

An Embedded Deep Learning NILM System: A Year-Long Field Study in Real Houses

Simone Mari¹, Member, IEEE, Giovanni Bucci², Member, IEEE, Fabrizio Ciancetta³, Member, IEEE, Edoardo Fiorucci⁴, Senior Member, IEEE, and Andrea Fioravanti⁵, Member, IEEE

Abstract—Nonintrusive load monitoring (NILM) systems are used to identify the energy consumption patterns of individual devices in an electrical system, but broadening their market availability is a significant challenge. In this article, an NILM system using edge processing is proposed, in which energy consumption data are processed directly on the device installed at the monitored facility. Specifically, it uses a sequence-to-point approach based on a convolutional neural network (CNN) implemented on an Arm Cortex-M7 microcontroller. This article also reports the results of an extensive 12-month testing phase. The NILM system was installed in two real houses in central Italy to evaluate its installation and potential application in real-world scenarios. This study presents a promising solution that enables the widespread adoption of NILM systems by reducing their implementation cost and complexity and addresses the privacy concerns associated with cloud-based data processing. The results of our real-world testing provide compelling evidence of the potential of the proposed NILM system in various applications, including smart homes, building automation, and industrial energy management.

Index Terms—Convolutional neural network (CNN), deep learning (DL), edge deployment, energy disaggregation, nonintrusive load monitoring (NILM).

I. INTRODUCTION

ANALYSIS of the energy consumption of each device in an electrical system is a valuable tool for identifying inefficient or malfunctioning devices and implementing appropriate energy-saving measures. This type of analysis can be achieved using either intrusive load monitoring (ILM) systems or nonintrusive load monitoring (NILM) systems.

ILM systems require a transducer to be installed on each device to measure its energy consumption accurately. They provide precise results but can be challenging to implement due to space constraints, which make installing transduction and communication systems difficult.

On the contrary, NILM systems measure the total absorbed power and disaggregate the individual contributions of each

device using the specific energy consumption models (signatures) of different electrical loads. Although simpler than ILM systems from a hardware standpoint, NILM systems require suitable algorithms to identify individual device absorptions accurately [1], [2].

The choice of a monitoring solution depends on the installation requirements. For example, an NILM system can be placed in a switch box inside a property or at a distance from it. Local NILM systems can acquire voltages and currents, process them in real time, and display or store results on a remote server. In contrast, remote systems can only use data available in the cloud, which has limited measurement frequencies due to storage and data transmission limitations.

Broadening the availability of NILM systems is a significant challenge in the industry. Companies developing NILM solutions [3], [4], [5] focus on business-to-business (B2B) services rather than the business-to-consumer (B2C) sale of hardware. This is primarily because NILM technology is mainly used for energy management and monitoring in commercial and industrial settings, not in residences. These companies typically offer a wide range of services to businesses and organizations, such as energy audit, monitoring, and reporting and energy efficiency consulting. This approach allows them to collaborate with customers closely to gain a deep understanding of their specific energy usage patterns and provide tailored solutions for reducing energy consumption and costs. However, this makes it difficult to compare NILM systems proposed by researchers with commercially available NILM systems, as the former systems are not readily accessible to the general public.

Various techniques have been proposed in the literature, which differ in the sampling frequency of signals, the approach used to recognize devices, and the algorithmic technique used. Such techniques have evolved, and the ensuing state-of-the-art is interesting to analyze.

Since Hart [6] proposed the first NILM system in the 1980s, significant developments have occurred. Over the next two decades, research on this topic focused on finding new signatures capable of uniquely identifying devices and developing classifiers capable of providing indications based on these signatures. This type of approach often involves detecting events before their classification.

After an event is detected, the features (and then the signature) associated with the appliance that caused it are extracted. This approach can therefore be divided into three basic steps: event detection, feature extraction, and load identification. These techniques can be grouped within a so-called

Manuscript received 31 March 2023; revised 11 September 2023; accepted 10 October 2023. Date of publication 27 October 2023; date of current version 7 November 2023. The Associate Editor coordinating the review process was Dr. Manyun Huang. (Corresponding author: Simone Mari.)

The authors are with the Department of Industrial and Information Engineering and Economics, University of L'Aquila, AQ 67100 L'Aquila, Italy (e-mail: simone.mari@graduate.univaq.it; giovanni.bucci@univaq.it; fabrizio.ciancetta@univaq.it; edoardo.fiorucci@univaq.it; andrea.fioravanti@univaq.it).

Digital Object Identifier 10.1109/TIM.2023.3328085

event-based framework, where “event” means a change in the electrical parameters of the aggregate signal.

Dong et al. [7] proposed a system based on the detection of power signal events and the associated parameters, such as the active power range, reactive power range, harmonic content range, presence or absence of spikes, number of phases (single or double), and event searching time. The detected events are linked to appliance operational cycles through a clustering algorithm. For the approach in [8], detected power signal events are linked to appliances by minimizing the discrepancy of various considered parameters, such as effective voltage, effective current, active power, reactive power, apparent power, power factor, total harmonic distortion, and voltage–current ($V-I$) trajectory. Over the years, efforts have been directed toward identifying features extractable from events that allow for unique appliance discrimination and preferably remain constant across various operating states. Therefore, Teshome et al. [9] decomposed the aggregate current into two orthogonal components and defined $V-I$ trajectories with respect to each of the two components.

This direction was further pursued by Gupta et al. [10], who continuously measured the electromagnetic interference effects produced by appliances during their activation or deactivation. This system continuously processes the voltage measured at a domestic socket to obtain its Fourier transform. The processing result reveals more significant harmonic content after the occurrence of appliance switching operations. The appliance causing the harmonic content is identified using a k -nearest-neighbor classifier. Similarly, in [11], the effects of switching operations on the absorbed current signal were evaluated to realize galvanically isolated measurement systems.

These methods share the following limitations. First, the efficiency of event detection algorithms that struggle to find the right tradeoff between false positives and false negatives is a problem. The noise in aggregate power signals often hinders the identification of minor loads. Previously proposed event-based systems consistently perform excellently because of features that can be measured accurately following an event, such as those described above. However, these systems have poor generalization ability, such as during operation on unseen houses after training or the availability of data from other houses [12]. Moreover, most of these systems struggle with computational requirements, which grow significantly with the loads to be disaggregated and are thus often unusable in real-world scenarios [9].

Finally, many of these systems detect the activities of appliances but cannot provide a quantitative indication of their energy consumption. The ability to obtain information about the status of various appliances nonintrusively is important for many applications where NILM systems are implemented, such as smart home automation and ambient assisted living. NILM systems are used to obtain the energy consumption details of individual appliances and may be applied to recommendation systems.

In around 2010, a new trend emerged in the field of NILM systems research, namely, systems that do not require an initial event detection phase. The input of such a system is a window of samples of the aggregate signal (therefore time series data);

these samples are processed continuously, without waiting for event occurrence. These systems are generally known in the literature as nonevent-based systems. The concept of features or signatures is unnecessary, as the models use the aggregate power signal itself in place of features.

Kolter et al. [13] was among the first researchers to propose systems belonging to this category by applying discriminative sparse coding to energy disaggregation. This approach involves training discriminative models for each category of appliances. The individual energy consumption is obtained as a combination of basis functions multiplied by activations.

Hidden Markov models (HMMs) are widely used for nonevent-based NILM systems. In particular, Kim et al. [14] first proposed a factorial HMM (FHMM) in which the behavior of each appliance is modeled through an independent HMM. In this way, following a training phase, the FHMM can infer the hidden states of appliances from the aggregate consumption signal, thus separating individual consumption. In [15], HMMs are employed in a Bayesian framework that combines multiple models of individual appliances to form a general model of appliances. Bonfigli et al. [16] proposed a bivariate FHMM that uses both active and reactive power consumption data. Paradiso et al. [17] showed that additional information, such as house occupancy and appliance usage times, can improve the disaggregation result.

In 2015, Kelly and Knottenbelt [18] proposed the use of deep learning (DL) in nonevent-based NILM systems. Although ANNs had been used in previous work on NILM, they were designed as classifiers in the load identification phase of event-based systems. In [18], for the first time, the aggregate power signal was processed through an ANN using moving-window processing, and the problem was addressed as a blind source separation problem. The authors showed that these types of structures outperformed combinatorial optimization [13] and FHMM models [14], [15]. DL algorithms have the interesting advantage of not requiring the manual extraction of features, such as individual appliance consumption, ON–OFF state transition, and operation duration, which are instead automatically learned using ANNs.

In recent years, Zhang et al. [19] made a concrete contribution to this field by developing a convolutional neural network (CNN) structure called the sequence-to-point approach.

The sequence-to-point approach is a framework, and it has been improved using bidirectional layers [20] and self-attention mechanisms [21]. However, the complexity of these models incurs significant computational costs during the training and inference phases. In particular, the performance gains of these models are more aligned with increases in training data rather than an increase in model complexity.

The availability of training data is a primary challenge in developing effective, efficient NILM systems. DL models for nonevent-based NILM systems require so-called labeled datasets, namely, temporal sequences of aggregate power (system input) and the corresponding temporal sequences of appliance-level power (system output). At present, the only dataset that provides a satisfactory amount of data is the REFIT dataset [22]. The model proposed in [19] achieves

an excellent balance between complexity and performance, outperforming both FHMM and previous DL models. This algorithm has been tested on several public datasets; in particular, its generalization capability was tested in [23] by providing a system with data on homes belonging to datasets other than the one presented in training. Thus, in this work, DL models are used to demonstrate the feasibility of a solution based on a small, low-power microcontroller for real-time energy consumption monitoring.

The potential of these systems cannot be fully evaluated using existing datasets. The variability of submonitored loads in different homes within the same dataset, along with the fixed sampling frequencies, restricts the evaluation process. Moreover, a major metrological challenge is the limited knowledge about, or even the impossibility of, determining the uncertainty associated with the measurement of quantities such as current and power in these datasets.

Cloud-based energy disaggregation systems generally entail sending energy consumption data to a remote server, where they are processed and made available for user access from anywhere via a Web browser. These solutions are convenient because they do not require dedicated hardware to process data, but they have some disadvantages. For example, sending data to a remote server may be affected by service interruptions because it involves some latency and depends on the quality of the Internet connection. In addition, processing data remotely involves high data management and security complexities.

In contrast, in an edge solution, such as the one proposed in this article, energy consumption data are processed directly on the device installed at the monitored facility. The processed data can be accessed locally or transmitted to a remote server via the Internet. This approach has several advantages over cloud solutions. For example, edge solutions can process data in real time, eliminating the latency present in cloud solutions. They are also easier to install and manage, as they do not require a reliable Internet connection. This also makes these types of systems cheaper than their cloud counterparts despite requiring more knowledge for managing real-time systems. In addition, data access is limited only to authorized devices at the monitoring site, so no costs are incurred for transmitting data or maintaining a remote server. In summary, edge solutions for NILM offer greater reliability, privacy, security, and convenience than cloud solutions.

This article, an extended version of [24], offers the following contributions.

- 1) An edge solution based on a small, low-power microcontroller is developed using the state-of-the-art model proposed in [19] for real-time energy disaggregation. This solution can be implemented within a home without requiring prior knowledge of the electrical system or connected loads.
- 2) A comprehensive architecture is outlined for a metering system suitable for capturing and monitoring various electrical measurements generated by household appliances. This system captures the aggregate consumption of the electrical system and each appliance and transmits the data wirelessly to a central concentrator. The measurements are collected and processed using a certified

class 0.2 single-phase onboard meter, which ensures the reliability of the metrological data.

- 3) The system's ability to evaluate and characterize NILM system performance is demonstrated using reliable data.
- 4) Results from the implementation of the system in two Italian households (for six months each) are presented. The experimental phase lasted one year, from January 2022 to February 2023 (inclusive).

II. NILM AS A NONLINEAR REGRESSION PROBLEM

Nonevent-based DL models are implemented in the proposed system. These models are based on a supervised learning mode. This category of systems involves showing both input samples and expected output samples during the training phase. These samples constitute labeled datasets. For nonevent-based NILM systems, a dataset is labeled when both the time sequences of the aggregate absorbed power signal and the time sequences of the power absorbed by individual appliances (or the time sequences of their ON/OFF state) are provided. The data pair (X_t, Y_t) is therefore available, where X_t indicates the reading of the aggregate absorbed power and Y_t indicates the reading of the absorbed power at the appliance level. A supervised nonevent-based model aims to learn the relationship between X and Y . Thus, the problem formulated in (1) can be approached as a nonlinear regression problem

$$X = f(Y). \quad (1)$$

A sequence-to-point approach is used in this work. Given an aggregate power reading window, an ANN is trained for exclusively predicting the midpoint of an appliance-level power reading window. In this way, the overall time series is obtained through sliding-window processing.

In this approach, with t denoting a generic moment in time and W denoting the window length, the power absorbed by the monitored appliance at the central point of the window $Y_{t+(W/2)}$ is estimated for each window $X_{t:t+W-1}$.

This approach is based on the premise that $Y_{t+(W/2)}$ can be depicted as a nonlinear regression of the input window $X_{t:t+W-1}$. Consequently, the estimated power consumed by a household appliance at any given time must be influenced not only by past power readings but also by future ones.

Several models used to depict the relationship f between the X and Y series have significant limitations. Models, such as FHMMs, can be considerably influenced by factors, such as the existence of unknown appliances, base loads, and noise. The implementation of these models therefore requires the explicit modeling of these variables.

In contrast, the use of DL models does not need their explicit modeling because these models separate the consumption profile of an appliance by treating everything else as background.

The necessary features, such as the individual appliance consumption, ON–OFF state transition, and operation duration, are learned automatically by the neural network; in other words, manual extraction is unnecessary.

This process does not require the use of specific information about consumption sources or their power profiles. A DL model can generalize to new consumption sources

that may appear over time. Thus, this model can be used to separate independent signals associated with different consumption sources in various situations without needing specific information about the consumption sources or their power profiles.

A. Model Configuration

The implemented sequence-to-point model consists of convolutional layers, so it is a CNN. CNNs are ANNs designed to exploit the inherent properties of certain 2-D data structures in which spatially close elements are correlated (local connectivity).

The same process can be applied to 1-D data sequences. A 1-D CNN effectively derives features from a fixed-length segment of a dataset; the location of a feature in the segment is unimportant. 1-D, 2-D, and 3-D CNNs work in the same way. Their difference lies in the structure of the input data and how the filter and the convolution kernel act on the data.

In this work, a suitable 1-D CNN is implemented to process the time sequence of the aggregate power signal and predict the midpoint of the time sequence of the absorbed power at the appliance level. Since the only dimension in a time series is time, the kernel flows in only one direction.

Let X be a time series of length N , represented as a 1-D vector $X = [x_1, x_2, \dots, x_N]$. A filter (or kernel) is a small window of values that flow along the time series to perform convolution. Let F be the filters to be used in the convolutional layer. Each filter has a length F_{len} .

The convolution operation between a filter f and a segment of the time series X is defined as follows:

$$(F * X)_i = \sum_{j=1}^{F_{len}} f_j \cdot x_{i+j-1} \quad (2)$$

where i is the index at which the filter is running along the time series and j is the index within the filter.

For each kernel, a 1-D convolutional layer extracts 1-D local patches (subsequences) of the original sequence through sliding-window processing. It then applies identical transformations to these patches. For each kernel, since the same transformation is applied to each patch, a pattern learned at a certain position in the sequence can also be recognized at a different position. This makes the 1-D CNN translation invariant (for temporal translations).

The proposed 1-D CNN involves the following building blocks.

- 1) *Input Layer*: In this layer, data are preprocessed through a sliding-window technique such that each input contains 599 samples of the aggregate active power reading. The data are sampled at a sampling rate of 1/8 Hz, so each input sequence covers a time interval of 4792 s.
- 2) *First 1-D Convolutional Layer*: In the first 1-D convolutional layer, 30 filters (or kernels) of length 10 (kernel size) are defined, which allows the ANN to learn 30 different features. In these layers, the step (stride) at which the kernel moves along the input sequence is also defined. The number of filters in each layer is a hyperparameter chosen by the programmer.

The features learned by, and thus the weights assigned to, each filter are the result of training. In this work, a stride equal to 1 is defined. A kernel of size 10 moving along an input sequence of size 599 with a stride equal to 1 produces an output sequence of 590 elements. However, a padding process is used to fill the input sequence with a certain number of zeros at the beginning and end of the sequence to output the same number of elements as the input sequence (599). The output of the first 1-D convolutional layer is thus an array of 30×599 neurons, that is, 30 output sequences resulting from applying 30 filters to the input sequence.

- 3) *Second 1-D Convolutional Layer*: The result of the first 1-D convolutional layer is directly fed into the second layer. Thirty filters are also defined to be trained in this layer. Although the input of this layer is now 2-D instead of 1-D (a 30×599 matrix), the transformation applied by this layer is still 1-D convolution. Therefore, the kernels move along a single (temporal) direction. Eight dimensions are chosen for the kernels in this layer. The kernels in this case are no longer vectors of a length equal to the imposed kernel size (1×8) but a 30×8 matrix.
- 4) *Third, Fourth, and Fifth 1-D Convolutional Layers*: Three more 1-D convolutional layers are added to learn higher level features. The kernel numbers of layers 2–4 are 40, 50, and 50, respectively, and the kernel sizes are 6, 5, and 5, respectively. The stride is kept equal to 1 for these three additional layers. The padding process is provided for all convolutional layers. All neurons in the five 1-D convolutional layers predict using the ReLU activation function.
- 5) *Flatten Layer*: This layer transforms the entire output matrix of the fifth and final 1-D convolutional layer into a single vector ($1 \times 29\,950$).
- 6) *Dense Layer*: This layer reduces the output dimension from 29 950 (from the flatten layer) to 1024. The neurons in this layer also use the ReLU activation function.
- 7) *Output Layer*: This layer provides the value of the midpoint of the appliance power reading and thus has a single neuron. This neuron receives the weighted sum of the 1024 output elements from the previous layer as its input. It then applies a linear activation function, which is effectively the same as applying no activation function at all.

Fig. 1 shows the overall structure of the implemented 1-D CNN.

B. Training Settings

Over the past decade, numerous public datasets have been released to enable researchers to assess and compare the performance of their NILM algorithms with that of new proposals. These datasets differ in sampling frequency, number of monitored homes, and availability of submonitored data (measurements obtained directly from household appliances or other loads). In some cases, submonitored data are accessible but not synchronized with the aggregate power measurement.

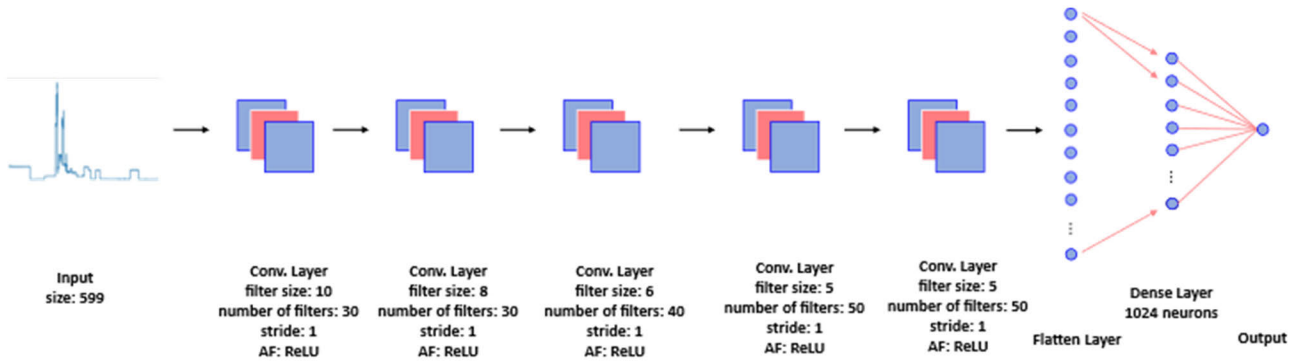


Fig. 1. Sequence-to-point approach proposed by Zhang et al. [19].

A dataset that fulfills the following requirements can be used to evaluate the performance of DL-based NILM systems:

- 1) a sampling frequency of 1 Hz or more, which enables the assessment of the impact of sampling frequency on NILM system performance through the ability to obtain the lowest frequencies, starting from the original data;
- 2) synchronous measurements of aggregate and appliance-level values of all quantities;
- 3) a sufficient number of houses (at least >1);
- 4) a sufficient number of classes of household appliances;
- 5) a sufficiently long acquisition period, which guarantees a sufficient amount of training data.

The most complete datasets in this regard are ECO [25] and ENERTALK [26]. ECO provides aggregate measurements of active power, voltage, current, and power factor at 1 Hz for six homes and 18 appliance classes; however, at the appliance level, only active power is available. ENERTALK provides both aggregate and appliance-level active and reactive power measurements at 15 Hz for 22 homes but has a small number of measured appliance classes.

The primary datasets used to train DL-based NILM systems and their distinguishing characteristics are given as follows.

- 1) The reference energy disaggregation dataset (REDD) [27] offers high- and low-frequency data, with voltage and current measurements at 15 kHz and 1 Hz, respectively.
- 2) The U.K. domestic appliance-level electricity (U.K.-DALE) dataset [28] comprises aggregate power and appliance-level data at 1/6 Hz for five houses.
- 3) The REFIT dataset [22] provides aggregate and appliance-level power measurements for 21 U.K. homes at a sampling rate of 1/8 Hz. However, the most common appliances used in NILM systems may not be available in all 21 homes.

REFIT is the largest among these datasets. Therefore, although its low sampling rate prevents the assessment of its influence on performance, its wide availability of examples allows for a sufficiently robust NILM system. Therefore, in the current study, the training phase is conducted using the REFIT dataset. The proposed system uses a CNN to identify and recognize individual household appliances. The focus is on developing

TABLE I
HOUSES FROM THE REFIT DATASET USED FOR THE TRAINING PHASE

	Training set		Validation set	
	Houses	Samples	Houses	Samples
Dishwasher	5, 7, 9, 13, 16, 20	35984834	18	5007721
Washing machine	2, 5, 7, 8, 9, 15, 16, 17	49588126	18	5007721
Fridge	2, 5, 9, 15	25559502	12	5859544

an NILM system capable of separating the electrical power consumption of the three major household loads (dishwasher, washing machine, and fridge) from the total consumption. This is because these loads are commonly targeted by NILM system developers for disaggregation in the market [5].

Therefore, each DL model should be trained using data from one household appliance. The REFIT dataset contains measured consumption data for 21 homes, but not all of them contain data for the three abovementioned loads. Table I shows the homes used to train each model and the total number of samples available.

Table I also shows the houses and the relative number of samples used as the validation set. The validation set plays a crucial role in the training process of a DL model, as it enables the evaluation of the model's ability to generalize to unseen data. Moreover, the model's performance during training can be assessed continuously using the validation set by monitoring it at the end of each epoch. This helps avoid a common problem in DL called overfitting [29], where the model becomes too tightly fit to the training data, hindering its ability to perform well on new data. For each appliance, validation is performed using the entire dataset of a single house, according to the methodology established in [23], during the model presentation.

The data for training are preprocessed before being used for training. Preprocessing involves normalizing the data using the following equation:

$$x'_k = \frac{x_k - \bar{x}}{\sigma_x} \quad (3)$$

where x_k is the k th sample, \bar{x} is the mean of the aggregate or appliance-level power reading, and σ_x is the standard deviation of the aggregate or appliance-level power reading. After the

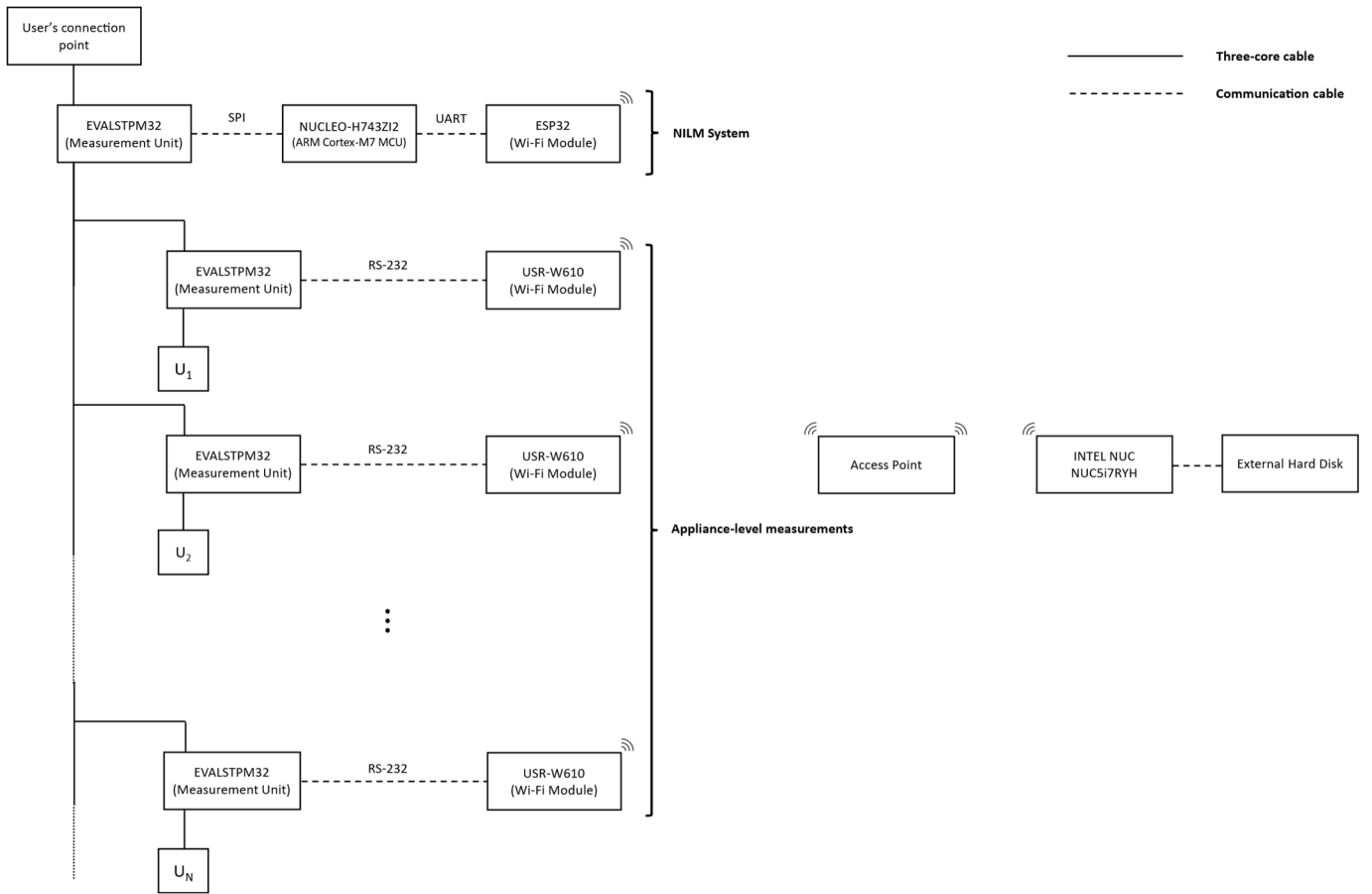


Fig. 2. Proposed architecture for the installation and performance evaluation of the NILM system.

data are normalized, they can be fed into the models for training.

The CNNs are implemented on a desktop computer (64-bit Windows 10 operating system) using TensorFlow [30] for model development and training. The adopted cost function, implemented in TensorFlow, is the mean squared error (MSE) (4) and is applied to each batch during the training phase

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad (4)$$

where Y_i is the actual value, \hat{Y}_i is the predicted value, and N is the number of processed samples.

The parameters of each CNN are updated after one iteration of every batch of data. The batch size chosen for training is 1000, so the neural network parameters are updated every 1000 samples. Each model is trained for ten epochs.

The Adam optimizer [31] is used to drive the model training process. Adam is a highly efficient optimization algorithm that is widely adopted by DL practitioners. It operates by using an adaptive learning rate that is dynamically adjusted based on the mean and variance of the gradients of the cost function according to the weights of the ANN. During the training phase, Adam continuously monitors the mean and variance of the gradients, thus effectively fine-tuning the learning rate. By increasing (decreasing) the learning

rate for slowly (rapidly) changing weights, Adam effectively eliminates oscillations and accelerates convergence toward the global minimum solution.

III. ARCHITECTURE OF THE PROPOSED SYSTEM

The proposed architecture is based on a distributed data acquisition system and communicates through a Wi-Fi network.

The first developed device (see the NILM system in Fig. 2) measures the aggregate active power and disaggregates it using the model described in Section II. This system also measures reactive power, voltage, and current, but these are not used for the performance evaluation of this NILM system. The NILM system consists of the following:

- 1) a measurement unit (an EVALSTPM32 board);
- 2) a processing unit (a NUCLEO-H743ZI2 board);
- 3) an ESP32 Wi-Fi module for connection to a wireless local area network (WLAN).

The main device is placed immediately downstream of the general power meter located at the user's connection point to ensure that the aggregate active power can be measured.

In addition, ad hoc power meters are developed, and each of them consists of the following:

- 1) a measuring unit (an EVALSTPM32 board);
- 2) a USB-W610 Wi-Fi module for connection to the WLAN.

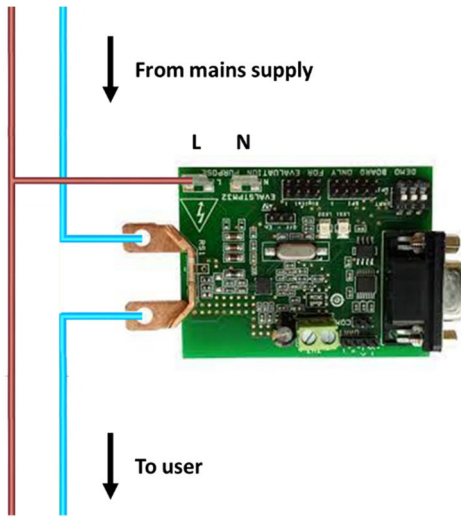


Fig. 3. Pin connections of the EVALSTPM32 board for voltage and current measurement.

These appliance-level power meters enable the measurement of electrical quantities related to the operation of individual household appliances (active power, reactive power, root-mean-square (rms) voltage, and rms current).

The established WLAN has a star topology; both the NILM system and the appliance-level power meters are connected to a concentrator via the Wi-Fi network. Specifically, the implemented communication and data archiving infrastructure, concisely schematized in the right part of Fig. 2, consists of the following:

- 1) an access point, which creates the wireless network where all nodes are connected;
- 2) an Intel NUC NUC5i7RYH system, which is the main part of the network and enables the management of the nodes connected to the network using a Python script, downloads the measurement data, and stores them in a MySQL database;
- 3) an external hard disk for storing the MySQL database.

Furthermore, a Web server is developed based on Node-RED to check the whole system and plot few-point measurement data.

The proposed system is shown in Fig. 2.

For a clear, concise understanding of the connections involved in the schematic in Fig. 2, Fig. 3 shows a detailed view of the EVALSTPM32 board's phase and neutral conductor connections for voltage and current measurements, respectively. In the voltage measurement circuit, the N terminal is deliberately left unconnected to the neutral conductor. This is because the shunt for current measurement is already placed at the same potential as the N terminal, which eliminates the need for a separate connection.

A. NILM System and Appliance-Level Power Meters

The pretrained DL models (Section II) are uploaded to the NUCLEO-H743ZI2 board. This is a high-performance microcontroller from STMicroelectronics that is based on the Arm Cortex-M7 architecture. This board is part of

the STM32H7 series and offers advanced features, including a floating-point unit, hardware encryption, and up to 2 MB of flash memory. The NUCLEO-H743ZI2 board also includes various communication interfaces, including Ethernet, USB, CAN, and various serial ports. It is designed for use in a wide range of applications, including industrial automation, motor control, and consumer electronics. This microcontroller can be programmed using different integrated development environments (IDEs) and supports a range of development tools and software libraries provided by STMicroelectronics.

The IDE used in this work is STM32CubeIDE. The models are implemented through X-CUBE-AI, an expansion package dedicated to AI projects running on STM32 Arm Cortex-M-based MCUs. The X-Cube-AI core engine, schematically shown in Fig. 4, offers an NN mapping tool to create and implement pretrained DL models for embedded systems with limited hardware resources.

The generated STM32 NN library can be integrated into an IDE project or a makefile-based build system. The code generator quantizes weights, biases, and activations from floating-point precision to 8-bit precision and maps them onto a specialized C implementation for supported kernels. This technique aims to reduce the model size, improve CPU and hardware accelerator latency, and reduce power consumption without compromising model accuracy.

A validation mechanism is provided to compare the accuracy of the generated model with that of the uploaded DL model using the same input tensors (fixed random inputs or custom dataset). The scheme of the validation engine is shown in Fig. 5.

The NUCLEO-H743ZI2 receives input data for models via the EVALSTPM32, which is the measurement unit. The EVALSTPM32 is a single-phase meter with a class 0.2 rating that uses a shunt transducer to acquire power line currents. This board offers SPI/UART pins that allow for interfacing with a microcontroller during application development. The EVALSTPM32 calculates the rms values of voltage and current; instantaneous voltage and current waveforms; and active, reactive, and apparent power and energies. This board is a mixed-signal IC family that comprises an analog section and a digital section. The analog section incorporates two low-noise, low-offset programmable-gain amplifiers, and up to four second-order 24-bit sigma-delta ADCs, among other components. The digital section consists of a digital filtering stage, a hardwired DSP, a digital front end, and a serial communication interface. The board's power data registers supply filtered or instantaneous measurements, with a wide-band bandwidth of 3.6 kHz, which can be used to measure up to the 72nd harmonic of a 50-Hz signal.

An SPI is used to connect the EVALSTPM32 to the NUCLEO-H743ZI2. This interface is configured with a clock frequency of 10 MHz and full-duplex transmission mode. The SPI is configured on the NUCLEO-H743ZI2 using the STM32CubeIDE HAL library. Once the SPI is configured, the firmware on the NUCLEO-H743ZI2 requests the active power, reactive power, rms current, and rms voltage readings from the EVALSTPM32 using the SPI protocol.

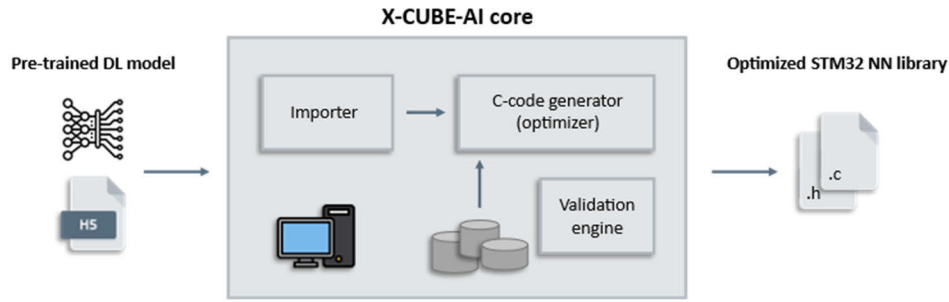


Fig. 4. X-Cube AI core engine.

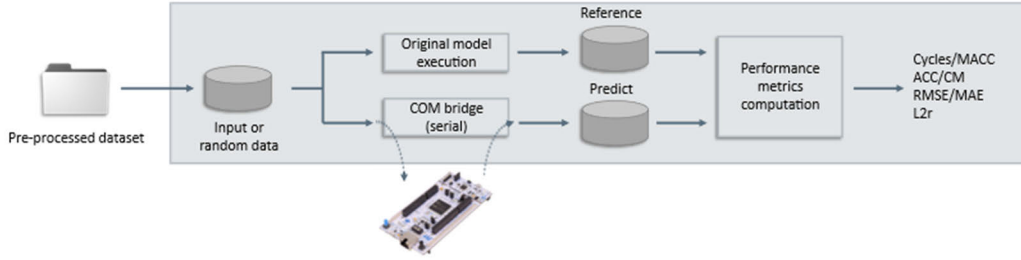


Fig. 5. AI validation firmware embedding the tested C model.

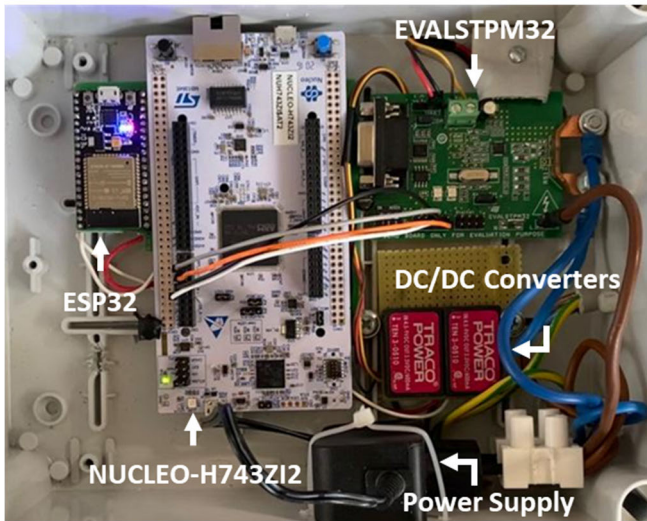


Fig. 6. Implemented NILM system.

Communication occurs when the NUCLEO-H743ZI2 sends the request and the EVALSTPM32 replies with the requested data. Once the data are read, the NUCLEO-H743ZI2 processes the active power data using the abovementioned models to obtain appliance-level active power information. After the data are processed, the NUCLEO-H743ZI2 uses the ESP32 module to connect to the Wi-Fi network. The NUCLEO-H743ZI2 board then sends the data to the concentrator using a TCP-based response–response protocol. The ESP32 module is configured as a client, and communication occurs asynchronously. The response–response protocol allows for bidirectional communication between the NUCLEO-H743ZI2 and the concentrator, thus enabling the sending of data in both directions. A photo of the described NILM system is in Fig. 6.

As schematized in Fig. 2, each appliance-level power meter consists of one EVALSTPM32 measurement unit connected to the WLAN via a USR-W610 converter. The USR-W610 is a serial-to-Wi-Fi and serial-to-Ethernet converter capable of bidirectional transparent transmission between RS-232/RS-485 and Ethernet/Wi-Fi. It allows the assignment of work details and the transparent transmission of serial data and TCP/IP data packets via a converter.

The USR-W610 can open a TCP socket as a server or a client. Each wireless module is set to receive TCP messages through the TCP socket’s server side to ensure good implementation of the proposed WLAN network architecture (Fig. 2). The TCP data are then converted by the wireless module for the RS-232 interface. The EVALSTPM32 board adopts a request–response serial communication handshake, so the bidirectional TCP server socket connection can address the requirement of the communication protocol. The modules are addressed by a static IP address stored in the USR-W610 network configuration to ensure point-to-point communication. Fig. 7 shows one of the appliance-level power meters, and Fig. 8 shows the overall installed system.

B. Central Concentrator and Web Server

The main part of the system is the Intel NUC NUC5i7RYH system, which manages the nodes connected to the network using a Python script, downloads measurement data, and stores them in the MySQL database. The system is based on a fifth-generation Intel Core i7-5557U processor (3.1–3.4 GHz Turbo Dual core, 4 MB cache, 28 W TDP). The Intel NUC supports Intel’s hyperthreading technology and has 16-GB DDR3 memory. The OS used to manage the system is Windows 7.

The application program is written in Python [32], an object-oriented programming language suitable for developing

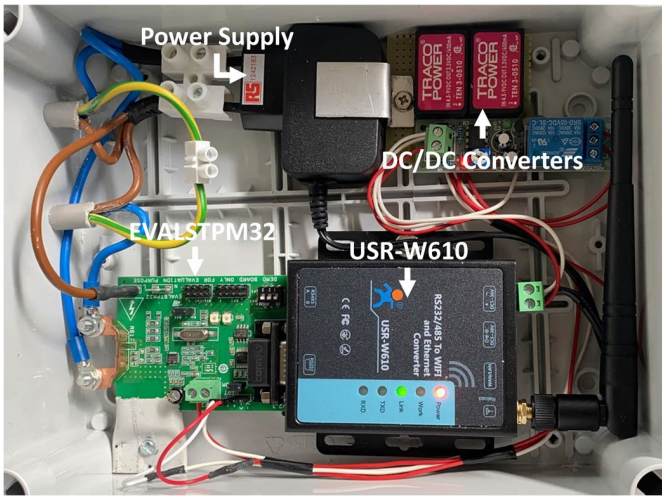


Fig. 7. Appliance-level power meters.



Fig. 8. Overall system installed in the two houses observed during the test phase.

distributed applications, scripting, numerical computing, and system testing.

The following tasks are performed by the main program.

- 1) Establish a connection with the NILM system and all appliance-level power meters.
- 2) Send the read request to each node connected to the network and receive the data.
- 3) Store the data in the MySQL database.

Each power meter is accessed through the multithreading approach. In particular, the Python script manages a number of threads equal to the number of power meters to ensure that all communication ends within the sampling time (8 s). Ultimately, all acquired data are stored in a MySQL record.

The timestamp is synchronized using an NTP protocol. The main program is shown in Algorithm 1.

The Web server allows data supervision during monitoring. The interface is developed in Node-RED [33], a programming tool for wiring together hardware devices, APIs, and online services. For each monitored load, this user interface shows four different graphs for rms voltage, rms current, active power, and reactive power. The graphs show a time window

Algorithm 1 Main Program

```

connect to host
while True do
  take current time
  read processed and unprocessed data from the NILM
  system
  for each appliance-level power meter do
    read Vrms, Irms, P and Q
    store data in a MySQL database
    if sample time is elapsed
      repeat
    else
      wait

```

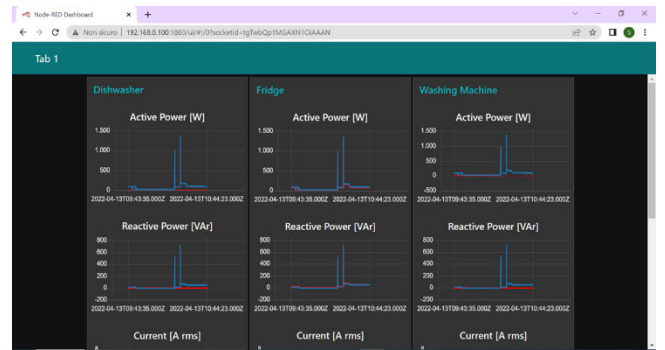


Fig. 9. Web server of the system developed in Node-RED.

corresponding to the last hour of acquisition (hence 450 points considering the sampling frequency of 1/8 Hz); they can be used to compare the aggregate quantities with the appliance-level quantities. The data are obtained directly from the MySQL database. Fig. 9 shows the Web server.

C. Calibration of the Measurement Unit

A series of measurements is initially performed using the power meters to ensure that their measured values are metrologically valid. A Fluke 6100 A electrical power standard is used for this initial verification process. This is a highly accurate voltage and current standard that provides a reference for measuring electrical power and energy. After the power meters are verified, more detailed measurements are obtained using a HARMONICS-1000, a single-phase measuring system for harmonics and flicker. This system enables the simultaneous generation of load voltage and current and thus a highly detailed assessment of the accuracy and reliability of power meters under different operating conditions.

The procedure used in the following calibration is based on the approach presented in the ISO publication “Guide to the Expression of Uncertainty in Measurement” [34].

In this calibration, the transducer response is provided as a function of the applied quantity in (5) to derive the K_i coefficients from a least-squares fit to the calibration data

$$R = K_0 + \sum_{i=1}^N K_i V^i \quad (5)$$



Fig. 10. Setup for calibrating the power meters.

where R is the transducer response, V is the applied quantity, K_i is the coefficients characterizing the transducer, and N is the polynomial order of the calibration function. A polynomial order of 1, which is equivalent to linear interpolation, is chosen in this work. In the measurements, a variable resistor of known magnitude is utilized as the load. Data are acquired over resistances of 23–136 Ω , and the load is powered at voltages of 220–240 V. In total, the voltage, current, and active power are acquired at 15 different points. Fig. 10 shows the calibration setup used in the measurement process. The transducer response R_j is measured for each input voltage V_j , current A_j , and active power P_j . The K_i coefficients in (5) are calculated using the least-squares fit of the measurement sets (V_j, R_j^V) , (A_j, R_j^A) , and (P_j, R_j^P) .

The uncertainty associated with the deviation of the measured data from the fit curve is represented by the standard deviation u_r as follows:

$$u_r^2 = \frac{\sum_{j=1}^N (d_j^2)}{(n - m)} \quad (6)$$

where d_j is the difference between the transducer inputs and the responses calculated using (5), n is the number of individual measurements in the calibration measurement set, and m is the order of the polynomial plus one.

The coefficient of variance c_V is calculated using (7) to account for measurement variability

$$c_V = \frac{u_r}{\mu} \quad (7)$$

where μ represents the mean of the set of measurements. The calibration results indicate that the coefficients of variance of voltage, current, and active power are 0.11%, 0.13%, and 0.87%, respectively.

The noise in the input channels is characterized to further ensure the reliability and accuracy of the measurement system. The intrinsic noise of acquisition and measurement systems can compromise load identification and measurement accuracy. The noise characterization results indicate that the level and type of noise in the input channels are below the resolution of the ADC measurement system. Therefore, noise is not a significant factor in the measurements and does not affect the accuracy of the results.

Overall, these calibration and noise characterization procedures ensure the reliability and accuracy of the measured values and provide a solid foundation for the subsequent data analysis and interpretation.

IV. EXPERIMENTAL RESULTS

As part of the development phase of the proposed system, the performance of the implemented prototype in real-world scenarios was tested. The NILM system and the appliance-level power meters were installed in two houses in central Italy—one in Marche and the other in Abruzzo.

From February 2022 to January 2023, the system collected, processed, and archived the data of each house (six consecutive months for each house). This extensive testing phase ensured that the system's performance was thoroughly analyzed and its effectiveness verified under various operating conditions.

The system is designed to acquire data at a sampling rate of 1/8 Hz. The DL algorithm, implemented using the NUCLEO-H743ZI2 board, processes time windows of 599 points, which are equivalent to 4792 s or approximately 80 min. The algorithm starts processing when a new active power measurement is available (every 8 s).

An important feature of the implemented system is that it can operate without requiring any prior knowledge of the appliances to be monitored. For evaluating the performance of the system, its ability to recognize the absorption patterns of the monitored appliances was initially analyzed, starting from the aggregate energy consumption. Various load profiles of the appliances were considered, and the responses provided by the system were examined.

The obtained results, plotted in Figs. 11 and 12, show the four consumption patterns of the monitored dishwashers and washing machines against the acquisition time. The consumption patterns depended on the various work cycles set for the appliances, which were considered during the analysis to gain valuable insights into the system's performance.

The cycle considerations for the dishwashers and washing machines could not be applied to the monitored fridges. Instead, their daily consumption was analyzed, as presented in Fig. 13, which illustrates four examples of daily consumption patterns. This figure provides crucial information about the system's ability to recognize the daily consumption patterns of fridges accurately. The left and right graphs in Fig. 13 refer to the first and second houses, respectively. In particular, the two fridges had different consumption patterns, but the DL algorithm adapted without prior knowledge of these appliances, demonstrating its excellent flexibility and adaptability.

These figures depict the system's ability to recognize the distinct consumption patterns of different

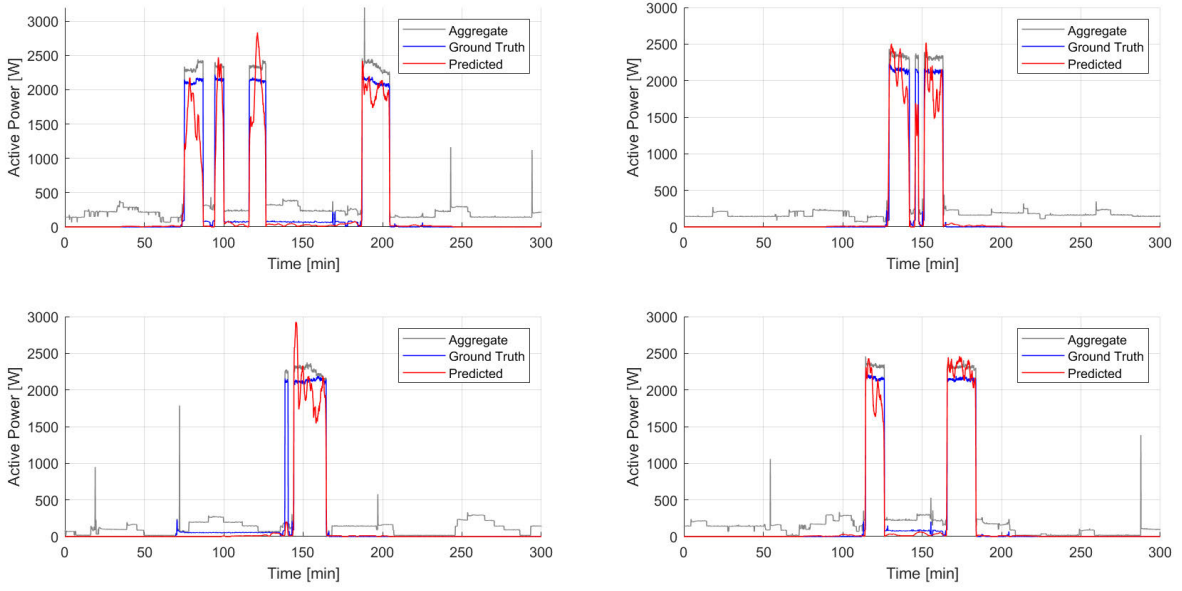


Fig. 11. Load profiles of dishwashers for different work cycles and corresponding disaggregation result.

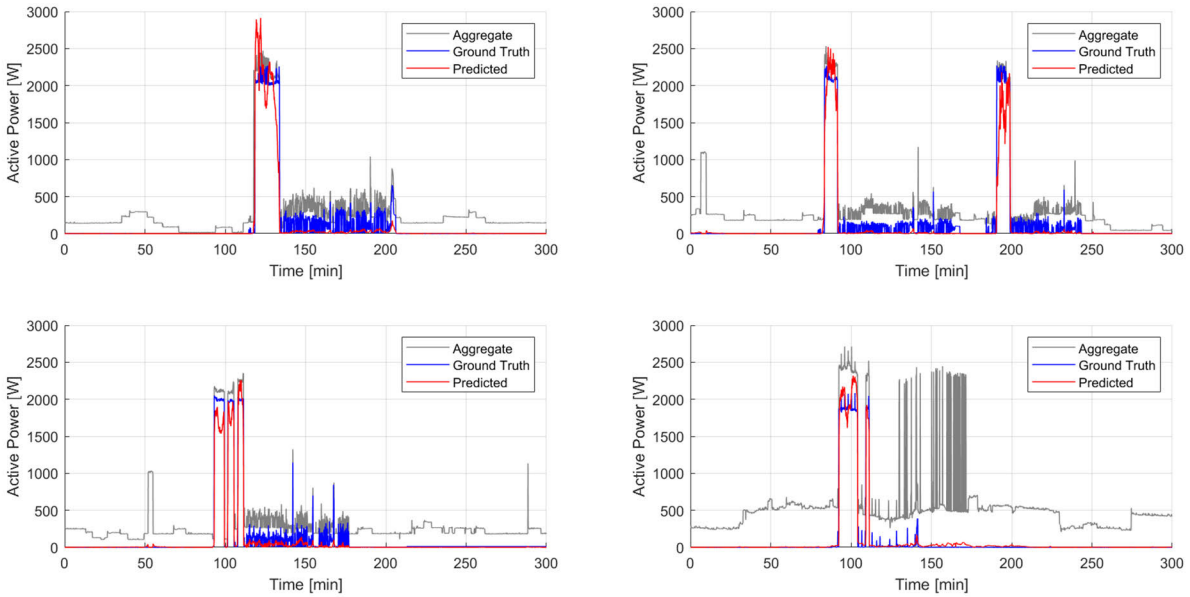


Fig. 12. Load profiles of washing machines for different work cycles and corresponding disaggregation result.

appliances accurately. System accuracy and reliability are qualitatively evident from their analysis.

For a quantitative evaluation of system performance, the accuracy of the system in estimating the energy consumption of the appliances in each work cycle was determined using metrics typically adopted for NILM systems.

The percentage relative error in estimating energy consumption, often called the signal aggregate error (SAE) in NILM literature, was computed using the following equation:

$$SAE = \frac{\hat{E}_c - E_c}{E_c} \cdot 100 \quad (8)$$

where \hat{E}_c indicates the estimated energy consumption per work cycle and E_c indicates the actual energy consumption

per work cycle, as measured using the appliance-level power meters. For each appliance, 50 work cycles were considered (25 for each house). Fig. 14 shows the percentage relative error (or SAE) in the estimation of energy consumption per work cycle. For the same reasons explained above, in the case of the fridges, daily consumption was considered work cycles.

The results are summarized in Table II, which presents the absolute and percentage relative errors in estimating the total energy consumed in the 50 work cycles (50 days for the fridges). These metrics offer a comprehensive evaluation of the system's accuracy in estimating energy consumption.

The analysis of the percentage relative error in (5) was extended to a half-yearly basis. An additional metric was

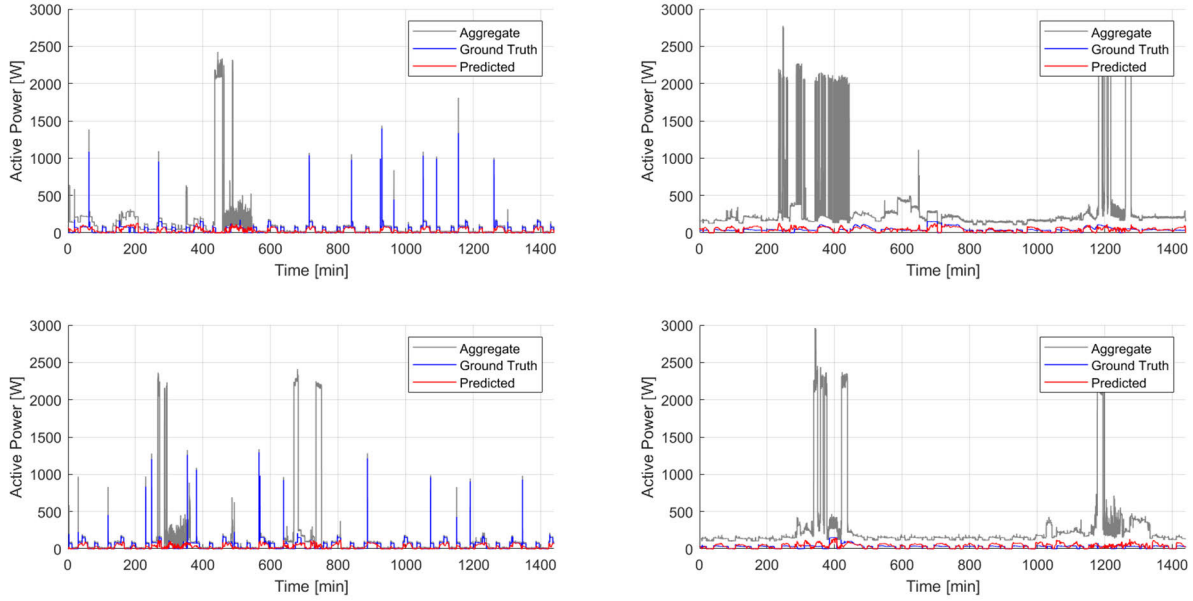


Fig. 13. Daily consumption patterns of fridges and corresponding disaggregation results.

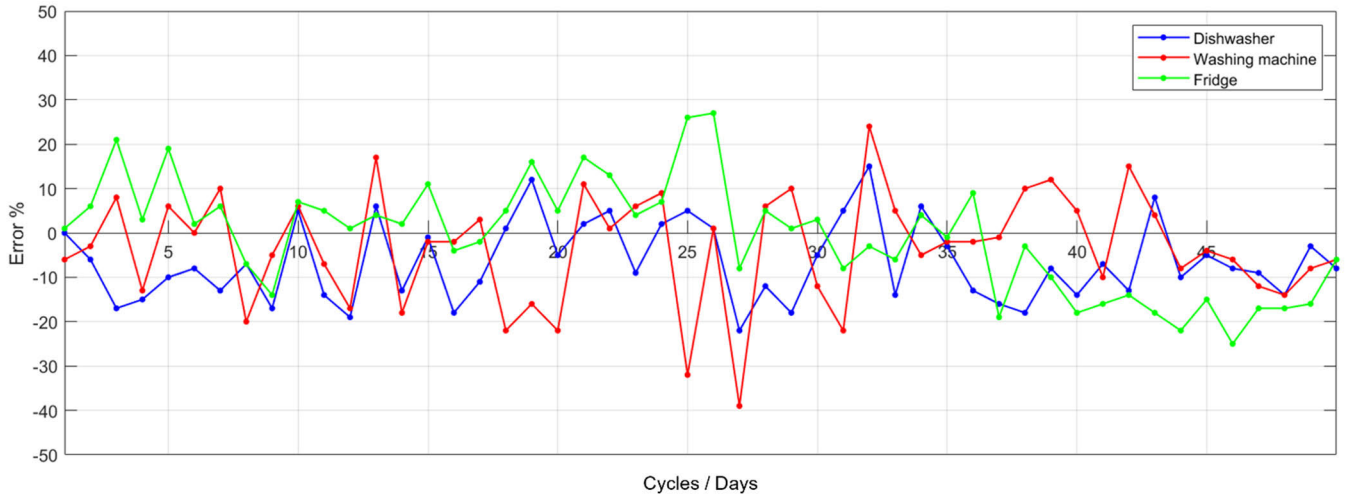


Fig. 14. Percentage relative error (or SAE) in the estimation of energy consumption per work cycle (for dishwasher and washing machine) and daily energy consumption (for fridge).

TABLE II
ABSOLUTE AND PERCENTAGE RELATIVE ERRORS IN
ESTIMATING TOTAL ENERGY CONSUMPTION

	Absolute error	SAE
Dishwasher	-4558,076 Wh	-7 %
Washing machine	-1166,067 Wh	-5 %
Fridge	-687,756 Wh	-2 %

considered to compare the proposed system with those in the literature.

The mean absolute error (MAE) was computed to evaluate the accuracy of the model's predictions. The MAE, computed using (9), measures the average absolute difference between the actual and estimated active power at each time step

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |\hat{p}_t - p_t| \quad (9)$$

where \hat{p}_t is the estimated power value at time step t , p_t is the measured power value at time step t , and T is the time interval considered.

Table III shows the percentage of relative errors and MAEs obtained for the different appliances over the entire dataset for each house (half-year basis). For a thorough explanation of the results in Table III, Table IV provides additional metrics obtained from a related study [23]. D'Incecco et al. [23] trained a sequence-to-point DL model on the REFIT dataset [22], as was done in this work, but they assessed its generalization capability using the U.K.-DALE [28] and REDD [27] houses as test sets. These datasets, along with REFIT, are widely used to evaluate the performance of energy disaggregation algorithms.

This study elucidates the model's robustness and applicability to real-world scenarios through the above assessment. The findings demonstrate that the model can generalize well

TABLE III
PERCENTAGE RELATIVE ERROR AND MEAN ABSOLUTE
ERROR ON A HALF-YEARLY BASIS

	House 1		House 2	
	SAE	MAE	SAE	MAE
Dishwasher	9 %	0,38	11 %	0,54
Washing machine	11 %	0,57	12 %	0,67
Fridge	7 %	2,43	10 %	1,47

TABLE IV
PERCENTAGE RELATIVE ERROR AND MEAN ABSOLUTE ERROR ACHIEVED
IN [23] BY TRAINING THE MODEL ON THE REFIT DATASET

	Tested on UK-DALE		Tested on REDD	
	SAE	MAE	SAE	MAE
Dishwasher	13 %	16,49	71 %	29,68
Washing machine	50 %	14,84	74 %	36,83
Fridge	9 %	17,00	2 %	38,63

to different datasets and households despite being trained on consumption patterns related to U.K. homes and installed in two Italian homes with different appliances and usage patterns. This indicates that the model can effectively adapt to different settings and perform well under various conditions.

The results presented in [23] were obtained by processing the data offline; in other words, the model was trained and tested on prerecorded data. In contrast, the results of the current article were obtained via the real-world implementation of the measurement system, where the model was applied to real-time acquired data. This further confirms the effectiveness and reliability of the model for on-field applications. Overall, these findings provide important insights into the model's generalization ability, applicability, and scalability, which are critical factors for its practical implementation and adoption.

V. CONCLUSION AND FINAL REMARKS

The proposed system offers an embedded solution energy disaggregation by leveraging a pretrained DL model. Specifically, this model is a sequence-to-point ANN designed to predict the midpoint of the window of an appliance-level power reading from the corresponding aggregate power window. The public dataset REFIT, which provides the consumption details of 21 U.K. houses, was used to train this model.

An architecture was developed to evaluate the performance of the proposed NILM system. An appliance-level power meter was installed on each appliance to be analyzed: dishwashers, washing machines, and fridges. A WLAN network with a star topology was established, where both the NILM system and the appliance-level power meters were connected to a concentrator via Wi-Fi.

The implemented NILM system consists of an EVALSTPM32 board for measuring electrical quantities, a NUCLEO-H743ZI2 board for processing the acquired signals, and an ESP32 Wi-Fi module for connecting the system to the WLAN. The appliance-level power meters

comprise an EVALSTPM32 board for measuring electrical quantities and a USR-W610 Wi-Fi module for connecting to the WLAN. Notably, the data processing section is not included in the appliance-level power meters.

The WLAN network is managed using an Intel NUC NUC5i7RYH, which downloads measurement data and stores them in a MySQL database. Testing was conducted for 12 months; the system was installed in two houses in central Italy, one in Marche and one in Abruzzo, and tested for six months for each house. During this time, the system processed the overall consumption of the houses to obtain details on the individual consumption of their dishwashers, washing machines, and fridges. Data analysis showed that the NILM system adapted satisfactorily to both houses despite the differences in the absorption profiles of the appliances. In addition, the system adapted when the same appliance had different duty cycles (as in the case of the dishwashers and washing machines). The six-month maximum relative percentage errors for the dishwashers, washing machines, and fridges were 11%, 12%, and 10%, respectively.

Overall, these results were highly satisfactory, especially compared with those of a prior study where the model was evaluated offline using prerecorded data. The outcomes of the present study were obtained through the real-world implementation of the measurement system, where the model was applied to real-time-acquired data. This confirmed the effectiveness and reliability of the model for on-field applications, specifically its generalization ability, applicability, and scalability. Such findings provided crucial insights into the practical implementation and adoption of the model.

The system was designed and implemented to demonstrate the feasibility of a solution based on a small, low-power microcontroller for real-time energy consumption monitoring. The microcontroller was chosen for its excellent performance and numerous integrated peripherals, which enabled the development of a highly energy-efficient system. Because of its Arm Cortex-M7 architecture and maximum clock frequency of 480 MHz, the microcontroller can handle complex DL algorithms while maintaining low power consumption.

The small size (20 × 20 mm) of the microcontroller enabled the creation of a compact energy consumption monitoring system. This feature is particularly important in areas needing a minimally invasive, low-environmental-impact solution, such as in the Internet of Things, where devices must be compact and nonintrusive.

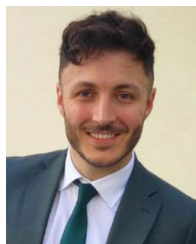
The proposed solution is not intended to become a commercial product. Instead, it is designed to demonstrate the feasibility of real-time energy consumption monitoring using a solution based on a small, low-power microcontroller. Nonetheless, the small size and low power consumption of the proposed system make it potentially suitable for commercial energy consumption monitoring solutions [34], [35].

Access to commercial energy consumption monitoring solutions through NILM remains limited, especially for end consumers. Companies developing NILM solutions [3], [4], [5] focus on B2B services rather than direct B2C hardware sales. This is mainly because NILM technology is primarily used for energy management and monitoring in commercial and

industrial settings, not homes. These companies typically offer a wide range of services, such as energy audit, monitoring, and reporting and energy efficiency consulting, to businesses and organizations. This approach allows them to collaborate closely with customers to gain a deep understanding of their specific energy usage patterns and provide tailored solutions for reducing energy consumption and costs. However, this increases the difficulty of comparing NILM systems proposed by researchers with commercially available NILM systems, as the former systems are not readily accessible to the general public.

REFERENCES

- [1] G. Bucci, F. Ciancetta, E. Fiorucci, S. Mari, and A. Fioravanti, "Multi-state appliances identification through a NILM system based on convolutional neural network," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, May 2021, pp. 17–21.
- [2] G. Bucci, F. Ciancetta, E. Fiorucci, S. Mari, and A. Fioravanti, "Deep learning applied to SFRA results: A preliminary study," in *Proc. 7th Int. Conf. Comput. Artif. Intell.*, Tianjin, China, Apr. 2021, pp. 302–307, doi: [10.1145/3467707.3467753](https://doi.org/10.1145/3467707.3467753).
- [3] Voltaware. (2019). *Intelligence for Your Home*. [Online]. Available: <https://www.voltaware.com>
- [4] Sense. (2018). *Sense Energy Monitor*. [Online]. Available: <https://sense.com/product>
- [5] NET2GRID. (2021). *NILM Systems*. [Online]. Available: <https://www.net2grid.com/>
- [6] G. W. Hart, "Prototype nonintrusive appliance load monitor," MIT Energy Lab., Electr. Power Res. Inst. Tech. Rep. 2, 1985.
- [7] M. Dong, Meira, W. Xu, and C. Y. Chung, "Non-intrusive signature extraction for major residential loads," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1421–1430, Sep. 2013.
- [8] G. Bucci, F. Ciancetta, E. Fiorucci, and S. Mari, "Load identification system for residential applications based on the NILM technique," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, May 2020, pp. 1–6, doi: [10.1109/I2MTC43012.2020.9128599](https://doi.org/10.1109/I2MTC43012.2020.9128599).
- [9] D. Teshome, T. D. Huang, and K.-L. Lian, "A distinctive load feature extraction based on Fryze's time-domain power theory," *IEEE Power Energy Technol. Syst. J.*, vol. 3, no. 2, pp. 60–70, Jun. 2016.
- [10] S. Gupta, M. S. Reynolds, and S. N. Patel, "ElectriSense: Single-point sensing using EMI for electrical event detection and classification in the home," in *Proc. 12th ACM Int. Conf. Ubiquitous Comput.*, Sep. 2010, pp. 139–148.
- [11] G. Bucci, F. Ciancetta, E. Fiorucci, S. Mari, and A. Fioravanti, "Measurements for non-intrusive load monitoring through machine learning approaches," *Acta IMEKO*, vol. 10, no. 4, pp. 90–96, 2021.
- [12] S. Dash and N. C. Sahoo, "Electric energy disaggregation via non-intrusive load monitoring: A state-of-the-art systematic review," *Electr. Power Syst. Res.*, vol. 213, Dec. 2022, Art. no. 108673, doi: [10.1016/j.epsr.2022.108673](https://doi.org/10.1016/j.epsr.2022.108673).
- [13] J. Z. Kolter, S. Batra, and A. Y. Ng, "Energy disaggregation via discriminative sparse coding," in *Proc. 23rd Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 1. Red Hook, NY, USA: Curran Associates, 2010, pp. 1153–1161.
- [14] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2011, pp. 747–758.
- [15] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *Proc. Nat. Conf. Artif. Intell.*, Toronto, ON, Canada, Jul. 2012, pp. 1–8. [Online]. Available: <http://eprints.soton.ac.uk/id/eprint/336812>
- [16] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, and F. Piazza, "Non-intrusive load monitoring by using active and reactive power in additive factorial hidden Markov models," *Appl. Energy*, vol. 208, pp. 1590–1607, Dec. 2017.
- [17] F. Paradiso, F. Paganelli, D. Giuli, and S. Capobianco, "Context-based energy disaggregation in smart homes," *Future Internet*, vol. 8, no. 1, p. 4, Jan. 2016.
- [18] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ. (BuildSys)*. New York, NY, USA: Association for Computing Machinery, 2015, pp. 55–64, doi: [10.1145/2821650.2821672](https://doi.org/10.1145/2821650.2821672).
- [19] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for non-intrusive load monitoring," in *Proc. AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, 2018, pp. 2604–2611.
- [20] Z. Jia, L. Yang, Z. Zhang, H. Liu, and F. Kong, "Sequence to point learning based on bidirectional dilated residual network for non-intrusive load monitoring," *Int. J. Electr. Power Energy Syst.*, vol. 129, Jul. 2021, Art. no. 106837, doi: [10.1016/j.ijepes.2021.106837](https://doi.org/10.1016/j.ijepes.2021.106837).
- [21] M. Yang, X. Li, and Y. Liu, "Sequence to point learning based on an attention neural network for nonintrusive load decomposition," *Electronics*, vol. 10, no. 14, p. 1657, Jul. 2021.
- [22] D. Murray, L. Stankovic, and V. Stankovic, "An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study," *Sci. Data*, vol. 4, no. 1, Jan. 2017, Art. no. 160122.
- [23] M. D'Incecco, S. Squartini, and M. Zhong, "Transfer learning for non-intrusive load monitoring," 2019, *arXiv:1902.08835*.
- [24] S. Mari, G. Bucci, F. Ciancetta, E. Fiorucci, and A. Fioravanti, "Advanced architecture for training and testing NILM systems," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, May 2022, pp. 1–6, doi: [10.1109/I2MTC48687.2022.9806577](https://doi.org/10.1109/I2MTC48687.2022.9806577).
- [25] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, "The ECO data set and the performance of non-intrusive load monitoring algorithms," in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Buildings*, 2014, pp. 80–89.
- [26] C. Shin, E. Lee, J. Han, J. Yim, W. Rhee, and H. Lee, "The ENERTALK dataset, 15 Hz electricity consumption data from 22 houses in Korea," *Sci. Data*, vol. 6, no. 1, pp. 1–13, Oct. 2019.
- [27] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. Workshop Data Mining Appl. Sustainability (SIGKDD)*, San Diego, CA, USA, vol. 25, 2011, pp. 59–62.
- [28] J. Kelly and W. Knottenbelt, "The U.K.-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five U.K. homes," *Sci. Data*, vol. 2, no. 1, Mar. 2015, Art. no. 150007.
- [29] X. Ying, "An overview of overfitting and its solutions," *J. Phys., Conf. Ser.*, vol. 1168, no. 2, Feb. 2019, Art. no. 022022, doi: [10.1088/1742-6596/1168/2/022022](https://doi.org/10.1088/1742-6596/1168/2/022022).
- [30] M. Abadi et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Accessed: Feb. 25, 2023. [Online]. Available: <https://www.tensorflow.org>
- [31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–15, doi: [10.48550/arXiv.1412.6980](https://doi.org/10.48550/arXiv.1412.6980).
- [32] Anaconda. (2020). *Anaconda Software Distribution*. Accessed: Feb. 13, 2020. [Online]. Available: <https://www.anaconda.com/distribution/>
- [33] *Node-RED*. Accessed: Feb. 25, 2023. [Online]. Available: <https://nodered.org/>
- [34] *JCGM 100: Evaluation of Measurement Data-Guide to the Expression of Uncertainty in Measurement (GUM)*, 1st ed., Bureau International des Poids et Mesures, 2012.
- [35] S. Mari, G. Bucci, F. Ciancetta, E. Fiorucci, and A. Fioravanti, "A review of non-intrusive load monitoring applications in industrial and residential contexts," *Energies*, vol. 15, no. 23, p. 9011, Nov. 2022, doi: [10.3390/en15239011](https://doi.org/10.3390/en15239011).
- [36] G. Bucci, F. Ciancetta, E. Fiorucci, S. Mari, and A. Fioravanti, "State of art overview of non-intrusive load monitoring applications in smart grids," *Meas., Sensors*, vol. 18, Dec. 2021, Art. no. 100145, doi: [10.1016/j.measen.2021.100145](https://doi.org/10.1016/j.measen.2021.100145).



Simone Mari (Member, IEEE) received the B.S., M.S., and Ph.D. degrees (Hons.) in electrical engineering from the University of L'Aquila, L'Aquila, Italy, in 2017, 2019, and 2023, respectively.

He presently serves as a Research Fellow in electrical and electronic measurements at the University of L'Aquila. His research primarily focuses on non-intrusive load monitoring, predictive maintenance systems utilizing artificial intelligence, automated measurement systems, and power measurements.



Giovanni Bucci (Member, IEEE) received the master's degree in electrical engineering from the University of L'Aquila, L'Aquila, Italy, in 1985.

He worked with Selenia Spazio, Rome, Italy, as a Designer of automatic test equipment until 1989. Since 2006, he has been a Full Professor of electrical measurements at the University of L'Aquila. His current research interests include smart sensors, multiprocessor-based measuring systems, digital measurement stations operating in real time, and power measurements. He has authored more than 180 scientific articles in these fields.



Edoardo Fiorucci (Senior Member, IEEE) received the Ph.D. degree in electrical and information engineering from the University of L'Aquila, L'Aquila, Italy, in 2004.

He is currently a Full Professor of electrical measurements with the University of L'Aquila. His research interests include power measurements, testing of electrical machines and power systems, pc-based instrumentation, measurement systems for power quality evaluation, and smart web sensors for distributed measurement applications. He has

authored more than 100 scientific articles in these fields.



Fabrizio Ciancetta (Member, IEEE) received the Ph.D. degree in electrical and information engineering from the University of L'Aquila, L'Aquila, Italy, in 2009.

He is currently an Assistant Professor of electrical measurements at the University of L'Aquila. His research and technical interests cover several aspects related to distributed measurement systems, signal processing, and multilevel inverter. He has authored about 70 journal and conference papers.



Andrea Fioravanti (Member, IEEE) received the master's degree in electrical engineering and the Ph.D. degree in electrical and information engineering from the University of L'Aquila, L'Aquila, Italy, in 2016 and 2021, respectively.

He is currently a Research Fellow and a Professor of electrical power systems at the Department of Industrial and Information Engineering and Economics, University of L'Aquila. His research interests are focused on the resilience of electrical power systems.