5 (top right), and above 0.5 (below).

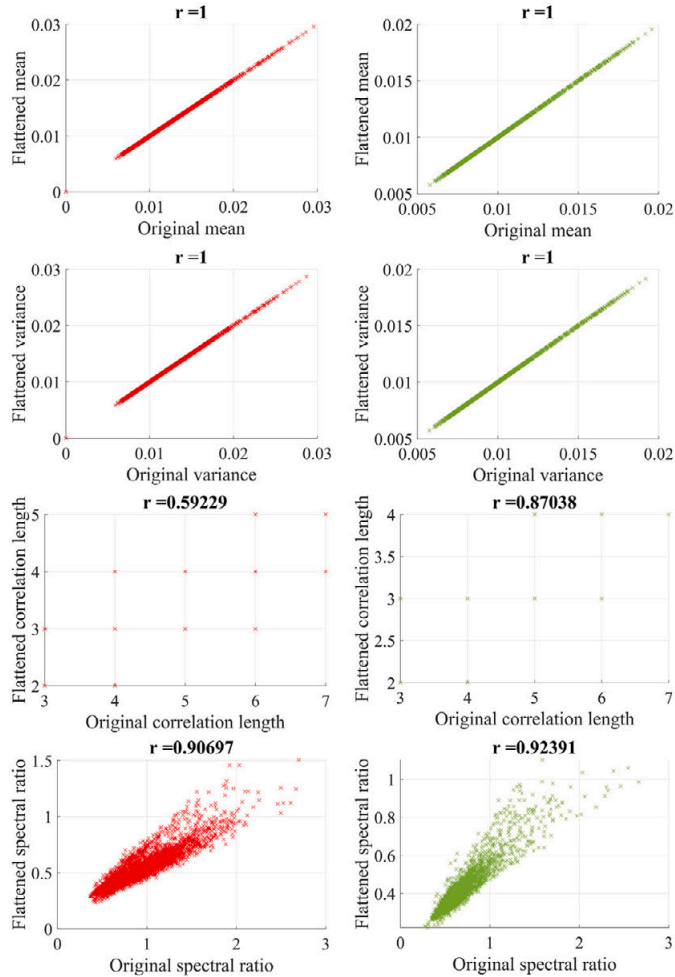


Fig. 24. Statistical and structural metrics for quantifying the impact of flattening for Swiss (left) and German (right) inputs.

therefore, acts as a pragmatic simplification that enables the use of a fully connected architecture for the DAE on high-dimensional data, as explored in the VI diagram images in this work, without losing

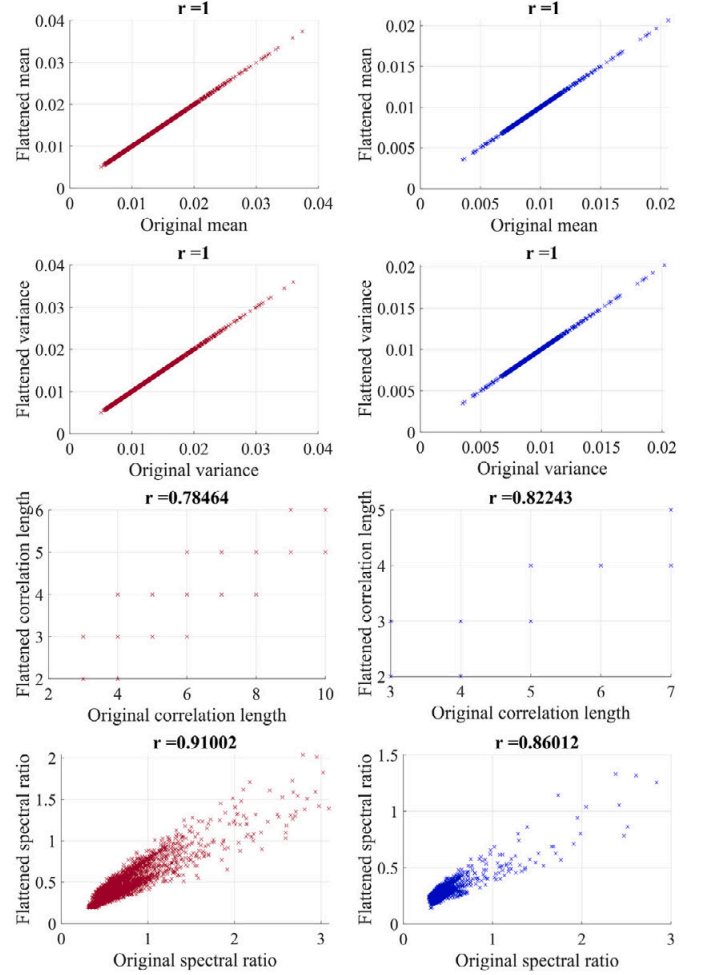


Fig. 25. Statistical and structural metrics for quantifying the impact of flattening for French (left) and Italian (right) inputs.

the waveform distortion signature that is key to the classification approach. Although this argument is valid and previous literature has used flattening combined with autoencoder-based models in a variety of applications [60–66], quantification of the input data for VI diagrams can back up the fact that the impact on the spatial structure of the images does not cause a significant loss of information critical to the novelty detection task. For this application, the DAE only needs to capture global patterns to generate meaningful reconstruction errors. Local adjacency information lost in flattening is not required because novelty detection principles explored using reconstruction error detect or classify via deviations in global statistical structure rather than fine spatial textures.

Figs. 24 and 25 illustrate the quantification, by means of scatter plots and Pearson correlation, of metrics that demonstrate the change (from original and flattened images) in first-order statistics, mean and variance, and higher-order structural properties such as correlation length and spectral ratios differ. Although there is no established standard to evaluate the impact of flattening high-dimensional images for fully connected DAE, those interpretable metrics that show maintenance of the overall distribution of the data and grounded classical image statistics [67,68] can quantify how flattening affects global spatial structure, providing empirical evidence that our approach preserves key statistical properties necessary for novelty detection approach on classification of load waveform distortion signatures for AC railways and is additionally validated on other benchmark loads. Moderate to high correlation is observed for structural properties of the image between both formats.

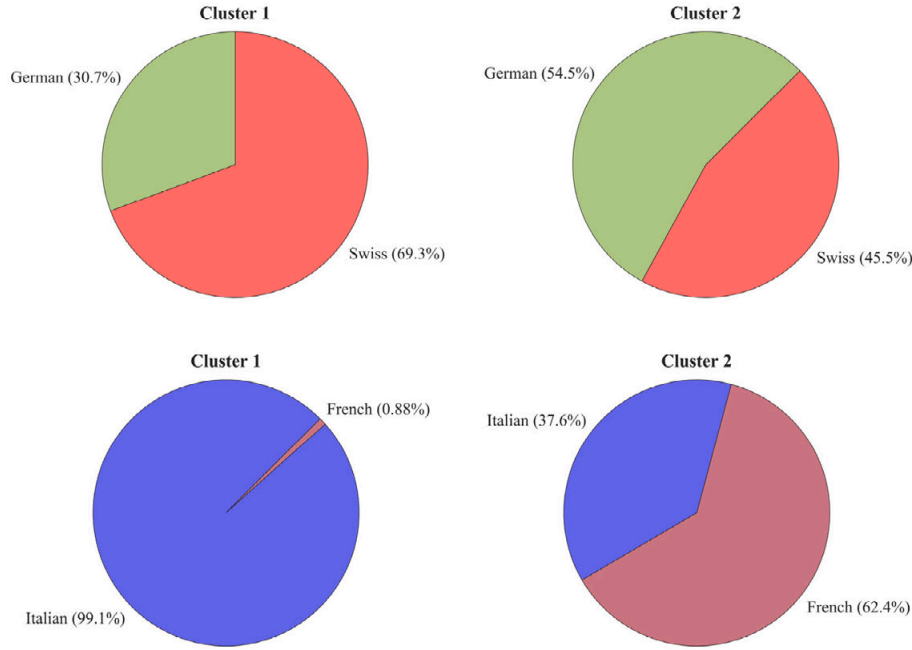


Fig. 26. Cluster separation utilizing k-means aiming binary classification between rolling stocks for the VI diagram cases: Swiss vs. German (top), and French vs. Italian (bottom).

Additionally, the justification for using autoencoder-based reconstruction error methodology for the waveform distortion signature classification over other traditional unsupervised machine learning techniques is related to the high-dimensional characteristics of the explored data. Traditional unsupervised methods, such as k-means clustering, cannot perform well, even with flattened input from VI, due to the size of each input observation, which is 72,000 for rolling stock classification cases. The so-called “curse of dimensionality” describes the phenomenon in which distances in high-dimensional spaces become less informative, and k-means typically yields poor or unstable clusters without first performing dimension reduction [69]. Such a previous step is often recommended, utilizing methods such as PCA or autoencoder-based methods, which allow for the assessment of a compact version of the input data that represents the main features associated with the observations. However, the objective of this paper is to explore high-dimensional data classification combining a novelty detection approach with DAE reconstruction error. Another aspect that can be mentioned regarding the k-means is the limitation of yielding stable classifications when each class can contain a variety of operating conditions and distortion patterns. Clustering techniques, such as k-means, can definitely identify patterns in the data; however, there is no guarantee that the unsupervised technique will accurately group the separation between two different types of loads, such as rolling stocks, especially when impacted by the high-dimensional characteristics of the data. Moreover, the variety of patterns and conditions (see Fig. 22) also affects the separation of the clusters.

The quantification and assessment of the traditional unsupervised performance on the VI input data can reveal both aspects associated with the possible utilization of k-means clustering for classification. Fig. 26 shows the result of forcing a binary cluster for the separation of the binary classification proposed in this work for the segregation of waveform distortion signatures from rolling stock, considering the VI diagram cases (input data with 72,000 in size). The results show that the clusters do not accurately separate the load signatures based only on the load monitoring or type of load characteristics, but rather show a mixed observation from both loads.

Additionally, the limitation of k-means in forming structured clusters with high-dimensional data is illustrated by the VI diagram cases, which

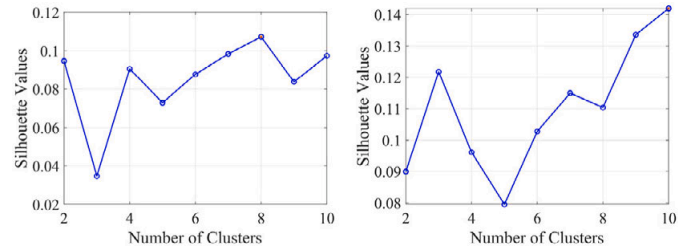


Fig. 27. Silhouette values assessment for both VI diagram cases, highlighting the poor performance of k-means on high-dimensional data: Swiss vs. German (left), and French vs. Italian (right).

utilize the flattened data again. For this purpose, Silhouette values [70] are used to indicate the quality of cluster separation for a given number of clusters, as illustrated in Fig. 27. Such results highlight the Silhouette value close to zero, or values below 0.25, which indicate a substantial lack of structure among the clusters and low quality performance of k-means in handling data with such high dimensions.

7. Conclusions

This work has presented a classification methodology for load monitoring based on RS waveform distortion signatures. The proposed methodology combines the benefits and advantages of unsupervised deep learning and RE performance classification for performing NILM in ERS, adopting a novelty detection approach. The methodology was applied to pantograph voltage and current measurements from four RS items, which were compared in pairs of the same kind of railway electrification. The methodology has shown suitable results with high accuracy (average of 98.81 % for spectrum input and 97.77 % for VI diagram input) and was subsequently validated with a load monitoring multi-class classification problem using LIT-SYN dataset (achieving an average accuracy of 99.13 % for spectrum input and 94.28 % for VI diagram input).

The paper has also discussed the main advantages and disadvantages. High accuracy comparable with state-of-the-art methods, automatic feature extraction capability avoiding additional signal processing and

parameter estimation, and suitable prediction time performance are highlighted as advantages. However, some limitations should also be acknowledged, such as finding one architecture that fits all DAE network needs for multiple class problems and the drawback of novelty-based detection techniques of worsening performance with strongly similar parameters (not the case for RS classification and equally extensible to many other classification methods, being implicit in the classification problem).

From a PQ ERS perspective, the results provide substantial capability for extensive application because the method can identify the main distortion features for each RS. The methodology in fact can achieve load disaggregation in an environment with a variety of disturbance sources, tracking operating conditions down to the sub-second duration. The method is also flexible accepting new classes (so new RS) with a minimal learning phase.

Future work should consider the application to measurements from other connection points in the system (such as the traction substation, which represents the point of common coupling) and the implementation of a monitoring system together with a railway authority (or infrastructure owner).

CRedit authorship contribution statement

Rafael S. Salles: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sarah K. Rönnerberg:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization. **Andrea Mariscotti:** Writing – review & editing, Visualization, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have cited both datasets that are public available.

References

- [1] EN 50388-1, Railway Applications — Fixed Installations and Rolling Stock — Technical Criteria for the Coordination Between Electric Traction Power Supply Systems and Rolling Stock to Achieve Interoperability, Brussels, Belgium, CENELEC, 2022.
- [2] H. Hu, Y. Shao, L. Tang, J. Ma, Z. He, S. Gao, Overview of harmonic and resonance in railway electrification systems, *IEEE Trans. Ind. Appl.* 54 (5) (2018) 5227–5245, <https://doi.org/10.1109/TIA.2018.2813967>
- [3] A. Mariscotti, Experimental characterisation of active and non-active harmonic power flow of AC rolling stock and interaction with the supply network, *IET Electr. Syst. Transp.* 11 (2) (2021) 109–120, <https://doi.org/10.1049/els2.12009>
- [4] A. Mariscotti, Non-intrusive load monitoring applied to AC railways, *Energies* 15 (11) (2022) 4141, <https://doi.org/10.3390/en15114141>
- [5] T. Hassan, F. Javed, N. Arshad, An empirical investigation of V-I trajectory based load signatures for non-intrusive load monitoring, *IEEE Trans. Smart Grid* 5 (2) (2014) 870–878, <https://doi.org/10.1109/TSG.2013.2271282>
- [6] P.A. Schirmer, I. Mporas, Non-intrusive load monitoring: a review, *IEEE Trans. Smart Grid* 14 (1) (2023) 769–784, <https://doi.org/10.1109/TSG.2022.3189598>
- [7] P.G. Papageorgiou, G.C. Christoforidis, A.S. Bouhouras, NilM in high frequency domain: a critical review on recent trends and practical challenges, *Renew. Sustain. Energy Rev.* 213 (2025) 115497, <https://doi.org/10.1016/j.rser.2025.115497>
- [8] L. Du, Y. Yang, D. He, R.G. Harley, T.G. Habetler, Feature extraction for load identification using long-term operating waveforms, *IEEE Trans. Smart Grid* 6 (2) (2015) 819–826, <https://doi.org/10.1109/TSG.2014.2373314>
- [9] J. Yuan, H. Wang, P. Wu, C. Wang, J. Chen, H. Jiao, A novel current signal feature and its application in noninvasive load monitoring, *IEEE Trans. Instrum. Meas.* 70 (2021) 1–10, <https://doi.org/10.1109/TIM.2020.3036651>
- [10] Y. Liu, X. Wang, L. Zhao, Y. Liu, Admittance-based load signature construction for non-intrusive appliance load monitoring, *Energy Build.* 171 (2018) 209–219, <https://doi.org/10.1016/j.enbuild.2018.04.049>
- [11] E. Akarslan, R. Doğan, A novel approach based on a feature selection procedure for residential load identification, *Sustain. Energy Grids Netw.* 27 (2021) 100488, <https://doi.org/10.1016/j.segan.2021.100488>
- [12] S. Ghosh, D. Chatterjee, Non-intrusive identification of harmonic polluting loads in a smart residential system, *Sustain. Energy Grids Netw.* 26 (2021) 100446, <https://doi.org/10.1016/j.segan.2021.100446>
- [13] X. Wu, X. Han, L. Liu, B. Qi, A load identification algorithm of frequency domain filtering under current underdetermined separation, *IEEE Access* 6 (2018) 37094–37107, <https://doi.org/10.1109/ACCESS.2018.2851018>
- [14] J.M. Vann, T. Karnowski, R. Kerekes, C.D. Cooke, A.L. Anderson, A dimensionally aligned signal projection for classification of unintended radiated emissions, *IEEE Trans. Electromagn. Compat.* 60 (1) (2018) 122–131, <https://doi.org/10.1109/TEMC.2017.2692962>
- [15] J.M. Vann, T. Karnowski, A.L. Anderson, Classification of unintended radiated emissions in a multi-device environment, *IEEE Trans. Smart Grid* 10 (5) (2019) 5506–5513, <https://doi.org/10.1109/TSG.2018.2883795>
- [16] D.N. Mylona, A.S. Bouhouras, A digital twin-based framework for load identification using odd harmonic current plots, *Appl. Intell.* 55 (2025) 635, <https://doi.org/10.1007/s10489-025-06512-3>
- [17] H. Mohsenian-Rad, W. Xu, Synchro-waveforms: a window to the future of power systems data analytics, *IEEE Power Energy Mag.* 21 (5) (2023) 68–77, <https://doi.org/10.1109/MPE.2023.3288583>
- [18] H. Lam, G. Fung, W. Lee, A novel method to construct taxonomy electrical appliances based on load signatures, *IEEE Trans. Consum. Electron.* 53 (2) (2007) 653–660, <https://doi.org/10.1109/TCE.2007.381742>
- [19] D. Teshome, T.D. Huang, K.-L. Lian, A distinctive load feature extraction based on Fryze's time-domain power theory, *IEEE Power Energy Technol. Syst. J.* (2016) 1, <https://doi.org/10.1109/JPEITS.2016.2559507>
- [20] B.M. Mulinari, L.D.S. Nolasco, E. Oroski, A.E. Lazzaretti, R.R. Linhares, D.P.B. Renaux, Feature extraction of V-I trajectory using 2-D Fourier series for electrical load classification, *IEEE Sens. J.* 22 (18) (2022) 17988–17996, <https://doi.org/10.1109/JSEN.2022.3194999>
- [21] P. Papageorgiou, D. Mylona, K. Stergiou, A.S. Bouhouras, A time-driven deep learning NILM framework based on novel current harmonic distortion images, *Sustainability* 15 (17) (2023) 12957, <https://doi.org/10.3390/su151712957>
- [22] E.L. de Aguiar, R. Bernardi, A.E. Lazzaretti, D.R. Pipa, E.G. Carati, R. Cardoso, Load identification with photovoltaic distributed generation and a novel public high-frequency dataset, *J. Control Autom. Electr. Syst.* 36 (2) (2025) 345–356, <https://doi.org/10.1007/s40313-025-01152-x>
- [23] Q. Wu, F. Wang, Concatenate convolutional neural networks for non-intrusive load monitoring across complex background, *Energies* 12 (8) (2019) 1572, <https://doi.org/10.3390/en12081572>
- [24] H. Grover, L. Panwar, A. Verma, B.K. Panigrahi, T.S. Bhatti, A multi-head convolutional neural network based non-intrusive load monitoring algorithm under dynamic grid voltage conditions, *Sustain. Energy Grids Netw.* 32 (2022) 100938, <https://doi.org/10.1016/j.segan.2022.100938>
- [25] F. Huo, Q. Luo, T. Yu, K. Wang, Z. Liang, M. Liang, Z. Pan, Non-intrusive load monitoring based on waveform restoration and feature fusion under complex backgrounds, *Meas. Sci. Technol.* 36 (4) (2025) 046115, <https://doi.org/10.1088/1361-6501/adc02a>
- [26] Y. Liu, X. Wang, W. You, Non-intrusive load monitoring by voltage-current trajectory enabled transfer learning, *IEEE Trans. Smart Grid* 10 (5) (2019) 5609–5619, <https://doi.org/10.1109/TSG.2018.2888581>
- [27] A. Mukaroh, T.-T.-H. Le, H. Kim, Background load denoising across complex load based on generative adversarial network to enhance load identification, *Sensors* 20 (19) (2020) 5674, <https://doi.org/10.3390/s20195674>
- [28] E. de Aguiar, A. Lazzaretti, B. Mulinari, D. Pipa, Scattering transform for classification in non-intrusive load monitoring, *Energies* 14 (20) (2021) 6796, <https://doi.org/10.3390/en14206796>
- [29] E.L. de Aguiar, L. da Silva Nolasco, A.E. Lazzaretti, D.R. Pipa, H.S. Lopes, ST-NILM: a wavelet scattering-based architecture for feature extraction and multilabel classification in NILM signals, *IEEE Sens. J.* 24 (7) (2024) 10540–10550, <https://doi.org/10.1109/JSEN.2024.3360188>
- [30] P. Held, S. Mauch, A. Saleh, D. Ould Abdeslam, D. Benyoucef, Frequency invariant transformation of periodic signals (FIT-PS) for classification in NILM, *IEEE Trans. Smart Grid* 10 (5) (2019) 5556–5563, <https://doi.org/10.1109/TSG.2018.2886849>
- [31] A. Faustine, L. Pereira, C. Klemenjak, Adaptive weighted recurrence graphs for appliance recognition in non-intrusive load monitoring, *IEEE Trans. Smart Grid* 12 (1) (2021) 398–406, <https://doi.org/10.1109/TSG.2020.3010621>
- [32] S. Yaemprayoon, J. Srinonchat, Exploring CNN model with inrush current pattern for non-intrusive load monitoring, *Comput. Mater. Contin.* 73 (2) (2022) 3667–3684, <https://doi.org/10.32604/cmc.2022.028358>
- [33] J. CHEN, X. WANG, Non-intrusive load monitoring using Gramian angular field color encoding in edge computing, *Chin. J. Electron.* 31 (4) (2022) 595–603, <https://doi.org/10.1049/cje.2020.00.268>
- [34] Y. Zhang, H. Wu, Q. Ma, Q. Yang, Y. Wang, A learnable image-based load signature construction approach in NILM for appliances identification, *IEEE Trans. Smart Grid* 14 (5) (2023) 3841–3849, <https://doi.org/10.1109/TSG.2023.3239598>
- [35] A.H. Eşlik, E. Akarslan, R. Doğan, A novel approach for residential load identification based on dynamic time warping, *Sustain. Energy Grids Netw.* 38 (2024) 101316, <https://doi.org/10.1016/j.segan.2024.101316>
- [36] T. Hjertberg, S.K. Rönnerberg, Identifying harmonic sources by profiling discrete harmonic cycle time, *Energies* 18 (4) (2025) 853, <https://doi.org/10.3390/en18040853>
- [37] N. Ehsani, F. Aminifar, H. Mohsenian-Rad, Convolutional autoencoder anomaly detection and classification based on distribution PMU measurements, *IET Gener. Transm. Distrib.* 16 (14) (2022) 2816–2828, <https://doi.org/10.1049/gtd2.12424>

- [38] G. Tanoni, E. Principi, S. Squartini, Non-intrusive load monitoring in industrial settings: a systematic review, *Renew. Sustain. Energy Rev.* 202 (2024) 114703, <https://doi.org/10.1016/j.rser.2024.114703>
- [39] A. Yaniv, Y. Beck, Advances in non-intrusive load monitoring for the industrial domain: challenges, insights, and path forward, *Renew. Sustain. Energy Rev.* 210 (2025) 115136, <https://doi.org/10.1016/j.rser.2024.115136>
- [40] D. Cruz-Rangel, C. Ocampo-Martinez, J. Diaz-Rozo, Online non-intrusive load monitoring: a review, *Energy Nexus* 17 (2025) 100348, <https://doi.org/10.1016/j.nexus.2024.100348>
- [41] M.A.F. Pimentel, D.A. Clifton, L. Clifton, L. Tarassenko, A review of novelty detection, *Signal Process.* 99 (2014) 215–249, <https://doi.org/10.1016/j.sigpro.2013.12.026>
- [42] EN 50463-2, *Railway Applications — Energy Measurement on Board Trains*, Brussels, Belgium CENELEC, 2017.
- [43] B. Bahrani, A. Rufer, M. Aeberhard, Catenary voltage support: adopting modern locomotives with active line-side converters, *IEEE Trans. Smart Grid* 3 (1) (2012) 377–387, <https://doi.org/10.1109/TSG.2011.2163090>
- [44] Z. Li, S. Liu, Interference mechanism analysis and mitigation measures with railway signalling equipment from harmonics in the traction system, *Transp. Saf. Environ.* 2 (4) (2020) 271–282, <https://doi.org/10.1093/tse/tdaa023>
- [45] M. Brenna, F. Foiadelli, D. Zaninelli, *Electrical Railway Transportation Systems*, IEEE Press Series on Power and Energy Systems, Wiley, 2018.
- [46] R.S. Salles, S.K. Rönnerberg, Review of waveform distortion interactions assessment in railway power systems, *Energies* 16 (14) (2023) 5411, <https://doi.org/10.3390/en16145411>
- [47] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O'Reilly Media, 2019.
- [48] A. Mariscotti, Data sets of measured pantograph voltage and current of European AC railways, *Data Brief* 30 (2020) 105477, <https://doi.org/10.1016/j.dib.2020.105477>
- [49] A. Mariscotti, Direct measurement of power quality over railway networks with results of a 16.7-Hz network, *IEEE Trans. Instrum. Meas.* 60 (5) (2010) 1604–1612.
- [50] K. Song, W. Mingli, S. Yang, Q. Liu, V.G. Agelidis, G. Konstantinou, High-order harmonic resonances in traction power supplies: a review based on railway operational data, measurements, and experience, *IEEE Trans. Power Electron.* 35 (3) (2020) 2501–2518, <https://doi.org/10.1109/TPEL.2019.2928636>
- [51] M.H.J. Bollen, S.K. Rönnerberg, Primary and secondary harmonics emission; harmonic interaction - a set of definitions, in: 2016 17th International Conference on Harmonics and Quality of Power (ICHQP), IEEE, 2016, pp. 703–708, <https://doi.org/10.1109/ICHQP.2016.7783333>
- [52] G. Dong, G. Liao, H. Liu, G. Kuang, A review of the autoencoder and its variants: a comparative perspective from target recognition in synthetic-aperture radar images, *IEEE Geosci. Remote Sens. Mag.* 6 (3) (2018) 44–68, <https://doi.org/10.1109/MGRS.2018.2853555>
- [53] C. Ge, R.A. de Oliveira, I.Y.-H. Gu, M.H.J. Bollen, Deep feature clustering for seeking patterns in daily harmonic variations, *IEEE Trans. Instrum. Meas.* 70 (2021) 1–10, <https://doi.org/10.1109/TIM.2020.3016408>
- [54] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*, 2014.
- [55] A. Singhal, I. Google, Modern information retrieval: a brief overview, *IEEE Data Eng. Bull.* 24 (Jan 2001).
- [56] M. Singh, S. Nagpal, M. Vatsa, R. Singh, A. Noore, Supervised cosmos autoencoder: learning beyond the Euclidean loss!, *arXiv:1810.06221*, 2018.
- [57] IEEE Std. 2841, IEEE Recommended Practice for Framework and Process for Deep Learning Evaluation (2022) <https://doi.org/10.1109/IEEESTD.2023.10097701>
- [58] D.P.B. Renaux, F. Pottker, H.C. Ancelmo, A.E. Lazzaretti, C.R.E. Lima, R.R. Linhares, E. Oroski, L.D.S. Nolasco, L.T. Lima, B.M. Mulinari, J.R.L. da Silva, J.S. Omori, R.B. dos Santos, A dataset for non-intrusive load monitoring: design and implementation, *Energies* 13 (20) (2020) 5371, <https://doi.org/10.3390/en13205371>
- [59] A.E. Lazzaretti, D.P.B. Renaux, C.R.E. Lima, B.M. Mulinari, H.C. Ancelmo, E. Oroski, F. Pöttker, R.R. Linhares, L.D.S. Nolasco, L.T. Lima, J.S. Omori, R.B. dos Santos, A multi-agent NILM architecture for event detection and load classification, *Energies* 13 (17) (2020) 4396, <https://doi.org/10.3390/en13174396>
- [60] T.S. Roy, A.H. Vijay, *A Robust Anomaly Finder Based on Autoencoders*, 2019.
- [61] S. Koda, F. Melgani, R. Nishii, Unsupervised spectral-spatial feature extraction with generalized autoencoder for hyperspectral imagery, *IEEE Geosci. Remote Sens. Lett.* 17 (3) (2020) 469–473, <https://doi.org/10.1109/LGRS.2019.2921225>
- [62] N. Merrill, A. Eskandarian, Modified autoencoder training and scoring for robust unsupervised anomaly detection in deep learning, *IEEE Access* 8 (2020) 101824–101833, <https://doi.org/10.1109/ACCESS.2020.2997327>
- [63] H. Fukuda, T. Kondo, Anomaly detection in random circuit patterns using autoencoder, *J. Micro/Nanopattern. Mater. Metrol.* 20 (4) (Nov 2021) <https://doi.org/10.1117/1.JMM.20.4.044001>
- [64] A. Gareev, V. Protsenko, D. Stadnik, P. Greshniakov, Y. Yuzifovich, E. Minaev, A. Gimadiev, A. Nikonov, Improved fault diagnosis in hydraulic systems with gated convolutional autoencoder and partially simulated data, *Sensors* 21 (13) (2021) 4410, <https://doi.org/10.3390/s21134410>
- [65] A. Banerjee, A. Das, S. Behra, D. Bhattacharjee, N.T. Srinivasan, M. Nasipuri, N. Das, Carp-DCAE: deep convolutional autoencoder for carp fish classification, *Comput. Electron. Agric.* 196 (2022) 106810, <https://doi.org/10.1016/j.compag.2022.106810>
- [66] A. Imsirovic, A. Rokni, *Autoencoders for Anomaly Detection in Commercial Truck Gearshifts* (Ph.D. thesis) 2024.
- [67] A. van der Schaaf, J.H. van Hateren, Modelling the power spectra of natural images: statistics and information, *Vis. Res.* 36 (17) (1996) 2759–2770, [https://doi.org/10.1016/0042-6989\(96\)00002-8](https://doi.org/10.1016/0042-6989(96)00002-8)
- [68] C. Robertson, Theory and practical recommendations for autocorrelation-based image correlation spectroscopy, *J. Biomed. Opt.* 17 (8) (2012) 080801, <https://doi.org/10.1117/1.JBO.17.8.080801>
- [69] I. Assent, Clustering high dimensional data, *WIREs Data Min. Knowl. Discov.* 2 (4) (2012) 340–350, <https://doi.org/10.1002/widm.1062>
- [70] L. Kaufman, P.J. Rousseeuw, *Finding Groups in Data*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ, USA, 1990, <https://doi.org/10.1002/9780470316801>