



OPEN

DATA DESCRIPTOR

HIFDA - High-Frequency Electrical Voltage and Current Signals from Household Appliances

Víctor M. Navarro¹  , Marta Barragán¹, Rubén Nieto², Jesús Ureña¹ & Álvaro Hernández¹ 

The increasing demand for efficient energy management in smart grids has led to the development of various Non-Intrusive Load Monitoring (NILM) techniques. These aim to disaggregate energy consumption and classify appliances using data from a single-point smart meter at a household's grid mains. With the use of machine learning methods, NILM solutions increasingly rely on datasets for training and validation. While datasets like WHITED, BLOND, and UK-DALE provide insights into consumption patterns, they face limitations such as lack of steady-state data, complicated ground-truth or low sampling rates, which hinder detecting low-power appliances. High sampling rates, however, improve classification accuracy and enable identifying these devices. This study introduces the High Frequency household electrical signals DATaset (HIFDA), a high-frequency dataset capturing steady-state signals from 14 household appliances at 100 kSPS, including the empty grid. Data, collected via a custom System-on-Chip (SoC) device, focuses on active consumption and includes multiple time windows. HIFDA, hosted on Zenodo, ensures its suitability for modern NILM research and other applications.

Background & Summary

The growing demand for efficient energy management and monitoring systems, driven by the rise of smart grids and intelligent power systems, has encouraged the definition and development of different techniques and methods in the last decades. One of them is Non-Intrusive Load Monitoring (NILM), which is intended to disaggregate the energy consumption of individual electrical loads (or appliances) by only analysing the electrical signals measured by a single-point smart meter at the entrance of the mains in a household or building^{1,2}. From this starting point, NILM techniques have expanded in different topics and fields, such as load classification, usage patterns of appliances, prediction of energy consumption, etc. Although the initial and main application of NILM techniques has been and is efficient energy management^{3,4}, they have also emerged as a feasible solution for non-invasive monitoring of elderly, in order to promote their independent living in their own houses⁵. In this context, NILM allows the detection and assessment of certain daily activities that can be associated to the on/off switching of some appliances.

Most previous works in NILM have often focused on detecting and disaggregating those appliances with high energy consumption, since they are more interesting from the point of view of energy saving and efficient management. Nevertheless, when dealing with applications related to activity monitoring in households, other loads with lower energy consumption can become even more relevant. For these appliances with low consumption, a key aspect for NILM performance is the sampling rate of the electrical signals involved. High sampling frequencies commonly provide with more information, which allows the capture of unique signatures that may differentiate one appliance from another, especially if it is a low-power device such as a laptop or a charger⁶.

Furthermore, machine learning^{7,8}, and particularly deep neural networks^{9,10}, have become a predominant alternative in NILM techniques in recent years, whether they involve classification or regression aspects. These alternatives always require a large set of data that supports a feasible training of the algorithm. For that purpose, a huge effort has been carried out in the last years to capture, format and make available electrical consumption data from different real households and appliances. Due to the primary interest focused on energy management, most of these datasets comprise long periods of time at low sampling rates, whereas they sometimes provide the

¹Electronics Department, University of Alcalá, Alcalá de Henares, 28805, Spain. ²Electronic Technology Department, Rey Juan Carlos University, Móstoles, 28933, Spain. ✉e-mail: victormanuel.navarro@uah.es; alvaro.hernandez@uah.es

Dataset	Region	Number of appliances	Sampling rate for aggregate signal	Sampling rate for individual appliance	Type	Available
BLOND-50 ¹²	Germany	53	50 kHz	6.4 kHz	Continuous	Yes
BLOND-250 ¹²	Germany	53	250 kHz	50 kHz	Continuous	Yes
BLUED ²³	USA	11	12 kHz	—	Continuous	No
COOL ²⁴	France	42	—	100 kHz	On/Off	Yes
PLAID ²⁵	USA	537	—	30 kHz	On/Off	Yes
REDD ¹⁴	USA	44	1 Hz	15 kHz	Continuous	No
SustDataED ²⁶	Portugal	17	12.8 kHz	0.5 Hz	Continuous	No
UK-DALE ¹³	UK	54	16 kHz	1/6 Hz	Continuous	Yes
WHITED ¹¹	Several	110	—	44.1 kHz	On/Off	Yes
HIFDA ¹⁵	Spain	14	—	50 kHz	Continuous	Yes

Table 1. Datasets providing electrical signals sampled at high rates for household appliances.

disaggregated consumption of some appliances as well. Nevertheless, it is also possible to find some well-known open-access high-frequency datasets for electrical signals, such as WHITED¹¹, BLOND¹², UK-DALE¹³, or REDD¹⁴, each providing valuable data on electrical consumption.

The WHITED dataset, with a sampling rate of 44.1 kHz, focuses on capturing transient electrical signals from over 100 household and industrial devices, making it ideal for detailed signal analysis and appliance disaggregation. Meanwhile, the BLOND-250 dataset, sampled at 50 kHz for individual appliances and 250 kHz for aggregate measurements, offers long-term high-resolution data from over 50 household appliances, thus enabling comprehensive studies of energy consumption patterns and appliance usage routines across different seasons. Its large set, over 1.5 TB, makes it suitable for machine learning applications. The UK-DALE dataset provides both high-frequency (16 kHz) voltage and current aggregate data and low-frequency (1/6 Hz) data for individual appliances from some UK households, covering up to 54 appliances per home. It is extensively used in NILM research for both short-term transient analysis and long-term monitoring. In contrast to UK-DALE, the REDD dataset, while also offering aggregated and individual current and voltage data, offers higher acquisition frequencies on individual appliances (15 kHz) rather than on the aggregate signal (1 Hz). Additionally, its data were obtained from houses located in the USA, so their fundamental grid frequency (60 Hz) is different from the European one (50 Hz). As can be noticed, these datasets vary in their sampling frequencies and appliance diversity, with WHITED excelling in device variety, and BLOND-250, UK-DALE and REDD providing long-term recordings. However, despite their strengths, they also present some drawbacks, such as the lack of steady-state data in WHITED, the lack of a straightforward solution for isolating individual appliances using the provided ground-truth in BLOND-250, the limited sampling frequency for individual electrical loads in UK-DALE and REDD, and the different fundamental frequency of the grid compared to the European one in the case of REDD, which could give rise to implementation incompatibility problems.

A further comparison between the main datasets focused on electrical signals sampled at high frequency rates is shown in Table 1, where the dataset discussed in this paper has also been included in the last row for convenience, as it will be compared with the existing datasets on a later explanation. As can be observed, there are notable differences between them, including the diverse sampling frequencies at which the considered appliances are captured. These sampling rates are not the same for the individual signals from each appliance as the aggregated signals (one or more simultaneous appliances), commonly sampled at higher frequencies, which makes characterizing and classifying the isolated appliances difficult. Another notable difference is the type of signal captured, where some datasets contain only the on/off transients of the appliance, whereas others contain a continuous capture of the appliance operation, which also includes its steady state. Table 1 also includes a section indicating the region of each dataset: this is a relevant aspect to take into account, since it determines the fundamental frequency of the electrical signals. Finally, Table 1 provides their availability for every data set, since some are no longer available to be downloaded, which increases the scarcity, already quite prevalent, of datasets with high sampling rates.

In this context, this work describes a new dataset for NILM applications containing steady-state electrical signals from individual common household appliances acquired at high sampling rates. A total of 14 appliances were measured at an acquisition rate of 100 kSPS, capturing the intervals when the target appliance was active and consuming energy (no idle states). Also, the empty grid (no active appliance) was captured under the same conditions. Data acquisition was performed in a supervised laboratory environment by using a custom analog front-end, capable of acquiring electrical signals at up to 500 kHz. It consists of a FPGA-based (Field-Programmable Gate Array) System-on-Chip (SoC) architecture, responsible for managing two Analog-Digital Converters (ADCs) and the corresponding two Analog Front-Ends (AFE), one for the voltage signal and another for the current one. The main novelty, and motivation for the development of HIFDA¹⁵ (High Frequency household electrical signals DATaset), is to provide with a suitable and large enough amount of samples to enable the proper training of deep neural networks and machine learning solutions that may use high-frequency harmonics from the involved electrical signals to identify household appliances, including those with low energy consumption. The possibility of working with these appliances implies a significant improvement in those applications dealing with daily activity monitoring and behavior routines recognition, which can be tracked more suitably with the identification of these low-power loads.

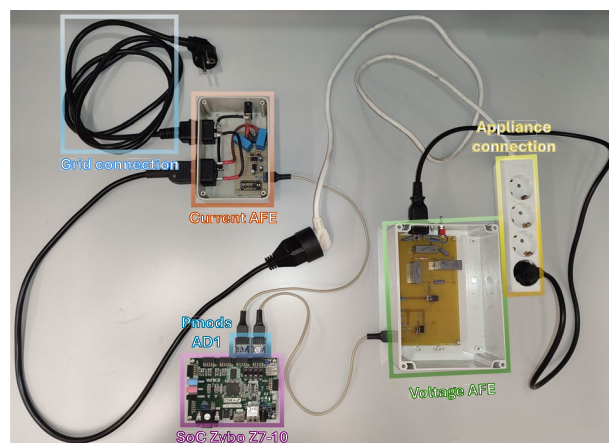


Fig. 1 Complete acquisition system, with all its elements assembled.

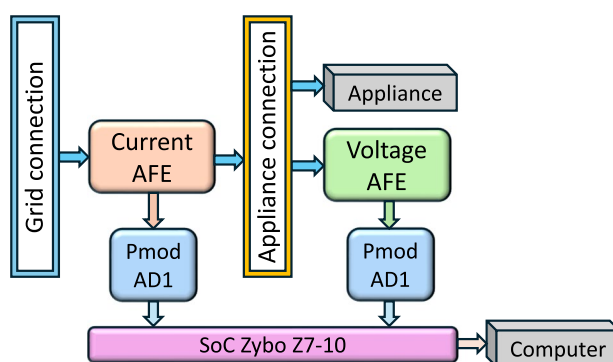


Fig. 2 Block diagram of the complete acquisition system layout, each element sharing the same color assigned in Figure 1.

Compared to other previous datasets, as can be observed in Table 1, HIFDA provides individual appliance data captured at a higher acquisition rate than most of them, the only exceptions being the COOLL (100 kHz) and Blond-250 (50 kHz) datasets. However, whereas the acquisition rate of COOLL is twice that of HIFDA, the information contained in COOLL is only about the on/off events of the appliance, with no data about the steady state, which may be problematic for the accurate recognition of appliances with different operating phases. HIFDA contains data of the steady state for each appliance, therefore providing more information about their general operation. On the other hand, the Blond-250 dataset presents the same sampling rate as HIFDA, although it has a complex ground truth that requires manual supervision in order to fully isolate the information related to every appliance. Since the HIFDA data have been captured in a fully supervised manner, it provides the end user with a direct and reliable ground truth.

Methods

This Section firstly details the experimental setup developed for measuring the corresponding electrical signals, voltage and current, for HIFDA dataset. Furthermore, a detailed description of the appliances involved in the measurements is provided.

Experimental Setup. The system with which data have been captured is based on a Zybo z7-10¹⁶ development platform. It contains a FPGA, where a SoC architecture has been defined and implemented. It also includes two AFEs and two AD7476A¹⁷ ADCs, which are combined to acquire the aforementioned voltage and current signals. Figure 1 shows the complete acquisition system with all its elements assembled, accompanied by an explanatory block diagram of its layout, shown in Fig. 2. As can be observed, the current AFE is connected in series between the mains and the appliances, whereas the voltage AFE is connected in parallel to the appliance's socket. Both AFEs divert a low voltage to the ADCs from two Pmod AD1 boards, which are in charge of acquiring the corresponding signals. The ADCs are linked to the Zybo Z7-10 platform, so the designed SoC architecture manages the converters and uploads the incoming acquired samples towards a computer through an available Ethernet link.

The AFEs include transformers and amplifiers to adjust the signals to a manageable range for the ADC (0 to 3.3 V). In the case of voltage, as can be observed in Fig. 3, a Notch filter is used to remove the 50 Hz component (the standard frequency in the European power grid), while allowing the amplification of higher frequency

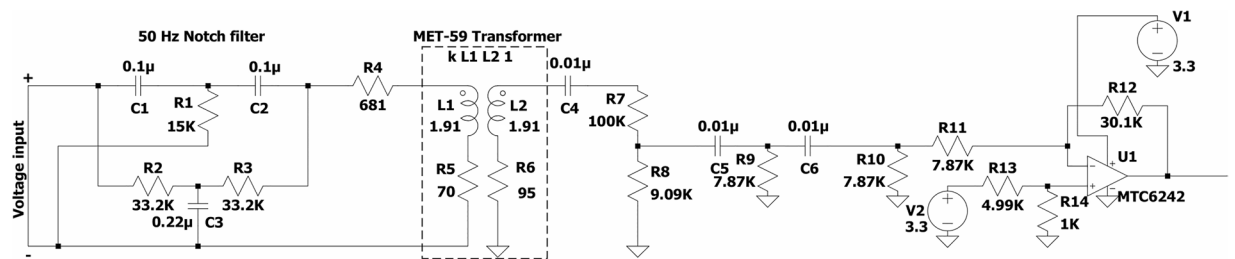


Fig. 3 Electrical scheme of the AFE dedicated to the acquisition of voltage signals.

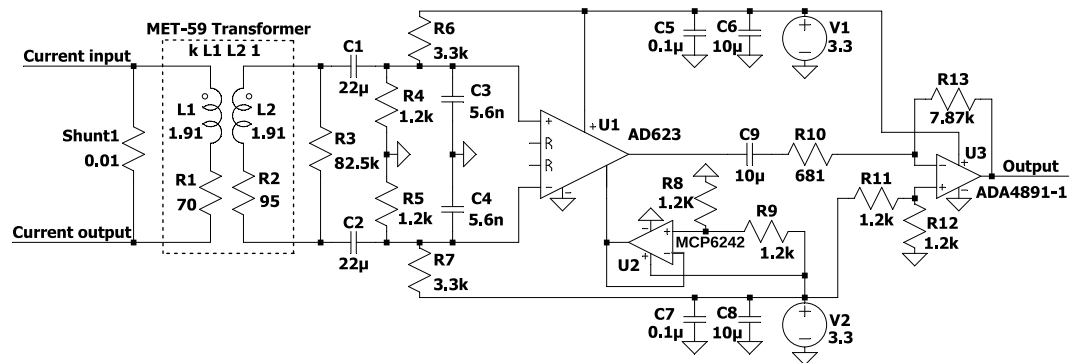


Fig. 4 Electrical scheme of the AFE used for the acquisition of current signals.

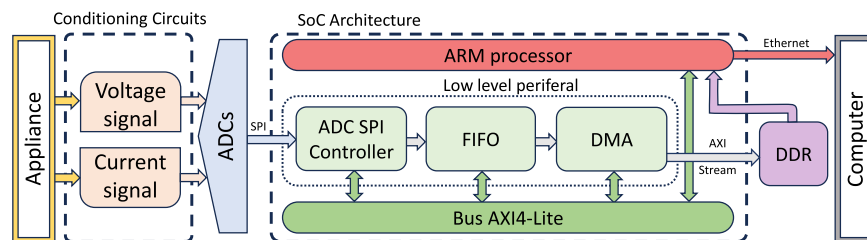


Fig. 5 Hardware block diagram including the Soc architecture.

harmonics. The Notch filter is then followed by a MET-59 transformer, another filtering, and an amplification stage based on the MTC6242 operational amplifier.

The current AFE, shown in Fig. 4, uses a shunt resistor to generate a small voltage drop through a MET-59 transformer, which is then filtered by a differential amplifier AD623 to eliminate as much ambient noise as possible, and further amplified to adequate the signal to the desired 0 to 3.3 V range, based on an ADA4891 operational amplifier.

Note that the ADCs can achieve an acquisition rate of up to 1 MSPS, with a 12-bit resolution, thus allowing the capture of key details from the electrical signals, including high-frequency harmonics. The use of separate conditioning circuits and ADCs for voltage and current helps minimize noise and increase system safety. The data acquired by the AD7476A ADCs are sent to the SoC architecture via the Serial Peripheral Interface (SPI) communication protocol. Although the ADCs are capable of achieving acquisition rates of up to 1 MSPS, the system was set up at 100 kSPS, as a trade-off in this application between the high-frequency information acquired and the amount of data samples to be handled. Consequently, the proposed acquisition system is thus able to capture the voltage signals with a final bandwidth ranging from 300 Hz to 50 kHz, and the current signals with a final bandwidth from 30 Hz to 50 kHz.

In regard to the SoC architecture shown in Fig. 5, initially, when the SoC is booting up, the ADC SPI controller is configured by means of an Advanced eXtensible Interface (AXI)4-Lite bus from the Advanced RISC Machine (ARM) processor, by writing the desired sampling rate (in kHz) on the internal registers of the AXI peripheral. The Direct Memory Access (DMA) is also configured via the AXI bus, defining the Double Data Rate (DDR) memory base address where the incoming samples will be stored. Since the proposal actually has two independent ADCs, two different base addresses have been defined: one for each ADC, making sure to leave enough memory space for the correct data storage. Once both modules are configured, the data acquisition process can begin.

Appliance	Brand	Model	Power (W)	Main load type
Air conditioner	Cecotec	ForceClima 9050	1010	Reactive
Charger	—	HNHB050300E5	120	Switched-source
Coffee maker	Mandine	MCM85225R-16	1000	Resistive
Computer	Tacens	RADIX III 720	720	Switched-source
Griddle	Solac	PA5255	2200	Resistive
Hair dryer	Philips	TYPE HP8251	2300	Resistive
Heater	Haverland	ST-1	2000	Resistive
Iron	Taurus	Artica 2800 zaffiro (ver 2)	2800	Resistive
Laptop	Toshiba	Satellite U200-141	360	Switched-source
Light	OSRAM	L22W/21-840 C G10Q	22	Resistive
Microwave	LG	MS1929G	800	Reactive
Monitor	LG	Flatron W2242T	240	Switched-source
Vacuum	Bosch	Series 2	700	Reactive
Washing machine	LG	F2WR5S09A0W	2200	Reactive — Resistive
Empty grid	—	—	0	—

Table 2. Description of the electrical appliances involved in the creation of the HIFDA dataset.

Appliance	No. of files for current				No. of files for voltage				Format
	10.24 ms	163.84 ms	1310.72 ms	Full time (aprox 5.4 s)	10.24 ms	163.84 ms	1310.72 ms	Full time (aprox 5.4 s)	
Air conditioner	23916	1520	200	50	23916	1520	200	50	.txt
Charger	23916	1520	200	50	23916	1520	200	50	.txt
Coffee maker	23916	1520	200	50	23916	1520	200	50	.txt
Computer	23916	1520	200	50	23916	1520	200	50	.txt
Empty grid	23916	1520	200	50	23916	1520	200	50	.txt
Griddle	23916	1520	200	50	23916	1520	200	50	.txt
Hair dryer	23916	1520	200	50	23916	1520	200	50	.txt
Heater	23916	1520	200	50	23916	1520	200	50	.txt
Iron	23916	1520	200	50	23916	1520	200	50	.txt
Laptop	23916	1520	200	50	23916	1520	200	50	.txt
Light	23916	1520	200	50	23916	1520	200	50	.txt
Microwave	23916	1520	200	50	23916	1520	200	50	.txt
Monitor	23916	1520	200	50	23916	1520	200	50	.txt
Vacuum	23916	1520	200	50	23916	1520	200	50	.txt
Washing machine	23916	1520	200	50	23916	1520	200	50	.txt

Table 3. Disposition of the HIFDA dataset files.

The data acquired by each ADC is received through the SPI link and processed by a low-level specific peripheral implemented in the FPGA. The 12-bit samples coming from each ADC channel are managed and temporarily stored by the ADC SPI controller, which extends the sign to provide two 16-bit data streams. These data are then stored in a FIFO memory with a datawidth of 32 bits: the 16 most significant bits correspond to the second channel in the ADC, whereas the 16 least significant bits are for the first channel. Whether the DMA is ready for data transfer, the samples gathered in the FIFO are sent using an AXI-Stream bus to the DDR memory for storage. The ARM processor then reads that information and processes it independently, by separating the data corresponding to each channel, and normalizing and converting each signal into the reference voltage range of 0 to 3.3 V, as described in (1). Further information about the involved SoC architecture can be found in a previous work¹⁸.

$$x_{sc} = \left(\frac{V_{ref} * x_{adc}}{2^{(N_{bits}-1)}} \right) + \frac{V_{ref}}{2} \quad (1)$$

Where x_{sc} is the resulting scaled sample; $V_{ref} = 3.3$ V is the reference voltage; x_{adc} is the acquired sample; and $N_{bits} = 12$ is the number of resolution bits. This process ensures that the acquired signals agree with the input voltage span in the acquisition, and, therefore, they are ready for the following processing dedicated to load identification and consumption analysis algorithms.

The acquired samples are uploaded through an Ethernet link available in the SoC architecture to a computer, where the incoming data are received and stored in files with a text format (.txt), using Matlab^{®19}. For the

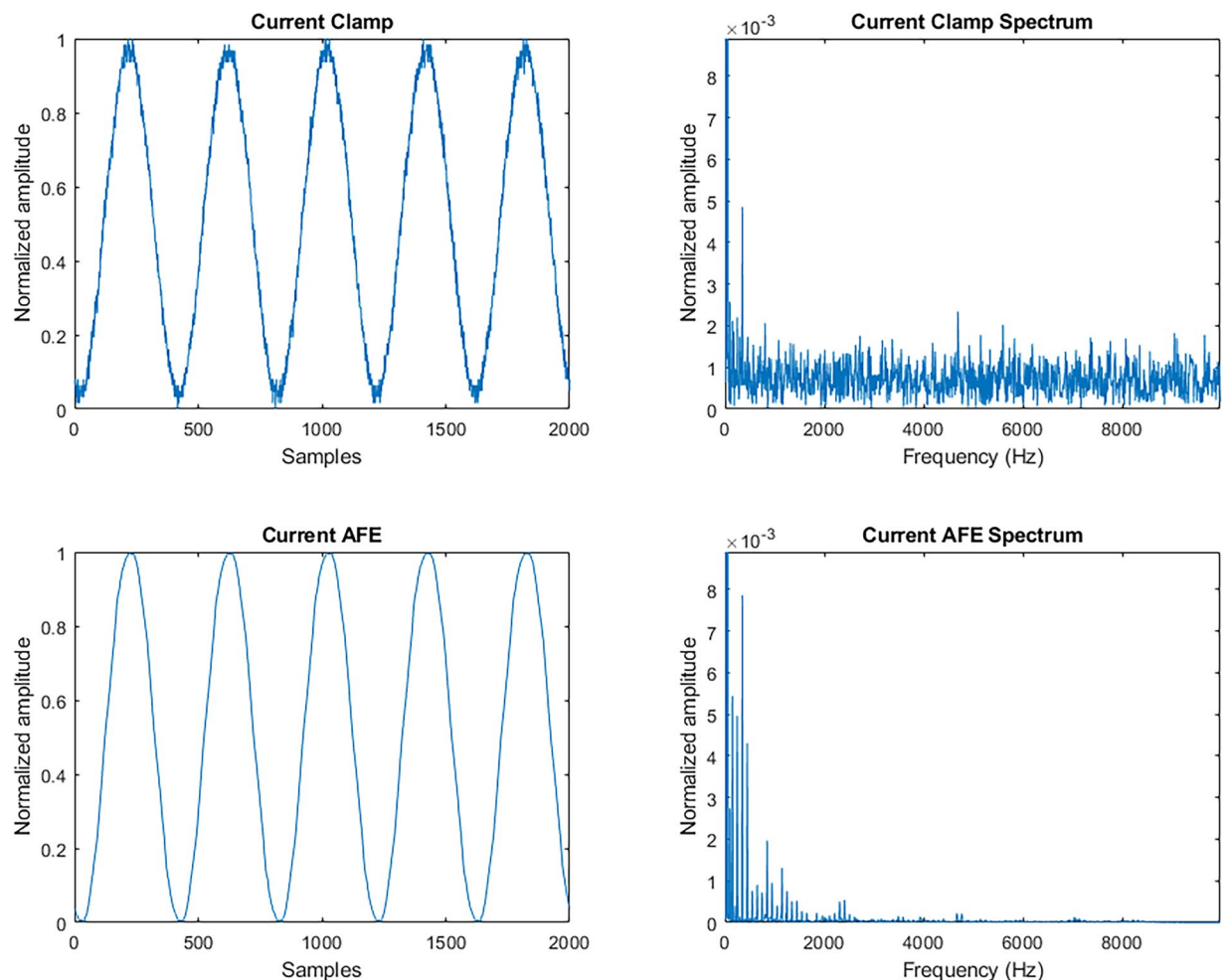


Fig. 6 Comparison between the current signals from a hair dryer acquired by the designed AFE (lower plots) and those obtained by a commercial current clamp (upper plots).

creation of this dataset, 50 windows of the voltage and current signals were taken for fourteen isolated devices, described in Table 2, as well as for the empty electrical grid. Each window contains approximately 540,000 samples acquired at 100 kSPS. This implies that the acquired windows have a length of 5.4 seconds. The appliances selected for the creation of this dataset correspond to examples with different electrical behavior, thus containing resistive loads, reactive ones, and switched source loads. They also present different power consumptions, involving appliances with consumptions ranging from 22 W to 2.8 kW. Low-power devices are particularly interesting, since the characteristics of their electrical signals are not easily distinguishable at low sampling rates (note that these appliances are not commonly considered in other previous datasets).

For convenience, the acquired windows have been split in order to create three additional datasets, containing time slots of 10.24 ms, 163.84 ms and 1310.72 ms. These lengths may be considered in the training of different neural network topologies and architectures. The size of these windows has been set so that the number of data contained in each window is a power-of-two length, since this facilitates the processing tasks for many algorithms based on the Fast Fourier Transform or the Wavelet Transform, reducing their computational cost. At the same time, it is pursued to create windows containing time slots with a length of at least 10 ms, 100 ms and 1000 ms at 100 kSPS, respectively. In any case, the user is free to extract windows of any desired size from the global information contained within the frames.

Additionally, in order to increase the possible utility of this dataset, a general characterization of the data has been made, which allows the conversion of the samples into its corresponding physical values. On the one hand, for the current signal, the obtained waveform data have been compared with the samples obtained in parallel with a Yokogawa 96001²⁰ commercial clamp ammeter. This was used to estimate a scaling factor for the acquired signals, which was used to experimentally fit the conversion from the acquired value x_{sc} into the measured current I_{meas} (2).

$$I_{meas} = (x_{sc} - 1.6462) * 12.55 \quad (\text{A}) \quad (2)$$

On the other hand, for the voltage signal, the input voltage signal is acquired, not only by the proposed setup, but also by a Tektronix MSO5204B²¹ commercial oscilloscope. Since a 50 Hz Notch filter in the proposed setup

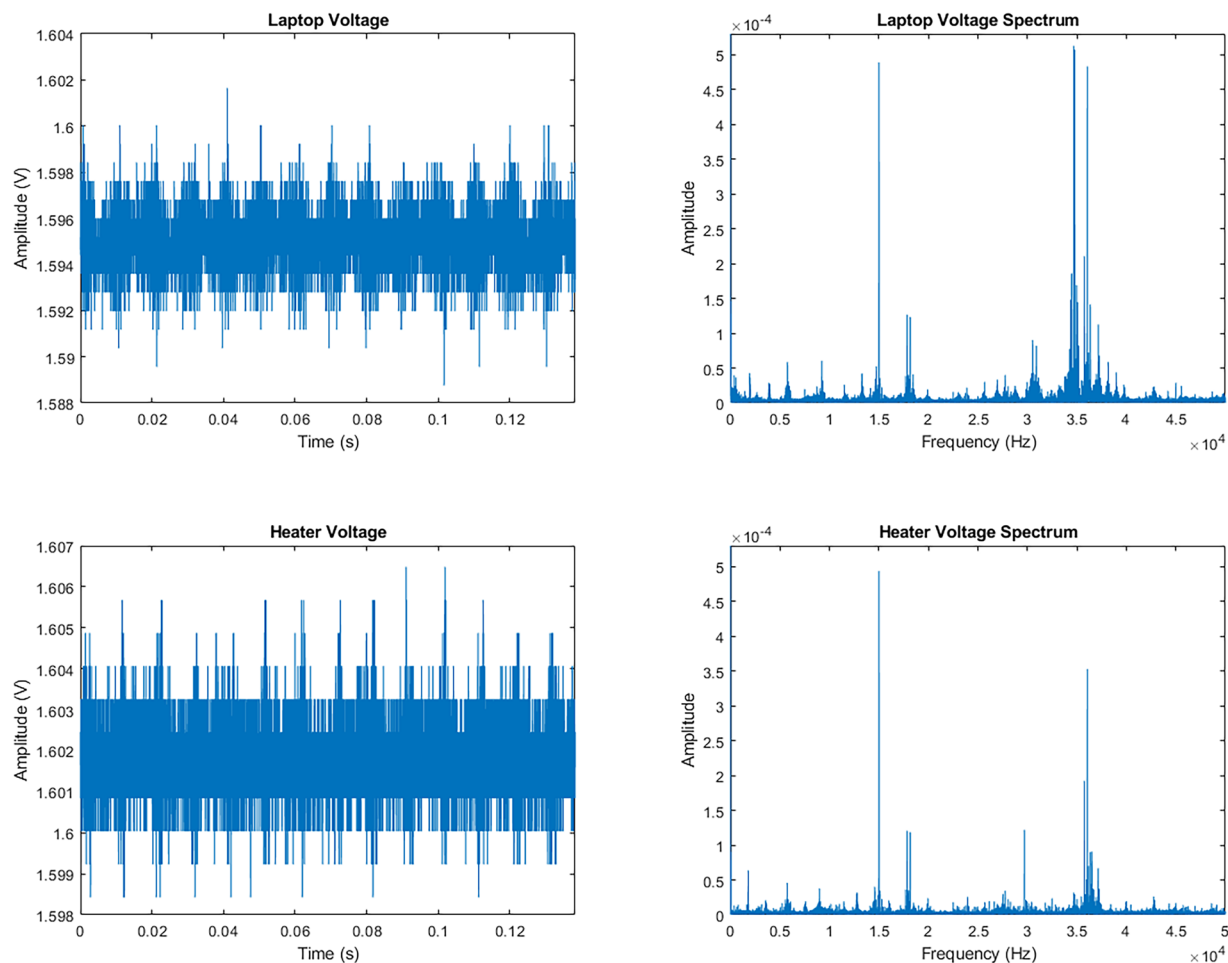


Fig. 7 Examples of voltage signals from the HIFDA dataset for a laptop and a heater, together with their corresponding frequency spectrum.

is in charge of filtering out the main frequency component, both signal cannot be compared directly. Instead, the Fast Fourier Transform is implemented to estimate the scaling factor for those frequency components existing in both acquired signals. In this way, it is also possible to experimentally adjust the conversion from the acquired sample x_{sc} into the measured physical voltage V_{meas} in Volts (3).

$$V_{meas} = (x_{sc} - 1.5992) * 33.64 \quad (\text{V}) \quad (3)$$

Data Records

The HIFDA¹⁵ dataset is available at Zenodo as a compressed ZIP file within an open-access repository (<https://doi.org/10.5281/zenodo.13884627>). It contains five main folders. One of this folders, “0.Img_Appliances”, provides with pictures of the fourteen appliances whose electrical signals have been captured to create this dataset. The rest of the folders correspond to the different window divisions described before. Inside every one of them, there is a “Current” folder and a “Voltage” one, where it is possible to find a subfolder for every appliance under analysis. This subfolder contains the associated data in multiple text files (.txt). The total number of files and their disposition can be observed in Table 3. It is worth noting that, even though the number of available files is higher for the smaller window configurations, the information in all these window length configurations is mostly the same, as all of them come from the full-time recordings. It is important to keep in mind that the bandwidth of the captured voltage signal ranges from 300 Hz to 50 kHz, so the fundamental component of the grid, located at 50 Hz, does not appear on the acquired samples. Meanwhile, the captured current signal has a bandwidth ranging from 30 Hz to 50 kHz, approximately, thus including the fundamental component.

Inside the ZIP file there is also a “Readme.txt” file that contains general information about the dataset and how to use it, including information about the bandwidth for every signal, the initial 0 V to 3.3 V scaling and the corresponding conversions to physical voltages and currents.

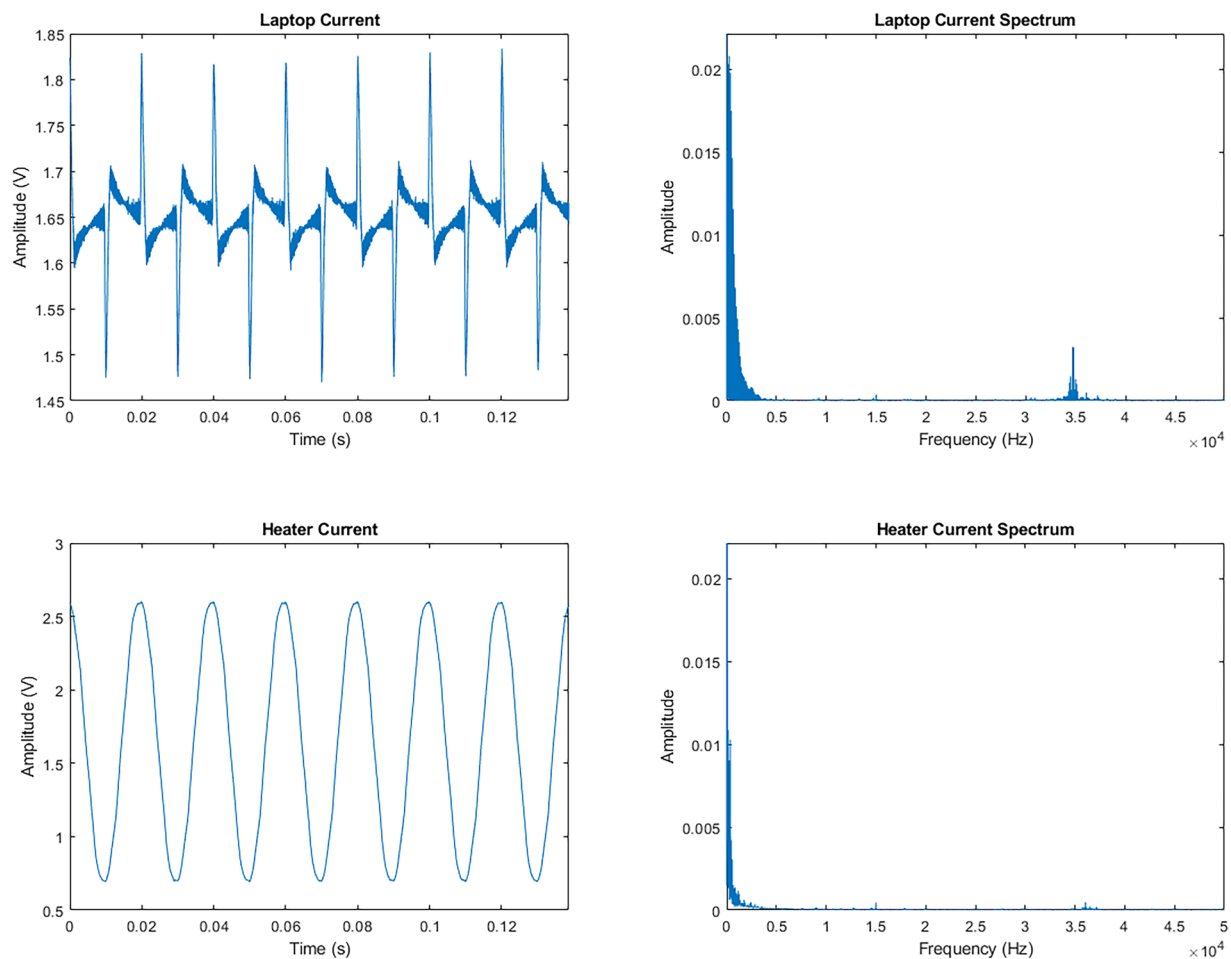


Fig. 8 Examples of current signals from the HIFDA dataset for a laptop and a heater, together with their corresponding frequency spectrum.

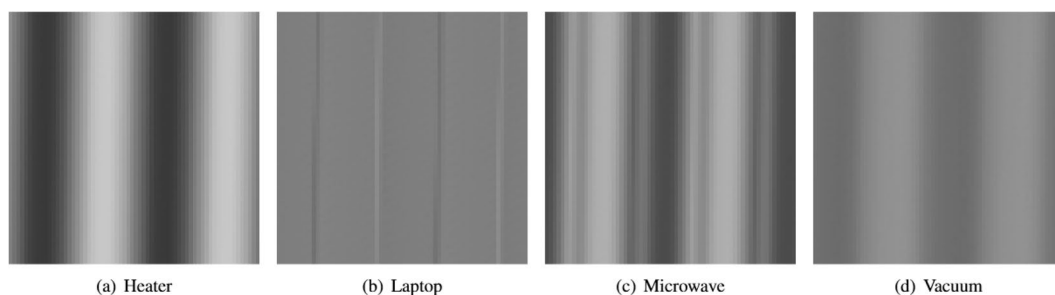


Fig. 9 Examples of the 64×64 images used to train the CNN, generated from the current signals of different appliances.

Technical Validation

In order to verify the technical quality and practical utility of the generated HIFDA dataset, a characterization of the AFE modules was carried out, where their Bode diagrams were calculated to validate that the obtained bandwidths correspond to the desired ones. This has been implemented using a signal generator and an oscilloscope, by inserting a sine signal in both AFEs and observing the amplitude of the output signal, with the shunt resistor removed in the case of the current AFE so as not to cause a short circuit. For more information, refer to previous work¹⁸.

Furthermore, as a method of verifying the waveform of the current signal, a comparison has been made between the current signals of several household appliances captured simultaneously with the developed system and those captured by means of a commercial current clamp. As an example, the results for a hair dryer are shown in Fig. 6, where a decimation of the data captured by the developed system was necessary to resample the data at 20 kSPS, since this is the upper limit for the bandwidth of the current clamp. In addition, both signals

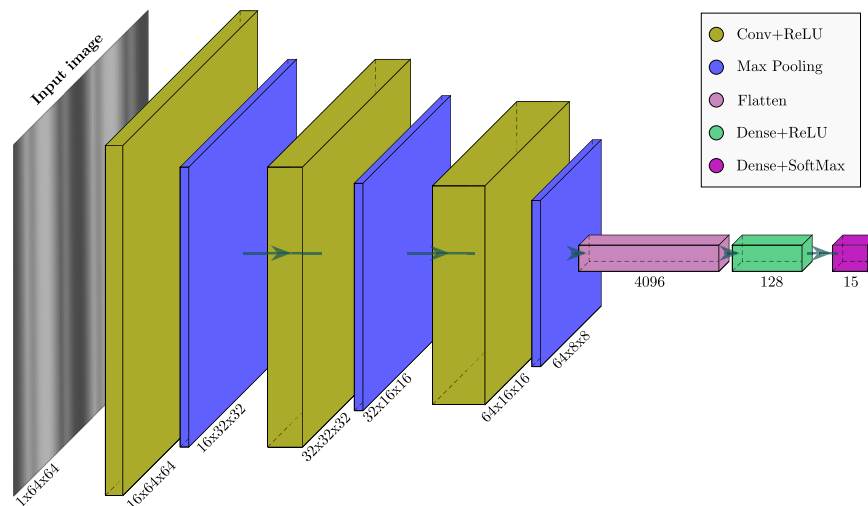


Fig. 10 Block diagram of the CNN architecture proposed for the validation of the HIFDA dataset in the appliance identification.

were normalized in order to establish a fair comparison. The signal obtained by the current AFE is cleaner than that provided by the current clamp, mainly due to its differential configuration, implemented to reduce ambient noise in the developed current AFE, which is not present in the commercial current clamp.

Other examples coming from the HIFDA dataset are depicted in Fig. 7 and Figure 8, where voltage and current signals for a laptop and a heater, along with their corresponding spectrum, are represented. As can be observed in Fig. 7, in the voltage captured from both appliances, the fundamental frequency at 50 Hz is clearly reduced by the applied Notch filter, whereas the signals around 35 kHz are more prevalent in the laptop due to the its switched-source nature, which inserts high-frequency harmonics in the grid. Furthermore, as Fig. 8 shows, the heater current signal waveform has a clean sine shape due to its mainly resistive behavior, whereas the laptop current signal waveform, due to its switched-mode source, contains peaks every half period accompanied by high-frequency noise around the 35 kHz range.

Since this dataset is intended to be used as a training set for deep neural networks and, in general, for machine learning algorithms applied to load identification purposes in the domain of NILM techniques, a validation has been carried out by training a Convolutional Neural Network (CNN), based on a model described in a previous work¹⁰. For clarity's sake, only 100 current captures for every household appliance were considered from the HIFDA dataset. A window of 4096 samples from every selected current captures is used to create the 64×64 input images that are provided to the convolutional neural network. These images are normalized from the 0 V to 3.3 V range to the 0 to 255 range, according to (4), and converted into the 64×64 images where normalised samples x_n are arranged in successive columns, x_{sc} being the original data.

$$x_n = \frac{255}{3.3} * x_{sc} \quad (4)$$

The training of the aforementioned CNN has been carried out, based on these 64×64 input images, thus containing a set of 100 images for each one of the 14 appliances, including also the empty grid. Figure 9 shows some examples of the images generated from the current signals of different appliances, where it is possible to distinguish the visual differences that may be used by the CNN to identify the different electrical loads.

The architecture of the CNN used in this test to classify the involved appliances is shown in Fig. 10, where it is possible to observe the internal structure and the purpose of each layer in the model. As has been already mentioned, the neural network input is a 64×64 grayscale image, followed by a 2D convolutional layer with a kernel size of 3×3 , activated with the ReLU (Rectified Linear Unit) function. Afterwards, there is a max pooling layer with a window size of 2×2 . This same structure is repeated two successive times. Finally, the model contains a flatten layer, a fully connected layer with 128 neurons, activated by ReLU, and a last dense layer with 15 neurons, that uses a SoftMax activation function to classify the fifteen classes corresponding to the 14 appliances and the empty grid from the HIFDA dataset.

Finally, the classification results of this test are shown in Table 4, which have been cross validated with 10 different training iterations. All of them have been carried out with a distribution of 70 percent of the generated input images for training, 20 percent for validation, and 10 percent for testing. It is worth noting that those appliances that have a higher rate of confusion among them are those with a resistive behavior and similar power, such as the griddle (2200 W), the hair dryer (2300 W) and the heater (2000 W). Otherwise, this is not the case of the iron (2800 W) or the coffee maker (1000 W) since, although they have a resistive load, they also feature a significant difference in power consumption. In the case of the loads with low energy consumption, such as the charger (120 W), the light (22 W), the monitor (240 W), and the laptop (360 W), the contrast is so low in the corresponding generated images that their signals are confused, even with the empty grid. In addition, it is common for those appliances with low energy consumption to be confused with those with similar switched-source

Appliance	Confusion Matrix															Precision	Recall	F1-score	Support
Air conditioner	71	0	4	0	0	0	0	0	0	0	0	0	0	25	0	0.6893	0.7100	0.6995	100
Charger	0	31	0	0	65	0	0	0	0	0	4	0	0	0	0	0.3010	0.3100	0.3054	100
Coffee maker	14	0	84	0	0	0	0	0	0	0	0	2	0	0	0	0.8842	0.8400	0.8615	100
Computer	0	7	0	35	9	0	0	0	0	16	15	0	18	0	0	0.5147	0.3500	0.4167	100
Empty grid	0	10	0	0	90	0	0	0	0	0	0	0	0	0	0	0.3543	0.9000	0.5085	100
Griddle	0	1	0	0	4	45	33	17	0	0	0	0	0	0	0	0.4286	0.4500	0.4390	100
Hair dryer	0	0	0	0	0	46	48	6	0	0	0	0	0	0	0	0.5783	0.4800	0.5246	100
Heater	0	0	0	0	0	14	2	83	0	0	0	0	0	0	1	0.7685	0.8300	0.7981	100
Iron	0	0	0	0	0	0	0	0	94	0	0	0	0	0	6	0.9495	0.9400	0.9447	100
Laptop	0	1	0	26	12	0	0	0	0	38	3	0	20	0	0	0.5205	0.3800	0.4393	100
Light	0	24	0	1	41	0	0	0	0	0	34	0	0	0	0	0.4789	0.3400	0.3977	100
Microwave	1	0	7	0	0	0	0	0	0	0	0	92	0	0	0	0.9684	0.9200	0.9436	100
Monitor	0	25	0	5	25	0	0	0	0	16	14	0	15	0	0	0.2778	0.1500	0.1948	100
Vacuum	10	0	0	0	0	0	0	0	0	0	0	0	0	90	0	0.7500	0.9000	0.8182	100
Washing machine	7	4	0	1	8	0	0	2	5	3	1	1	1	5	62	0.8986	0.6200	0.7337	100
Accuracy																		0.6080	1500
Macro avg.																0.6242	0.6080	0.6017	1500
Weighted avg.																0.6242	0.6080	0.6017	1500

Table 4. Classification results from the CNN applied to the loads included in the HIFDA dataset.

behavior, since they would have a similar shape on the generated images fed into the neural network, this being the case of the computer, laptop, monitor, and charger. As for the air conditioner (1010 W), the vacuum (720 W) and the coffee maker (1000 W), their power consumptions are similar and, even though their type of load may be different, as this CNN uses only the current signal, the difference between a resistive and a reactive signal is minimal, since the signal phase shift cannot be obtained without the use of voltage. All of these are common issues in most NILM techniques, and a challenge still open to research on, supported by datasets such as HIFDA.

Usage Notes

The HIFDA dataset can be reused by unzipping the uploaded ZIP file and importing the data using any software package or environment capable of dealing with text files (.txt), such as Matlab®, Python²², or any other packages and/or frameworks. Depending on the particular purpose for which this dataset is to be applied, data may be used unaltered or the suggested scaling might be considered to transform the available values into physical signals.

Code availability

No custom code is needed.

Received: 6 December 2024; Accepted: 19 March 2025;
Published online: 29 March 2025

References

1. Kong, X. *et al.* Household energy efficiency index assessment method based on non-intrusive load monitoring data. *Appl. Sci.* **10**, 3820, <https://doi.org/10.3390/app10113820> (2020).

2. Kumar, P. *et al.* Smart grid metering networks: A survey on security, privacy and open research issues. *IEEE Commun. Surv. & Tutorials* **21**, 2886–2927, <https://doi.org/10.1109/COMST.2019.2899354> (2019).

3. Raiker, G. A. *et al.* Internet of things based demand side energy management system using non-intrusive load monitoring. *IEEE Int. Conf. on Power Electron., Smart Grid and Renew. Energy (PESGRE)*, 1–5, <https://doi.org/10.1109/PESGRE45664.2020.9070739> (2020).

4. Yan, Q., Xudong, W. & Zun, W. Applications of NILM in the optimization management of intelligent home energy management system. *IEEE Sustain. Power Energy Conf. (iSPEC)*, 2300–2305, <https://doi.org/10.1109/iSPEC50848.2020.9351198> (2020).

5. Alcalá, J. M., Ureña, J., Hernández, A. & Gualda, D. Assessing human activity in elderly people using non-intrusive load monitoring. *Sensors* **17**, 351, <https://doi.org/10.3390/s17020351> (2017).

6. Sun, M. *et al.* Non-intrusive load monitoring system framework and load disaggregation algorithms: A survey. *Int. Conf. on Adv. Mechatron. Syst. (ICAMechS)*, 284–288, <https://doi.org/10.1109/ICAMechS.2019.8861646> (2019).

7. Loukas, E. P. *et al.* A machine learning approach for NILM based on odd harmonic current vectors. *Int. Conf. on Mod. Power Syst. (MPS)*, 1–6, <https://doi.org/10.1109/MP.2019.8759666> (2019).

8. Abbas, M. Z. *et al.* Non-intrusive status prediction of residential appliances using supervised learning techniques. *Int. Conf. on Energy Conserv. and Effic. (ICECE)*, 1–6, <https://doi.org/10.1109/ICECE58062.2023.10092511> (2023).

9. Nambiar, L., KumarGopal, V., Liu, Q. & Pradeep, R. Energy disaggregation for NILM applications using shallow and deep networks. *Int. Conf. for Convergence Technol. (I2CT)*, 1–6, <https://doi.org/10.1109/I2CT45611.2019.9033955> (2019).

10. Tapiador, M., de Diego-Otón, L., Hernández, A. & Nieto, R. Implementing a CNN in FPGA programmable logic for NILM Application. *Conf. on Des. of Circuits and Integr. Syst. (DCIS)*, 1–6, <https://doi.org/10.1109/DCIS58620.2023.10335989> (2023).

11. Kahl, M., Ul Haq, A., Kriebhbaumer, T. & Jacobsen, H. A. WHITED - A worldwide household and industry transient energy data set. *Int. Work. on Non-Intrusive Load Monit.*, <https://www.cs.cit.tum.de/dis/resources/whited/> (2016).

12. Kriebhbaumer, T. & Jacobsen, H. A. BLOND, a building-level office environment dataset of typical electrical appliances. *Sci. Data* **5**, 180048, <https://doi.org/10.1038/sdata.2018.48> (2018).

13. Kelly, J. & Knottenbelt, W. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Sci. Data* **2**, 150007, <https://doi.org/10.1038/sdata.2015.7> (2015).
14. Kolter, J. Z. & Johnson, M. J. REDD: A public data set for energy disaggregation research. *Work. on data mining applications sustainability (SustKDD)* **5**, 180048 (2011).
15. Navarro, V. M., Barragán, M., Hernández, A., Ureña, J. & Nieto, R. HIFDA - High-frequency electrical signals from household appliances dataset. *Zenodo*, <https://doi.org/10.5281/zenodo.14886758> (2025).
16. Digilent, Inc. Zybo Z7 board reference manual. Product Specification (2018).
17. AD7476A, 2.35 V to 5.25 V, 1 MSPS, 12-/10-/8-Bit ADCs in 6-Lead SC70, reference data sheet. Analog Devices (2016).
18. Navarro, V. M. *et al.* SoC architecture for high-frequency acquisition of household electric signals. *Int. Instrumentation Meas. Technol. Conf. (I2MTC)*, <https://doi.org/10.1109/I2MTC60896.2024.10560829> (2024).
19. MATLAB, manufacturer's webpage. MathWorks (2024).
20. 96001 clamp-on probe reference manual. Yokogawa (2024).
21. Mixed Signal Oscilloscopes MSO5000B, DPO5000B Series Datasheet. Tektronix (2018).
22. Python, manufacturer's webpage. Python (2024).
23. Anderson, K. *et al.* BLUED: A fully labeled public dataset for event-based non-intrusive load monitoring research. *Work. on data mining applications in sustainability (SustKDD)*, https://www.inferlab.org/wp-content/uploads/2012/08/2012_anderson_SustKDD.pdf (2011).
24. Picon, T. *et al.* COOLL: Controlled On/Off Loads Library, a public dataset of high-sampled electrical signals for appliance identification. *arXiv:1611.05803v1 [cs.OH]*, <https://doi.org/10.48550/arXiv.1611.05803> (2016).
25. Medico, R. *et al.* A voltage and current measurement dataset for plug load appliance identification in households. *Sci. Data* **7**, 49, <https://doi.org/10.1038/s41597-020-0396-4> (2020).
26. Ribeiro, M. and Pereira, L. and Quintal, F. & Nunes, N. SustDataED: A public dataset for electric energy disaggregation research. *Proc. ICT for Sustain.* <https://doi.org/10.2991/ict4s-16.2016.36> (2016).

Acknowledgements

This work was supported by the Spanish Ministry of Science, Innovation and Universities MCIN/AEI/10.13039/501100011033 (ALONE project, ref. TED2021-131773B-I00, INDRI project, ref. PID2021-122642OB-C41, and AGINPLACE project, ref. PID2023-146254OB-C43).

Author contributions

V.M. designed the SoC architecture and the voltage AFE, V.M. and M.B designed the current AFE and created the dataset, J.U. and A.H. assisted on and supervised the design of both AFEs, R.N. and A.H. supervised the design of the SoC architecture and analysed experimental results. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to V.M.N. or Á.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025