IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 21, 2020, accepted November 16, 2020, date of publication November 20, 2020, date of current version December 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3039639

Automated Non-Intrusive Load Monitoring System Using Stacked Neural Networks and Numerical Integration

SUELENE DE JESUS DO CARMO CORRÊA AND ADRIANA R. G. CASTRO
Institute of Technology, Federal University of Pará, Belém 66075-110, Brazil

Corresponding author: Suelene de Jesus do Carmo Corrêa (suelene@ufpa.br)

This work was supported in part by the PROPESP from Federal University of Pará, and in part by the Federal Institute of Education, Science and Technology of Pará.

ABSTRACT Population growth and new consumer needs, among other factors, have lead to growing energy demand, without a concomitant increase in energy generation. This way, reduction and rationalization of energy consumption, especially by residential users, have become a global concern generating a need for developing techniques for efficient management and distribution of the available energy. Non-Intrusive Load Monitoring (NILM) techniques have provided valuable information about energy consumption for power generation companies as well as consumers. Such information is important for making decisions related to sustainable use of energy resources. This study proposes an automated system based on Artificial Neural Network for performing some of the NILM tasks. A stacked neural network was developed to extract features of power signals of appliances to identify those in operation during a given period. This information is then used to disaggregate individual appliance loads through the total aggregate signal, and consumption is calculated through numerical integration. The system was tested using real data from two databases about appliances with On/Off, multi-level, and variable consumption patterns collected in low frequency. The performance metrics, resulting from identification and disaggregation tasks, demonstrate the efficiency of the proposed system.

INDEX TERMS NILM, AANN, feature extraction, load identification, pattern recognition.

I. INTRODUCTION

A significant challenge facing the world today is energy conservation. This challenge arose due to growing energy demand, as a result of many factors encompassing areas such as population growth, new consumer needs, environmental problems, among others [1], [2]. This scenario will overwhelm the capacity of existing energy resources and systems, generating an energy crisis that would compromise economic development in general. This crisis tends to grow as global energy demand is expected to grow one-third by 2040, while global economy is expected to grow by 150%, according to the International Energy Agency (IEA) [3].

Reduction and rationalization of energy consumption is an urgent necessity, especially with regard to residential use of electricity, since it is a sector of high demand and with good possibilities of intervention. Studies like [4]–[6] indicate that the maximum energy savings can be achieved by providing consumers more explicit information about their daily consumption.

This scenario has made it necessary to develop strategies for efficient management and distribution of the available energy. In this regard, non-intrusive techniques for monitoring energy consumption have been the focus of studies in recent decades [1], [2], [7], [8]. These techniques attempt to create systems to provide useful information for consumers and power generation companies [9], such as power consumed by each device and real-time energy feedback to support their decisions on planning and sustainable use of energy resources, to optimize their smart grid operations and propose tariffs according to users energy consumption habits [10].

NILM systems have the characteristic of using the aggregate consumption signal, captured by a central energy meter, to obtain individual appliance consumption. This approach does not interfere in the routine of the residents and costs less than intrusive load monitoring (ILM) because it does not require sensors for every appliance.

The associate editor coordinating the review of this manuscript and approving it for publication was Le Hoang Son.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

NILM systems are based on four fundamental tasks: event detection, extraction of characteristics, appliance identification and load disaggregation [11].

Considering appliance identification and load disaggregation, different techniques have been proposed in literature, such as: Discrete Fourier Transform [12]; Decision trees [13]; Principal Component Analysis (PCA) [14], [15], Genetic Programming [16], Artificial Neural Networks (ANN) [17]–[19], Deep Artificial Neural Networks (DNN) [20]–[26], Hidden Markov Models (HMM) [27]–[31], Integer and Quadratic Programming [32]–[34], Transfer learning [23], among others. Many of these techniques make use of several consumption parameters obtained from data collected at high frequency, which require expensive meters, that are not viable for residential use.

In this paper, a system based on a particular structure of Artificial Neural Network is proposed, a stacked Neural Network. This structure is composed of an Autoassociative Neural Network (AANN) and a Multi-layer Perceptron Neural Network (MLP) [35], to deal with the extraction of characteristics, appliance identification and load disaggregation tasks for NILM. Additionally, the calculation of individual appliance consumption is explored by applying numerical integration [36].

For the testing and validation of the system, two public databases with data collected under low frequency (<=1Hz) were used: the Reference Energy Disaggregation Dataset (REDD) [37], with data from six houses collected during 119 days and the UK Domestic Appliance Level Electricity (UK-DALE) [38], with data from five houses whose collection period varies from 2 months to 3,5 years. The proposed system was developed to identify seven of the most important appliance categories used in residences.

Important contributions of this study are the development of a system capable of using a single parameter (consumed Power) to perform the extraction of features, appliance identification and load disaggregation tasks over loads with different and complex operating behaviors and the use of power consumption signals collected in low frequency, which makes possible to use low-cost hardware for signal measurements and processed data storage. Moreover, the study analyzes the efficiency of using an AANN to extract the hidden features from appliance consumption data to improve the results of the identification and disaggregation tasks.

The structure of this paper is as follows; Section 2 presents related studies. Section 3 shows the system proposed and the methodology used in its development. Section 4 shows the experimental setup. Section 5 shows the results achieved, which allows the analysis of the feasibility of the system for feature extraction, appliance identification, and load disaggregation tasks, and Section 6 presents conclusions.

II. RELATED STUDIES

Some computational techniques and methods have been proposed to perform NILM tasks automatically. In this section, a summary of some studies that contributed significantly to this field of study is presented.

The first NILM study was carried out by Hart [39], who developed a finite state machine to represent a single appliance, where power consumption varied discretely at each step change. Hart's method used supervised learning and is based on optimization. This method is suitable for the identification of only a small number of appliances, since the complexity increases exponentially with the number of appliances, thus rendering it computationally costly [1].

Wang and Zheng [15] proposed a method that uses high frequency data and multiple features (current, voltage, consumed power and reactive power, power factor, variance of consumed power, among others) for residential appliance identification. The appliances were categorized according to working style and classified using mean-shift clustering. The identification accuracy with laboratory data was above 80%.

Semwal *et al.* [14] designed a NILM system which used high frequency data to identify the ON state of 8 appliances. The authors used amplitudes of current harmonics as features for the disaggregation task, applied PCA to reduce the dimension of the features, and compared three algorithms (SVM, ANN, and Bayes) for the identification task. Most of the appliances used in the study have a single level of operation and low consumption pattern, only one has high and variable consumption, and there are no appliances with similar consumption patterns.

Wong *et al.* [28] proposed a particle-based distribution truncation (PDT) method with a duration dependent hidden semi-Markov model (HSMM) for NILM. The method allows appliance state duration characteristics to be incorporated into the state transition model, eliminating the need to search over all possible state durations during the inference stage. The study was tested on REED data and got F-score of 0,826.

Johnson *et al.* [29] introduced a Factorial Hierarchical Dirichlet Process HSMM (F-HDP-HSMM) with explicit duration, developed sampling algorithms for efficient posterior inference and used modular Gibbs sampling that can be used for larger hierarchical Bayesian models to learn all levels of the model. The study was applied to perform energy disaggregation and exploited prior knowledge about each appliance power modes, levels and duration to learn the model details during inference instead of learning from training data.

Aiad *et al.* [30] applied for energy disaggregation a Factorial HSMM (FHMM) that considered mutual devices interactions and embedded the information about interactions into FHMM representations of the aggregated data. The hidden states in the FHMM were inferred through the Viterbi algorithm. The model was tested using REDD data and showed enhanced results when compared with the standard FHMM.

Makonin *et al.* [31] presented a disaggregation algorithm that used a super-state HMM and a variant of Viterbi algorithm which preserves dependencies between loads and can disaggregate multi-state loads. The study used low sampling rates and can run in real time. The disaggregation performance was high with CAE of 0.949 for test with five appliances from REDD dataset considering real noise conditions.

Kong *et al.* [32] proposed a hybrid NILM solver based on HMM and used low sampling rates, consumed power measurements, Integer Programming (IP) and Constraint Programming (CP) techniques. The first contribution of the study was to formulate the NILM problem as a segmented integer quadratic programming (SIQP) problem under the Viterbi decoding framework. The proposed framework significantly improved the computation efficiency when the problem size gets to the real-world size. The method was evaluated on REDD data and despite the low F-score of 0.51, it showed a satisfactory result in the task of disaggregation with a CAE of 0.816 dealing with 6 appliance categories.

Rottondi *et al.* [33] and Dash *et al.* [34] proposed optimization-based approaches for energy disaggregation of the main electrical appliances of residences. In [33] the authors applied quadratic programming; in [34] the authors applied integer programming. Both used ultra low frequency data (10 to 15 min resolution) and their results were poor compared to studies that use low and high frequency data.

Zhao *et al.* [40] applied Graph Signal Processing (GSP) to treat the energy disaggregation task. The authors made use of novel GSP concepts, applied at both, physical signal level via graph-based filtering and data level, via effective semi-supervised GSP-based feature matching. The study used the consumed power as appliance signature and was evaluated on REDD dataset.

Nascimento [20] applied recurrent and convolutional DNNs to identify and infer the consumption of three appliances (microwave, dishwasher, and refrigerator), making use of the sliding windows approach, low-frequency data from REDD dataset and consumed power as a signature. The study obtained accuracy of 68% to 99.8% depending on the appliance, algorithm, and test scenario. The main drawbacks of this method were the testing of only three appliances and the need of one network per appliance.

Kelly [21] proposed DNNs called long short-term memory (LSTM), denoising autoencoder (DAE) and a network that regresses the start time, end time and average power demand of each appliance activation (RECT). The study used windows of samples with fixed size to perform energy disaggregation. The author used one neural network per appliance, which makes the system computationally expensive (days of processing on a fast GPU) to run on an embedded processor inside a smart meter or in-home-display. The algorithms were tested on UK-DALE dataset.

Zhang et. al [22] proposed a sequence to point algorithm for a convolutional DNN to perform energy disaggregation. The study used a sliding window approach where the window of the input sequence is used to predict the output signal.The study was tested using low-frequency data of five appliances from REDD and UK-DALE.

Incecco *et al.* [23] proposed sequence to point algorithms for convolutional DNNs with appliance transfer learning (ATL), cross-domain transfer learning (CTL) and cross-domain transfer learning with fine tuning (CTL-FT) to perform energy disaggregation. The study used one convolutional DNN per appliance.

Harell *et al.* [26] presented a causal 1-D convolutional DNN to perform energy disaggregation on low-frequency data from AMPds dataset [41]. The study used various components of the complex power signal (current, active power, reactive power, and apparent power) as appliance signatures and obtained high performance in the disaggregation task.

Liu *et al.* [42] applied an event-based hybrid energy disaggregation approach using an improved multi-layer Hungarian algorithm to match transient features of appliances and a supervised adaptive resonance theory mapping neural network (ARTMAP) to cluster steady features for classification and identification of appliances. The study used a modified dual sliding window-based algorithm (DSWC) to detect transient events and convert the electrical features of appliances into a bipartite graph matching problem. Current, voltage, consumed power, reactive power and current harmonics collected at high frequency were the features used in the study.

Morais *et al.* [19] applied an ensemble of concurrent autoassociative neural networks to perform appliance identification using low-frequency data from REDD dataset and consumed power as appliance signature. The approach used one autoassociative neural network per appliance.

The cited studies show that most NILM techniques are based on a learning paradigm where the training stage is intrusive, requiring the use of plug level sensors for collecting data that are used for learning appliance specific models to be used for disaggregation during the operational phase, when the plug level sensors are removed. In contrast, some studies based on multi-label classification paradigm that do not require actual consumption data from all the appliances and can act on aggregate data have emerged [24], [43]–[45].

In [24] Massida *et al.* presented a disaggregation approach based in a convolutional DNN and multi-label classification to recognize the state of activation of the appliances and estimate their consumption. Although the study got good results, it considered only three appliances in the experiments and it was not able to deal with short consumption duration appliances like microwave, oven, and electric kettle, which are frequently used in residences.

Some studies like [40], [46] applied completely unsupervised techniques for energy disaggregation. However, the results obtained through the use of supervised techniques are significantly more accurate than these [47].

III. PROPOSED SYSTEM

The system proposed in this paper focuses on some steps of NILM, as shown in Fig. 1 and is composed by: a) A module to analyze the aggregate power signal for event detection; b) A module formed by an Autoassociative Neural Network [48], [49] to extract appliance features, and a classification MLP to perform appliance identification in a stacked structure and c) A module for load disaggregation based on numerical integration [36].

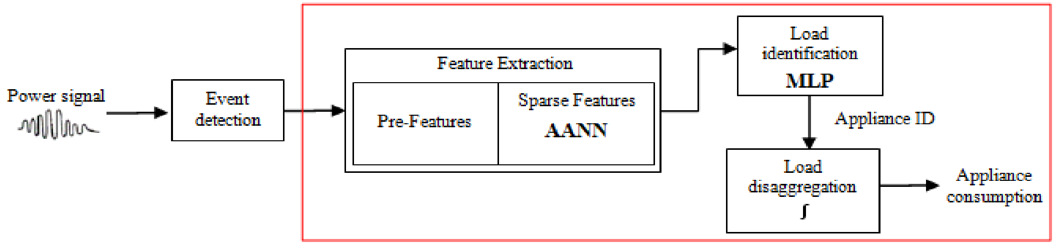

FIGURE 1. Logical architecture of the proposed system.

A. EVENT DETECTION MODULE

The event detection process consists of analyzing the consumed power signal of appliances and detecting the times when the appliances change their operational state. This module is based on the event detection approach presented in [39], [50], and analyzes the changes in the power levels of the signals. It is necessary to base the activities of the other modules that are the main contributions of this study.

State changes of appliance operation are identified by calculating the variation ΔP of the consumed power P between two consecutive instants t and $t+1$ calculated by $\Delta P = P_{t+1} - P_t$. If this variation is greater than a threshold and positive, it is an energizing event and if it is negative, it is a de-energization.

B. APPLIANCE IDENTIFICATION MODULE BASED ON ARTIFICIAL NEURAL NETWORKS

In this study, a stacked neural network model, formed by an AANN and a classification MLP, was created to extract useful data from an energy signal to characterize appliance load behaviors aiming to identify different categories of appliances that can compose an aggregate power signal. The appliances responsible for each signal that forms the aggregate consumption in a house can have different load behaviors and are classified into 4 main types [11]: On/Off; Finite State machines; Continuously Variable Devices and Permanent Consumption Devices. This variety of load behaviors makes the appliance identification process a challenging task.

In order to evaluate the effectiveness of the appliance identification module and the importance of the feature extraction module proposed, two models were trained and tested.

1) MODEL 1 (MD1) - BASED ON AANN AND MLP

Fig. 2 shows the proposed architecture for MD1, and comprises the following activities:

a: EXTRACTION OF SIMPLE FEATURES (PRE-FEATURES)

After the event detection process, the proposed system needs to obtain a set of representative features that can determine the behavior of the appliances taking into account their different modes of operation, allowing the identification of the activity of a specific appliance within an aggregate signal. The appliance signature used in this proposal for feature extraction is the consumed power signal of each appliance.

For all events detected by the previous module, the system collects simple features of the appliances (named here as pre-features), that are consecutive samples of power from the moment an appliance is turned on until the beginning of the stability period of the signal, from the moments the appliance changes the level of energy consumption and from the final moments of operation when it is turned off. This set of pre-features X_m selected from the power signal, is a matrix with 7 consecutive samples of power with a 3-second gap (REDD) or a 6-second gap (UK-DALE) and $time(t)$ associated with the events. Only seven samples are considered since according to the observation of different appliance data considered in this study, this quantity is sufficient to capture the characteristic behavior of the appliance consumption when they are going through this transition of states, while still capturing the beginning of the stability period, which is also important to characterize appliances. The pre-features selection procedure can be described by:

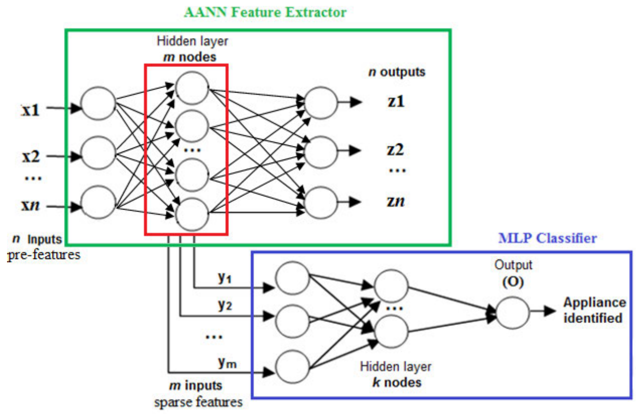

FIGURE 2. MD1 - Stacked neural network architecture proposed for appliance identification.

$$X_m = [x(t_m), \cdots, x(t_{m+(S-1)}); t_m, \cdots, t_{m+(S-1)}]^T \quad (1)$$

if an energizing event is detected, and if a de-energization event is detected the pre-features are described by:

$$X_m = [x(t_{m-(S-2)}), \cdots, x(t_{m+1}); t_{m-(S-2)}, \cdots, t_{m+1}]^T \quad (2)$$

where m is the number of the event identified between the total number of events M, S is the number of consecutive samples captured to represent the event (7 in this proposal), t_m is the moment of occurrence of a given event, and $x(t)$ is the power value measured at a given time.

b: EXTRACTION OF SPARSE FEATURES

In this study, the broad set of pre-features makes possible to collect even more representative features of the appliances behavior that can improve the performance of the identification task. This collection can be made through an AANN.

AANNs use unsupervised learning algorithms that apply backpropagation, setting the target values to be equal to the inputs [51]–[53], and do not use labels. They are used mainly to perform denoising, dimensionality reduction and feature extraction by learning efficient coding to obtain encoded representations for a set of data [49], [53]. The concept of AANN implies that the number of neurons in input and output layer neurons must be the same. The input data $x \in [0, 1]^d$ with dimension d are encoded to the hidden representation $y \in [0, 1]^{d'}$ through a deterministic mapping expressed by:

$$y = f_{\theta_1}(x) = f_{\theta_1}(W_1 x + b_1) \tag{3}$$

where $\theta_1 = \{W_1, b_1\}$. W_1 is a d' x d weight matrix, and b_1 is a bias vector of the hidden layer [53].

The resulting latent representation y is then mapped back to a reconstructed vector $z \in [0, 1]^d$ in input space. This mapping is expressed by:

$$z = g_{\theta'_1}(y) = g_{\theta'_1}(W'_1 y + b'_1) \tag{4}$$

with $\theta'_1 = \{W'_1, b'_1\}$. W'_1 is a weight matrix, and b'_1 is a bias vector of the output layer.

Each training sample x^i is mapped to a corresponding y^i and a reconstructed z^i, where f_{θ_1} and $g_{\theta'_1}$ are activation functions. The parameters of this model are optimized to minimize the average reconstruction error, the encoding-decoding information loss expressed by:

$$\begin{aligned} \theta_1^*, \theta_1'^* &= \underset{\theta_1, \theta'_1}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} L(x^i, z^i) \\ &= \underset{\theta_1, \theta'_1}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} L(x^i, g_{\theta'_1}(f_{\theta_1}(x^i))) \end{aligned} \tag{5}$$

where n is the total number of errors and L is a loss function like the squared error given by:

$$L(x, z) = ||x - z||^2 \tag{6}$$

In this study, the AANN is trained to learn sparse representations of the input data through the use of a larger number of hidden nodes than input nodes, which means representing high-dimensional original data using a sparse linear combination of the training samples, similarly to sparse coding [54], [55].

The pre-features of each appliance are used as inputs to train the AANN, which will store the characteristic data distribution of the appliances in a weight matrix θ. After training, the hidden layer nodes will contain features with a higher dimension than those provided as inputs to the network. The outputs of the AANN hidden layer y are the sparse features that can better represent the appliance's behavior and are described by (3).

c: APPLIANCE IDENTIFICATION

The sparse features y extracted using the AANN are used as inputs to a classification network, a simple MLP, that will provide as outputs o the identification of the appliances that are responsible for the events detected in the power signal as shown in Fig. 2. Each node of the MLP hidden layer k and output layer o has outputs calculated by equations (7) and (8).

$$k = f_{\theta_2}(y) = f_{\theta_2}(W_k y + b_k) \tag{7}$$

$$o = g_{\theta'_2}(k) = g_{\theta'_2}(W_o k + b_o) \tag{8}$$

where f_{θ_2} is the activation function, $W_k y$ is the weight matrix and b_k is the bias vector of the hidden layer; $g_{\theta'_2}$ is the activation function, $W_o k$ is the weight matrix and b_o is the bias vector of the output layer.

The output of the MLP, o, is a probability vector that indicates which appliance category an input pattern, X_m, given by (1) and (2) related to a detected event belongs to.

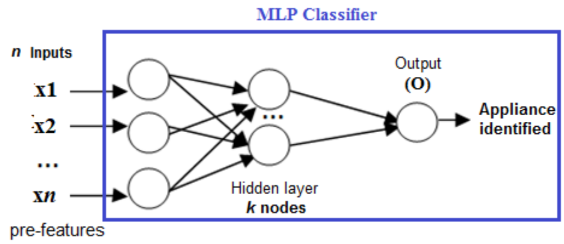

FIGURE 3. MD2 - Simple MLP architecture proposed for appliance identification.

2) MODEL 2 (MD2) - WITHOUT AANN

This model is the configuration of the network where the MLP classifier was trained to use the simple features (pre-features), X_m, given by (1) and (2), captured after the event detection to identify the appliances, without going through the sparse feature extraction process performed by the AANN used by MD1. This model is trained and tested to evaluate if the proposed sparse feature extraction process made in MD1 improves the classification performance of the identification system. Fig. 3 shows the MD2 structure.

C. LOAD DISAGGREGATION AND NUMERICAL INTEGRATION

Given an aggregate power signal P with a sequence of T samples measured in a period $P = \{\rho(1), \rho(2), \ldots, \rho(T)\}$ and considering K appliances in a house, the disaggregation process infers the power signatures P of each appliance that composes the aggregate power signal. The value of the aggregate signal at time t can be defined by:

$$\rho(t) = \epsilon^{(t)} + \sum_{k=1}^{K} p_k^{(t)} \tag{9}$$

where, $p_k^{(t)}$ is the power signature of a specific appliance at time t and $\epsilon^{(t)}$ represents noise and unmeasured appliance-level signals, such that, the power signature P of each of the K appliances through the entire period T are defined by Eq. (10).

$$\begin{aligned} P(t=1) &= \{\epsilon^{(1)}, p_1^{(1)}, p_2^{(1)}, \cdots, p_K^{(1)}\} \\ P(t=2) &= \{\epsilon^{(2)}, p_1^{(2)}, p_2^{(2)}, \cdots, p_K^{(2)}\} \\ &\vdots \\ P(t=T) &= \{\epsilon^{(T)}, p_1^{(T)}, p_2^{(T)}, \cdots, p_K^{(T)}\} \end{aligned} \tag{10}$$

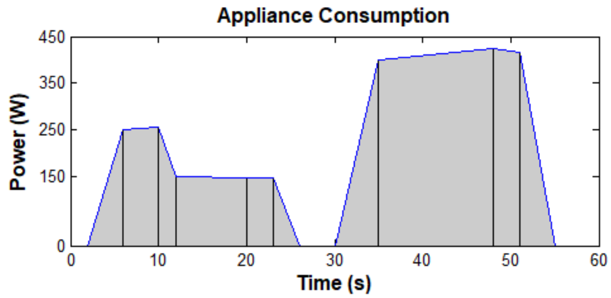

FIGURE 4. Appliance consumption viewed as a set of trapezoids.

This module receives information from the Appliance identification and Event detection modules to perform the disaggregation task. After the neural network classifier identifies which appliances are responsible for the events that occurred in the aggregate power signal at a given time, it is possible to separate/disaggregate all the events of a given appliance from the others that form the aggregate signal.

The disaggregated events of each appliance are then associated with the set of pre-features X_m provided by the feature extraction module (see equations (1) and (2)). This way, an estimation of the individual power signature of appliance k can be represented by:

$$P_k = \{p_k^{(1)}, p_k^{(2)}, \cdots, p_k^{(T)}\}; \quad 1 \leq k \leq K \tag{11}$$

Impossible sequences of events (i.e., consecutive de-energization events without the corresponding energization events) generated due to misclassifications from the previous module can be found and discarded through an analysis of conflicting values associated to the same time stamp, consecutive up events starting from the same value, variance, standard deviation and mean of the data samples associated with the events.

These respective samples of P and time (t) (considering that between two events there is the stability period) allow the generation of a graph that shows the approximate temporal consumption behavior of the appliance, as shown in Fig. 4. The area under the discrete curve formed by the values of P through time t represents the overall consumption of the appliance and can be calculated through numerical integration using the trapezoidal method [36]. Numerical integration is the approximate computation, using numerical techniques, of an integral of a function with a real variable defined in the [a,b] interval [36].

IV. EXPERIMENTAL SETUP

Some experiments were designed and tested to evaluate the feasibility of using stacked neural networks and numerical integration to enable the identification of different appliances and the disaggregation of their loads from an aggregate power signal in an automated way. First, the database and appliance categories selected for the study are presented, and then the model architecture and performance evaluation metrics are described.

TABLE 1. Quantity of patterns from REDD and UK-DALE used for training, validation and testing.

REDD				UK-DALE			
$N°$	Appliance	Type	Qtd.	$N°$	Appliance	Type	Qtd.
1	Refrigerator	A,B	217	1	Refrigerator	A,B	232
2	Microwave	C,D	206	2	Microwave	C,D	255
3	Stove	C	183	3	Stove	C	160
4	Oven	C	176	4	Oven	C	260
5	Dishwasher	C,D	256	5	Dishwasher	C,D	250
6	Air condit.	C	194	6	Washer	C,D	270
7	Washer/Dryer	C,D	250	7	Washer/Dryer	C,D	285
#	Total		1482	#	Total		1712

A - Always-on, B - On/Off, C - Multi state, D - Continuous variable

A. APPLIANCE MEASUREMENT DATABASES

The data used in this study for testing the proposed system are from the Reference Energy Disaggregation Dataset (REDD) [37] and UK Domestic Appliance-Level Electricity (UK-DALE) [38], large publicly available datasets created for promoting research in the NILM field, which is often difficult to carry out due to lack of data. These datasets contain individual and aggregate appliance data collected at low frequency (<= 1 Hz); the data come from six and five different houses, respectively. These datasets were selected for this study because they allow better comparison of results since they are used by most of the recent studies in the NILM field due to their completeness, complexity, and actualization.

This study considers seven appliances categories that have different consumption behavior and that are widely used in residences. The measurements of individual consumption of each appliance selected from the datasets, after passing through the event detection module, were used to create the database for training and testing of the models proposed in this study. The complete database resulted in 1482 patterns (10374 samples) from REDD and 1712 (11984 samples) patterns from UK-DALE, considering all houses of the datasets.

Since the quantity of data is high, the database was divided using the hold-out technique [56], [57] with stratification to ensure that each fold has the same proportion of observations with a given categorical value [58] into 70% of data for training, 15% for validation and 15% for test. In each of these sets, the data have a similar complexity and obey the same distribution. The stratification helps to make validation more stable and it is especially useful for unbalanced datasets and multiclass classification.

Table 1 shows the number of data patterns for each selected appliance from REDD and UK-DALE datasets.

B. APPLIANCE CATEGORIES

Seven appliances categories were selected from the databases to evaluate the identification and disaggregation accuracy of the proposed system, including those which are the most energy-consuming in homes. Table 1 shows this selection.

According to [8] the three reasons for modeling only the top-k energy consuming appliances instead of all appliances are: first, the disaggregation of such appliances provides the highest power value; second, such appliances contribute with the most remarkable features, and therefore the remaining appliances can be considered to contribute only with noise;

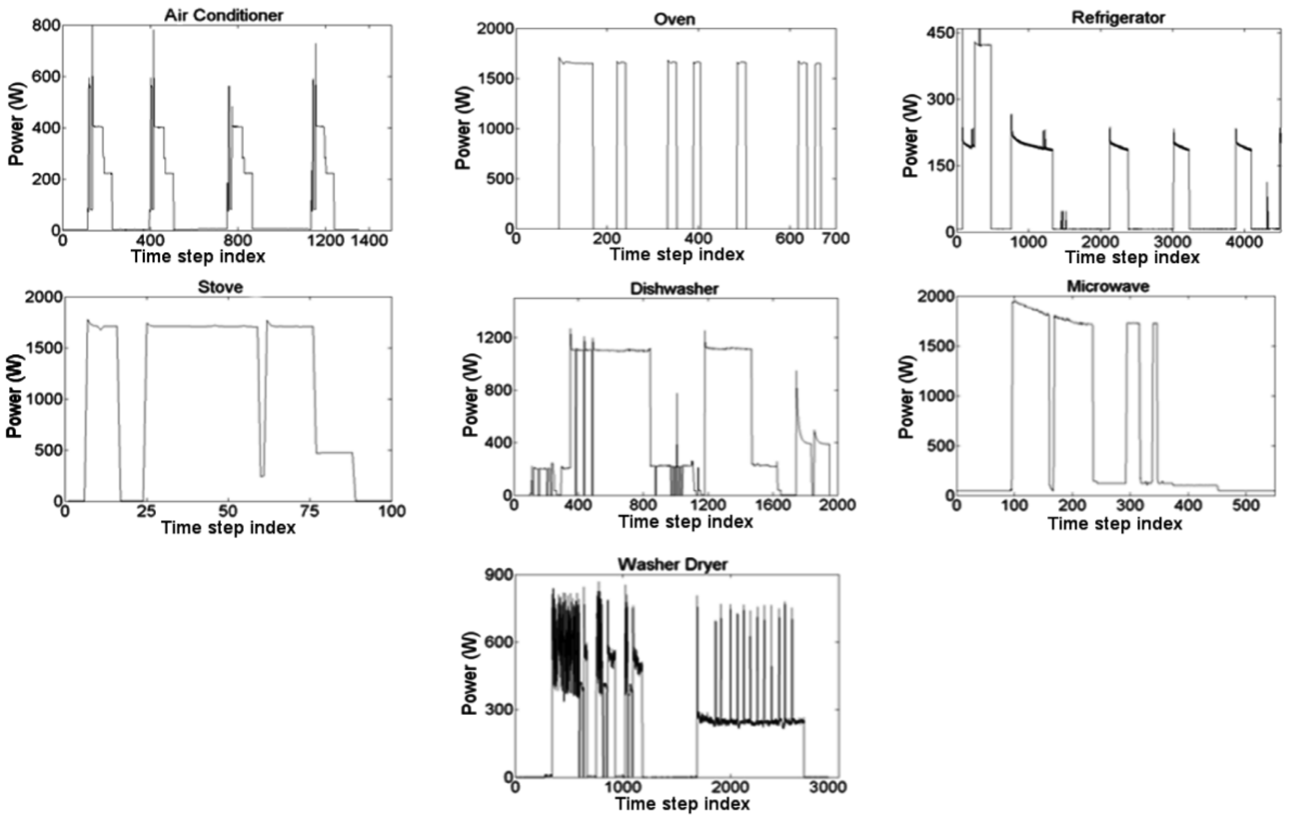

FIGURE 5. Examples of the operational behavior of appliances used in this study.

third, each additional modeled appliance might contribute significantly to increase the complexity of the task.

Fig. 5 presents examples of the characteristic consumption profile of appliances used in this study. The selected appliances comprise various types of operation modes (On/Off, Always-On, Multi-state, Continuously variable) [11], [21], similar consumption in certain levels of operation, and some have more complex behavior such as microwave, dishwasher and washer/dryer, which are multi-state but can present moments of continuously variable consumption, which increases the complexity of the identification task.

C. NEURAL NETWORK MODEL ARCHITECTURES

As discussed in previous sections and shown in Fig. 2, two neural network models were trained and tested to evaluate the performance of the proposed system.

The configuration parameters of the system architectures were chosen based on experimentation. The AANN and MLP were trained by varying the number of neurons in the hidden layer and the type of activation function (sigmoid logistic, hyperbolic tangent, linear). The loss function was the mean squared error.

The optimization method used was the scaled conjugate gradient (SCG) [59]. The value of the SCG parameter that determines the change in weight for second derivative approximation was 5.0e-5, and the value of the parameter used for regulating the indefiniteness of the Hessian was 5.0e-7. The parameters learning rate and momentum are calculated automatically at each iteration using second order information. The minimum performance gradient was 1e-6.

To avoid networks losing their capability of generalization, the early stopping technique [35], [57] was used to stop training when the validation error increases for a specified number of iterations indicating data overfitting.

According to the tests performed, the network configuration with best results is described below for MD1 and MD2.

1) MD1

AANN - The number of inputs and outputs of the AANN for MD1 that performs sparse feature extraction is 7, consisting of the pre-features, 7 consecutive power samples obtained after the event detection process. The model has one hidden layer, with 50 hidden neurons. The activation functions used were the Logistic sigmoid for the hidden layer and linear for the output layer. The training phase was set to 2000 epochs and the weights were initialized with random values sampled from a uniform distribution in the interval [0, 1].

MLP - In this model, the number of inputs of the classification MLP is 50; these inputs are the sparse features extracted with the AANN, that is, the outputs of the AANN hidden layer. This MLP has one hidden layer with 101 nodes, and the output layer has 7 neurons, equal to the number of appliances to be identified. The activation functions used were Logistic sigmoid for the hidden layer and Softmax for the output layer.

The training phase was set to 3000 epochs and the weights were initialized with random values sampled from a uniform distribution in the interval $[-R, +R]$, where $R = 1/\sqrt{n_{in}}$ and n_{in} is the number of inputs.

2) MD2

MLP - In this model, the classification MLP has seven inputs, the pre-features provided by the feature extraction module. The model has one hidden layer with 100 neurons, and the output layer has 7 neurons.The activation functions used were Logistic sigmoid for hidden layer and Softmax for the output layer. The training phase was set to 3000 epochs and the weights were initialized with random values sampled from a uniform distribution in the interval $[-R, +R]$, where $R = 1/\sqrt{n_{in}}$ and n_{in} is the number of inputs.

D. EVALUATION METRICS

Metrics used for performance evaluation of the proposed system concerning the classification task are: Accuracy (Acc), Sensitivity (Recall), Precision (Prec), Specificity (Specif), F-score and Corrected Assigned Energy (CAE). These metrics are common for evaluation of classifiers on NILM [32], [60] allowing comparisons with other studies.

The following terms are defined to calculate the metrics: True positive (TP), number of times an appliance is correctly classified as ON; True negative (TN), number of times an appliance is correctly classified as OFF; False positive (FP), number of times an appliance is incorrectly classified as ON and False negative (FN), number of times an appliance is incorrectly classified as OFF.

Accuracy (Acc): percentage of positive and negative samples correctly classified in all positive and negative samples.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

Sensitivity (Recall): percentage of positive samples correctly classified in all positive samples.

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

Precision (Prec): percentage of samples correctly predicted as positive on the total samples predicted as positive.

$$Prec = \frac{TP}{TP + FP} \tag{14}$$

Specificity (Specif): percentage of negative samples correctly identified on the total negative samples.

$$Specif = \frac{TN}{TN + FP} \tag{15}$$

F-measure or F-score: is the weighted average of Precision and Recall. It takes both false positives and false negatives into account. F-score is usually more useful than accuracy, especially if the class distribution is uneven.

$$F - score = \frac{2 * Prec * Recall}{Prec + Recall} \tag{16}$$

Other ways to visualize the results of the classification task are the AUC-ROC curve and the confusion matrix.

The AUC - ROC curve is a measure used to evaluate the performance of classifiers at various thresholds settings. ROC is a probability curve, and AUC represents the degree of separability. It indicates how much the model can distinguish between classes and A higher AUC indicates a better model. The ROC graph for multiclass problems shows a curve for each class in the same area. An excellent model with a high measure of separability has AUC near to 1. A model with the worst measure of separability has AUC near to 0 [60].

The Confusion matrix is a table used to visualize the performance of an algorithm for the problem of statistical classification. The matrix rows represent the instances in a predicted class and the columns represent the instances in an actual class (or vice versa). It makes easy to evaluate if the system is confusing two classes [60].

To evaluate the disaggregation performance, a metric to calculate the proportion of the total energy correctly assigned to an appliance (CAE) is used, and it is defined as:

$$CAE = \frac{1 - \sum_{t=1}^{T} \sum_{k=1}^{K} | \widehat{p}_k^{(t)} - p_k^{(t)} |}{2 \sum_{t=1}^{T} \overline{p}^{(t)}} \tag{17}$$

where $\widehat{p}_k^{(t)}$ is the estimated power consumption for the kth appliance at time t, $p_k^{(t)}$ is the actual power consumption for the appliance and $\overline{p}^{(t)}$ is the sum of the power consumption of all the appliances at each time.

V. EXPERIMENTAL RESULTS

This section summarizes the test results and presents them using the metrics described in section IV.

For a more precise analysis, two test scenarios were used. In the first one, the goal is to evaluate the general performance of the classification module of the system using data from all houses of REDD and UK-DALE datasets. The second scenario aims the evaluation of the system in performing the disaggregation task on the aggregate power signal formed by the seven appliances of house 1 from REDD, and houses 1 and 5 from UK-DALE. The selection of these houses is because they present at least half the number of appliances used in the study in operation during the same time interval.

A. SCENARIO 1 - CLASSIFICATION TEST

Tables 2 and 3 present the performance metrics for MD1 and MD2 using test data from REDD. MD1 obtained accuracy of 0,962, F-score of 0.879 and MD2 obtained accuracy of 0.942, F-score of 0.824. Also, according to precision and recall measures for MD1 the appliances more affected with false positives and negatives were stove and washer/dryer and for MD2 were stove, microwave, stove and dishwasher.

Tables 4 and 5 present the performance metrics for MD1 and MD2 using testing data from UK-DALE. MD1 obtained accuracy of 0.951, F-score of 0.858 and MD2 obtained accuracy of 0.946, F-score of 0.842. The measures of precision indicate that for MD1 the appliances with more false

TABLE 2. Proposed system results for appliance identification using AANN - MD1 - REDD - Scenario 1.

$N°$	Appliance	Accur	Recall	Prec	Specif	F-Score
1	Refrigerator	0,990	1,000	0,939	0,988	0,969
2	Microwave	0,965	0,935	0,853	0,970	0,892
3	Stove	0,932	0,786	0,733	0,955	0,759
4	Oven	0,985	0,926	0,962	0,994	0,943
5	Dishwasher	0,960	0,838	0,939	0,988	0,886
6	Air conditioner	0,965	0,897	0,867	0,976	0,881
7	Washer/Dryer	0,937	0,778	0,848	0,970	0,812
#	General	0,962	0,880	0,877	0,977	0,879

TABLE 3. Proposed system results for appliance identification without the AANN - MD2 - REDD - Scenario 1.

$N°$	Appliance	Accur	Recall	Prec	Specif	F-Score
1	Refrigerator	0,952	0,935	0,806	0,956	0,866
2	Microwave	0,928	0,774	0,774	0,957	0,774
3	Stove	0,938	0,750	0,808	0,970	0,778
4	Oven	0,973	0,926	0,893	0,981	0,909
5	Dishwasher	0,914	0,730	0,794	0,956	0,761
6	Air conditioner	0,942	0,828	0,800	0,963	0,814
7	Washer/Dryer	0,947	0,833	0,882	0,974	0,857
#	General	0,942	0,825	0,822	0,965	0,824

TABLE 4. Proposed system results for appliance identification using AANN - MD1 - UK-DALE - Scenario 1.

$N°$	Appliance	Accur	Recall	Prec	Specif	F-Score
1	Refrigerator	0,976	1,000	0,875	0,971	0,933
2	Microwave	0,962	1,000	0,800	0,955	0,889
3	Stove	0,981	0,833	1,000	1,000	0,909
4	Oven	0,967	1,000	0,833	0,960	0,909
5	Dishwasher	0,953	0,714	1,000	1,000	0,833
6	Washer	0,910	0,763	0,725	0,940	0,744
7	Washer/Dryer	0,910	0,650	0,813	0,967	0,722
#	General	0,951	0,852	0,864	0,970	0,858

TABLE 5. Proposed system results for appliance identification without the AANN - MD2 - UK-DALE- Scenario 1.

$N°$	Appliance	Accur	Recall	Prec	Specif	F-Score
1	Refrigerator	0,947	0,943	0,786	0,948	0,857
2	Microwave	0,990	1,000	0,941	0,988	0,970
3	Stove	0,975	0,833	0,952	0,994	0,889
4	Oven	0,975	1,000	0,875	0,970	0,933
5	Dishwasher	0,952	0,714	1,000	1,000	0,833
6	Washer	0,888	0,711	0,659	0,924	0,684
7	Washer/Dryer	0,892	0,650	0,722	0,945	0,684
#	General	0,946	0,836	0,848	0,967	0,842

positives were washer, washer/dryer and microwave, while for MD2 were washer, washer/dryer and refrigerator. Also, the recall measures indicate that for MD1 and MD2 the appliances with more false negatives were washer/dryer, washer and dishwasher.

Thus, comparing the results shown in Tables 2 to 5, it is possible to notice that the overall performance of the classifier is lower when the AANN is not used.

The AUC-ROC curves shown in Fig. 6 and 7 give a detailed view of MD1 and MD2 performance using REDD data. It is possible to observe that MD1 and MD2 got high performance for all classes since AUC values are close to 1, the optimal value.

The AUC-ROC curve shown in Fig. 8 and 9 show the MD1 and MD2 performance using UK-DALE data. The models got high performance for all classes too.

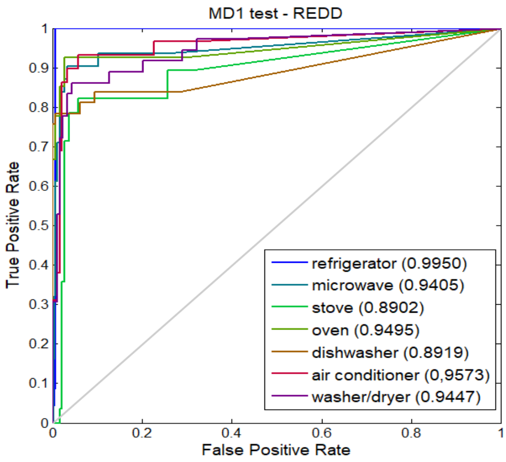

FIGURE 6. Resulting AUC-ROC - MD1 - REDD.

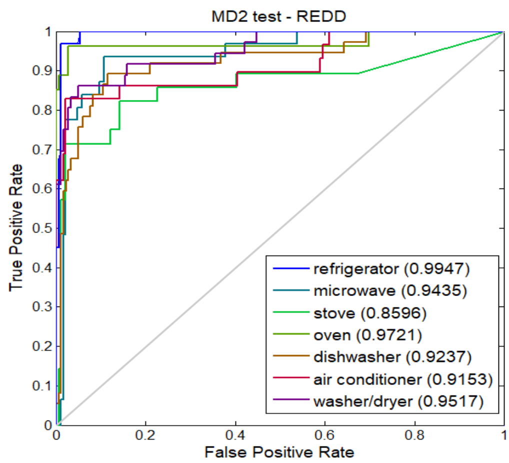

FIGURE 7. Resulting AUC-ROC - MD2 - REDD.

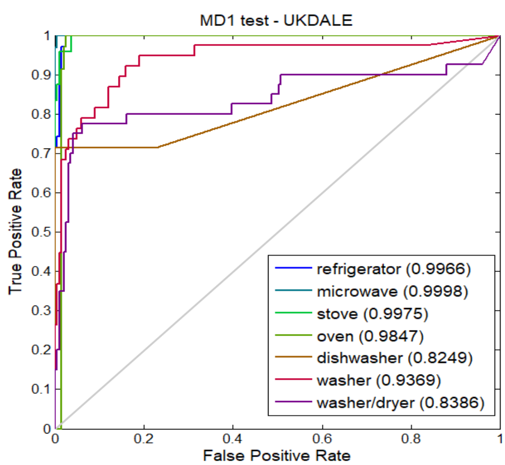

FIGURE 8. Resulting AUC-ROC - MD1 - UK-DALE.

The AUC values show that overall MD2 performs slightly lower than MD1, corroborating the results of Tables 2 to 5.

The confusion matrix shown in Fig. 10 details the MD1 performance using test data from REDD. The system correctly classified 100% of the 31 patterns of class 1 (refrigerator); the appliance class with the highest percentage

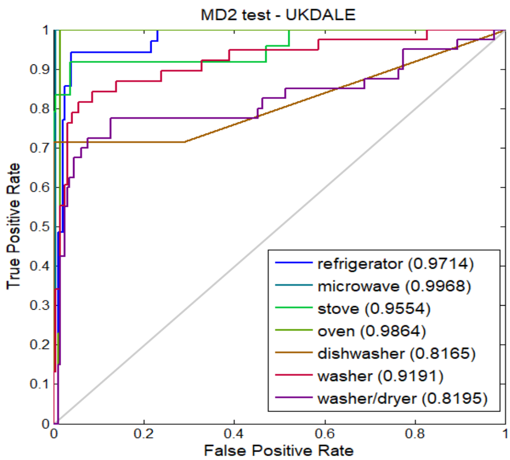

FIGURE 9. Resulting AUC-ROC - MD2 - UK-DALE.

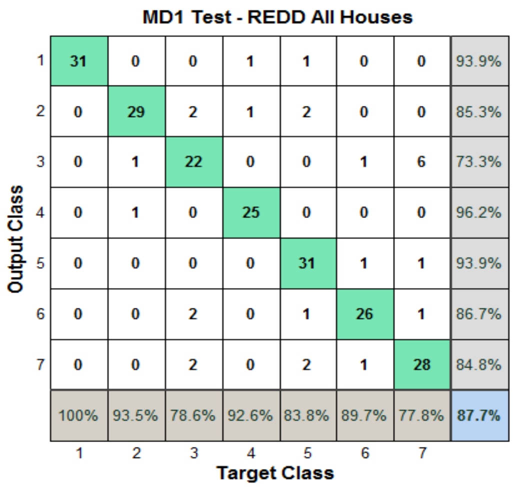

FIGURE 10. Resulting confusion matrix - MD1 - REDD.

of missclassification was 7 (washer/dryer), where 16.7% of the 36 patterns where classified as class 3 (stove), 2.75% as class 5 (dishwasher) and 2.75% as class 6 (air conditioner).

The confusion matrix shown in Fig. 11 presents the MD2 performance using test data from REDD. It shows that the system correctly classified 92.6% of the 27 patterns of class 4 (oven) and 7.4% were erroneously identified as class 2 (microwave); the appliance class with the highest percentage of incorrect classification was 5 (dishwasher), with 10.8% of the 37 patterns classified as 1 (refrigerator), 2.7% as 2 (microwave), 2.7% as 3 (stove), 8.1% as 6 (air conditioning) and 2.7% as 7 (washer/dryer).

The confusion matrix shown in Fig. 12 presents the MD1 performance using test data from UK-DALE. The system correctly classified 100% of the 35 patterns of class 1 (refrigerator); the appliance class with the highest percentage of incorrect classification was the 7 (washer/dryer), with 10% of the 40 patterns classified as 2 (microwave), 10% as 4 (oven) and 15% as 6 (washer).

The confusion matrix shown in Fig. 13 shows the MD2 performance using test data from UK-DALE. The system correctly classified 100% of the 32 patterns of class 2 (microwave), the 32 patterns of class 2 (microwave) and the

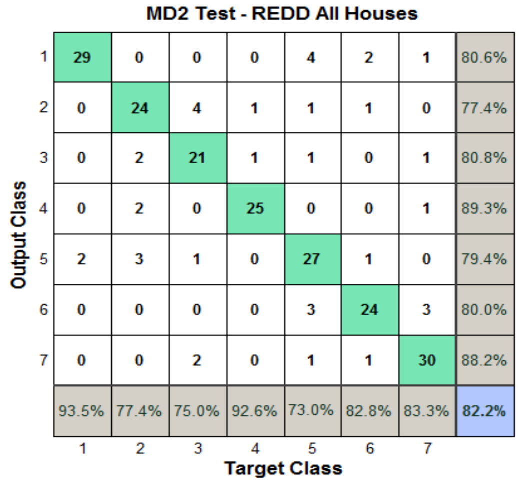

FIGURE 11. Resulting confusion matrix - MD2 - REDD.

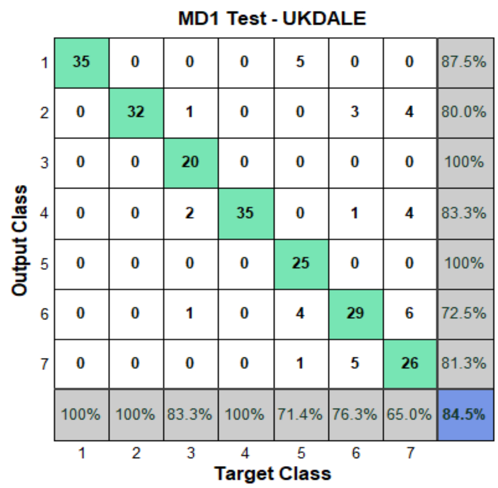

FIGURE 12. Resulting confusion matrix - MD1 - UK-DALE.

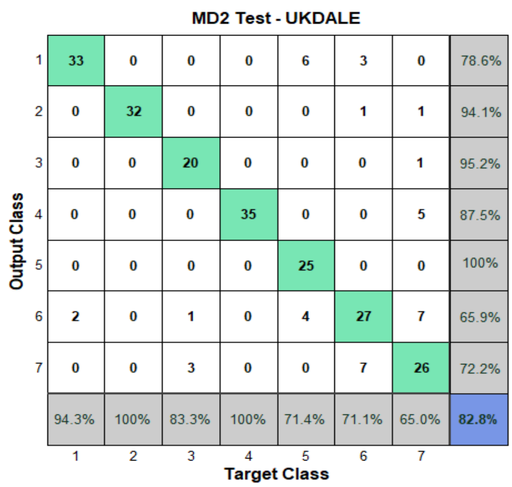

FIGURE 13. Resulting confusion matrix - MD2 - UK-DALE.

35 patterns of class 4 (oven); the appliance class with more misclassifications was 7 (washer/dryer), where 17.5% of the 40 patterns were classified as 6 (washer), 12.5% as 4 (oven), 2.5% as 3 (stove) and 2.5% as 2 (microwave).

TABLE 6. Comparative results of NILM approaches for the appliance identification task - REDD and UK-DALE datasets.

Dataset	Author	Technique	$N°$ Appl. categories	Appliance categories	Recall	Prec	F-Score
REDD	This study	AANN and MLP	7	refrigerator, microwave, oven, dishwasher, washer/dryer, stove, air conditioner	0,880	**0,877**	**0,879**
		MLP	7		0,825	0,822	0,824
	[32] Kong et al.	IP-HMM, and CP-HMM	6	refrigerator, microwave, oven, dishwasher, washer/dryer, kitchen outlet	0,583	0,463	0,516
	[28] Wong et al.	PDT-HMM	6	refrigerator, microwave, stove, dishwasher, kitchen outlet, lighting	**0,893**	0,768	0,826
	[40] Zhao et al.	GSP	6	refrigerator, microwave, stove, dishwasher, kitchen outlet, lights	0,779	0,76	0,769
UK-DALE	This study	AANN and MLP	7	refrigerator, microwave, oven, dishwasher, washer/dryer, stove, washer	**0,852**	**0,864**	**0,858**
		MLP	7		0,836	0,848	0,842
	[21] Kelly	DAE-DNN	5	kettle, fridge, washer, microwave, dishwasher	0,766	0,586	0,664
		RECT-DNN	5		0,534	0,786	0,636
		LSTM-DNN	5		0,710	0,396	0,508

The analysis of the errors shown in the confusion matrices indicates that most of the incorrect classifications are due to extreme similarities between some patterns of certain consumption levels of multi-state appliances like washer/dryer and dishwasher and washer/dryer and washer, that are in fact the same category and have the same operation mechanism when performing in the wash function only. The impact of these wrong classifications is mitigated during the disaggregation process when impossible sequences of events are identified and discarded.

Table 6 presents a comparison of the results of this study with studies that use other techniques to perform NILM tasks, like [32], with Integer Programming (IP) and Constraint Programming (CP) HMM, [28] with Particle-based Distribution Truncation (PDT) HMM, [40] with Graph Signal Processing (GSP) and [21] with a long short-term memory (LSTM) DNN, a denoising autoencoder (DAE) DNN and a DNN which regresses the start time, end time and average power demand of each appliance activation (RECT).

The results presented in Table 6 indicate that the proposed system got high performance in the appliance identification task considering seven appliance categories comparing to [21], [28], [32], [40] which perform the task for six or five appliance categories.

Concerning the comparison of the results, it must be taken into consideration that all the studies selected for comparison use data from REDD or UK-DALE dataset, though the particular data used by each author for training and testing may be different. So, a direct comparison of the results must be made with care.

B. SCENARIO 2 - DISAGGREGATION TEST

Data from REDD house 1 and data from UK-DALE houses 1 and 5 have been used to evaluate the disaggregation ability of the proposed system. It is important to highlight that the proposed system was trained and validated with data that covers the entire period of collection of the databases used. Scenario 2 points to specific periods that allow a more direct comparison with other studies and a better visualization of the disaggregation details.

TABLE 7. Proposed system results for appliance identification using AANN - REDD - Scenario 2.

$N°$	Appliance	Accur	Recall	Prec	Specif	F-Score
1	Refrigerator	0,966	0,800	0,889	0,987	0,842
2	Microwave	0,966	0,897	1,000	1,000	0,945
3	Stove	0,989	NaN	0,000	0,989	NaN
4	Oven	1,000	1,000	1,000	1,000	1,000
5	Dishwasher	0,896	0,884	0,884	0,906	0,884
6	Air conditioner	1,000	NaN	NaN	1,000	NaN
7	Washer/Dryer	0,966	NaN	0,000	0,966	NaN
#	General	0,969	0,895	0,896	0,978	0,896

REDD House 1 - aggregate power signal data from house 1 of REDD (starting in 2011/4/18) covering approximately 6h. This period was chosen to be in line with the test period reported in [32]. It should be noted that in this selected period of the day, only four of the seven appliances considered in this study were switched ON at least once.

UK-DALE Houses 1 and 5 - One-week data from UK-DALE house 1 (starting in 2014/6/30) and house 5 (starting in 2012/11/11) were combined to generate an aggregate signal containing all appliances used in this study, since not all the selected appliances are present in the same house.

Tables 7 and 8 present the results of the identification task achieved by MD1 and MD2 for house 1 of the REDD dataset. MD1 obtained accuracy of 0.969, F-score of 0.896 and MD2 obtained accuracy of 0.942, F-score of 0.823. The NaN values presented in Tables 7 and 8 represent the non-applicable calculations due to division by zero related to the appliances that were not used in the considered test period.

TABLE 8. Proposed system results for appliance identification without the AANN - REDD - Scenario 2.

$N°$	Appliance	Accur	Recall	Prec	Specif	F-Score
1	Refrigerator	0,878	0,700	0,467	0,900	0,560
2	Microwave	0,975	0,931	1,000	1,000	0,964
3	Stove	0,975	NaN	0,000	0,975	NaN
4	Oven	1,000	1,000	1,000	1,000	1,000
5	Dishwasher	0,823	0,721	0,861	0,906	0,785
6	Air conditioner	1,000	NaN	NaN	1,000	NaN
7	Washer/Dryer	0,975	NaN	0,000	0,975	NaN
#	General	0,942	0,823	0,823	0,965	0,823

TABLE 9. Proposed system results for appliance identification using AANN - UK-DALE - Scenario 2.

$N°$	Appliance	Accur	Recall	Prec	Specif	F-Score
1	Refrigerator	0,977	1,000	0,717	0,975	0,835
2	Microwave	0,981	0,913	0,839	0,987	0,874
3	Stove	0,998	1,000	0,600	0,998	0,750
4	Oven	0,991	0,970	0,780	0,992	0,865
5	Dishwasher	0,992	0,571	1,000	1,000	0,727
6	Washer	0,835	0,787	0,918	0,902	0,847
7	Washer/Dryer	0,844	0,829	0,829	0,690	735
#	General	0,945	0,867	0,792	0,958	0,828

TABLE 10. Proposed system results for appliance identification without the AANN - UK-DALE - Scenario 2.

$N°$	Appliance	Accur	Recall	Prec	Specif	F-Score
1	Refrigerator	0,963	1,000	0,629	0,960	0,772
2	Microwave	0,958	0,913	0,664	0,962	0,768
3	Stove	0,989	1,000	0,214	0,989	0,353
4	Oven	0,981	0,970	0,627	0,981	0,762
5	Dishwasher	0,992	0,619	1,000	1,000	0,765
6	Washer	0,784	0,690	0,926	0,920	0,791
7	Washer/Dryer	0,798	0,799	0,623	0,798	0,700
#	General	0,924	0,856	0,669	0,944	0,751

Since the appliances stove and refrigerator were not in operation during the test period, the precision of 0.000 indicates the occurrence of at least 1 false positive.

Tables 9 and 10 present the test results of the identification task achieved by MD1 and MD2 for the UK-DALE dataset. The MD1 obtained F-score of 0.828, and MD2 obtained F-score of 0.751.

Fig. 14 shows the actual load consumption profile for the disaggregation test period using appliance data from REDD house 1. Figures 15 and 16 present the estimated profile after disaggregation using MD1 and MD2, respectively. Observing the colored areas of each appliance consumption and comparing to the ground truth, it is possible to note differences in the areas of dishwasher and refrigerator. For MD1 these differences are smaller then for MD2.

Fig. 17 shows the actual load consumption profile for the disaggregation test period using the first two days of the test period using UK-DALE. Comparing the ground truth to the colored areas of Fig. 18 and Fig. 19 it is possible to observe differences in the estimated areas of washer, washer/dryer and and dishwasher. For MD1 these differences are smaller than for MD2.

The differences observed between the true and the estimated consumption profiles are due to the events that were incorrectly identified during the classification phase. Most of these incorrect classifications occur due to extreme similarities in the consumption profile of some

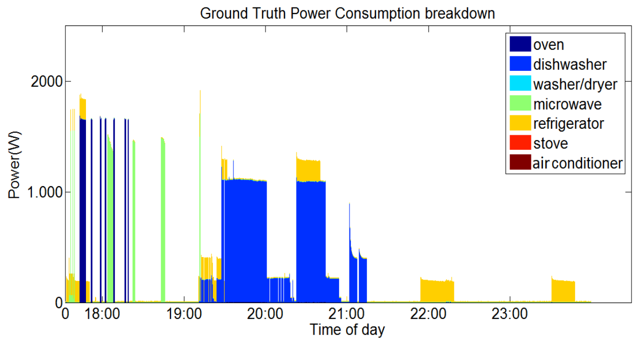

FIGURE 14. True power consumption breakdown for test scenario 2 - REDD.

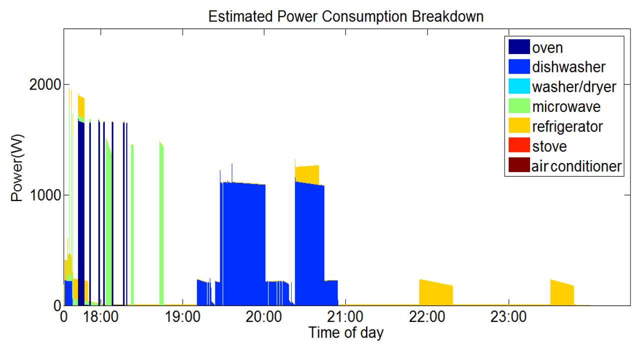

FIGURE 15. Estimated power consumption breakdown for test - MD1, Scenario 2 - REDD.

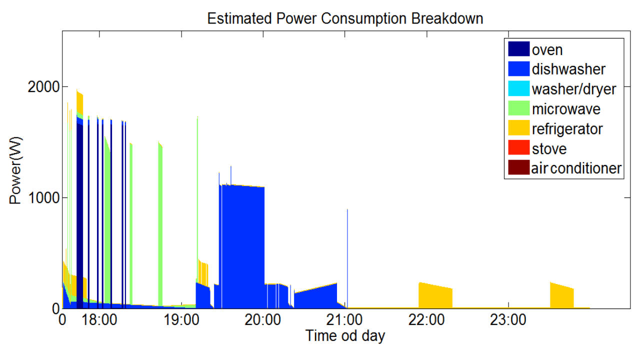

FIGURE 16. Estimated power consumption breakdown for test - MD2, Scenario 2 - REDD.

levels of operation of different appliances, as explained in subsection III.C.

To confirm that the values estimated by the proposed system have a high approximation to the real value of consumption, the metric Correctly Assigned Energy (CAE), which represents the energy correctly disaggregated, has been calculated and compared with the results obtained by other techniques presented in literature.

Table 11 shows the results of the CAE metric obtained by the proposed system and other studies like [32] that uses IP and CP with HMM, [40] with GSP, [30] with FHMM, [31] that uses HMM, [29] with F-HDP-HSMM, [21] that uses DNNs (DAE, RECT, LSTM) and [23] that uses DNNs with sequence-to-point learning with ATL and CTL. The results presented indicate that the proposed system has high

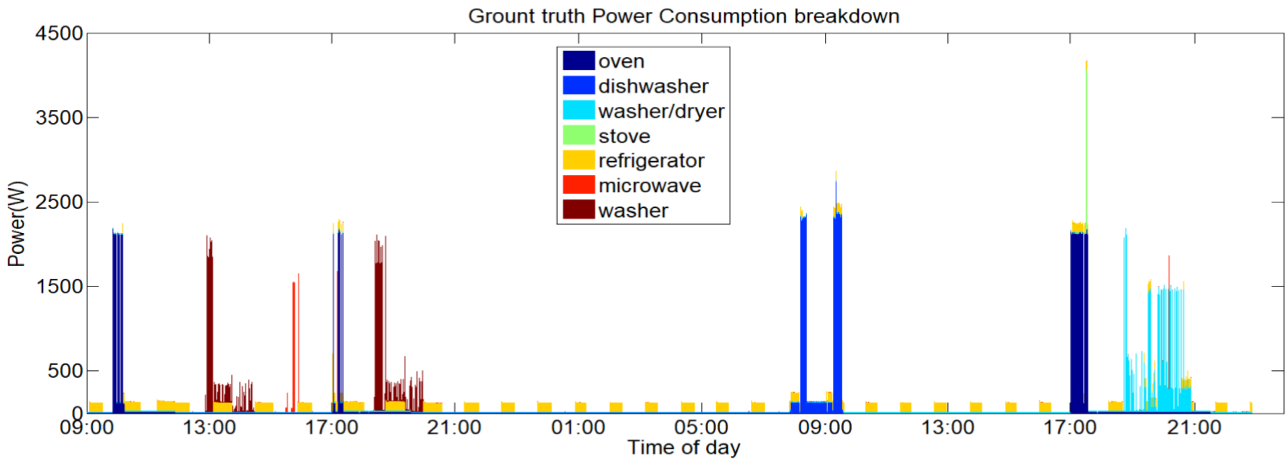

FIGURE 17. True power consumption breakdown for test scenario 2 - UK-DALE.

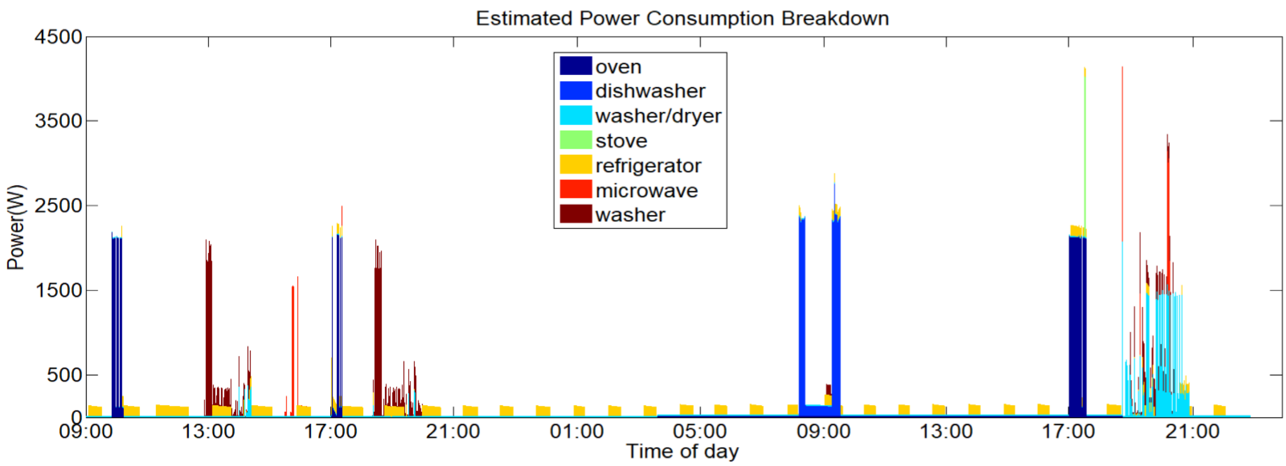

FIGURE 18. Estimated power consumption breakdown for test - MD1, Scenario 2 - UK-DALE.

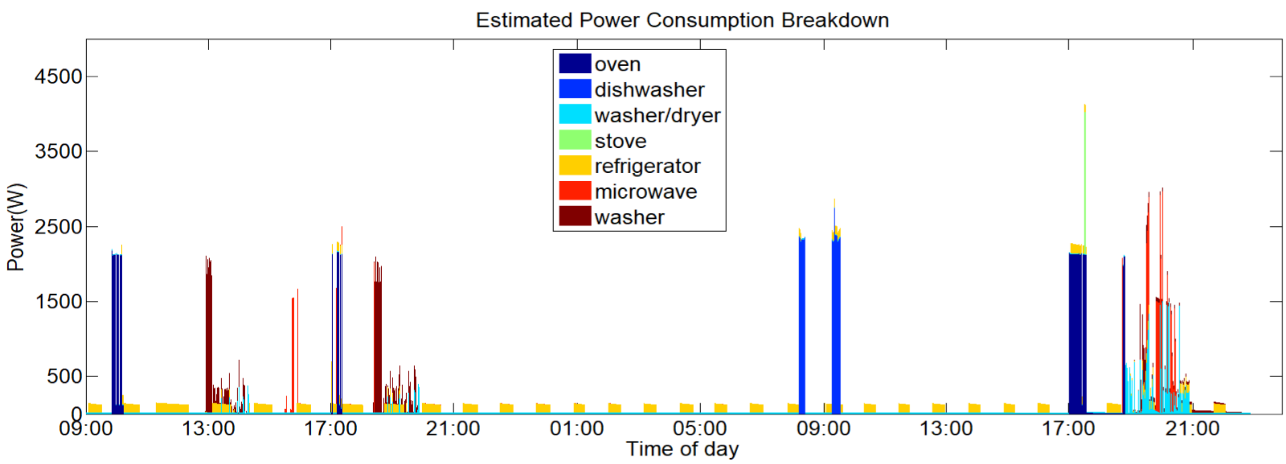

FIGURE 19. Estimated power consumption breakdown for test - MD2, Scenario 2 - UK-DALE.

performance in the disaggregation task, achieving a CAE of 0.965 and 0.842 for MD1 and MD2 respectively using data from REDD, values that are higher than the CAE of obtained by [29], [30], [32], [40], and CAE of 0.961 and 0.921 for MD1 and MD2 respectively using data from seven UK-DALE appliances, values comparable to the CAEs obtained by [21] that uses only five appliances to perform the disaggregation.

TABLE 11. CAE results for the disaggregation task using REDD and UK-DALE datasets.

Dataset	Author	Technique	$N°$ Appl. categories	Appliance categories	CAE
REDD	This study	AANN and MLP	7	refrigerator, microwave, oven, dishwasher, washer/dryer, stove, air conditioner	**0,965**
		MLP	7		0,894
	[32] Kong et al.	IP-HMM and CP-HMM	6	refrigerator, microwave, oven, dishwasher, washer/dryer, kitchen outlet	0,816
	[40] Zhao et al.	GSP	6	refrigerator, microwave, stove, dishwasher, kitchen outlet, lights	0,772
	[30] Aiad et al.	FHMM	6	refrigerator, microwave, stove, washer/dryer, kitchen outlet, lights	0,689
	[31] Makonin et al.	HMM	5	refrigerator, microwave, dishwasher, lights, furnace	0,949
	[29] Johnson et al.	F-HDP-HSMM	5	refrigerator, microwave, dishwasher, lights, furnace	0,815
UK-DALE	This study	AANN and MLP	7	refrigerator, microwave, oven, dishwasher, washer/dryer, stove, washer	0,961
		MLP	7		0,921
	[21] Kelly	DAE-DNN	5	kettle, fridge, washer, microwave, dishwasher	0,980
		RECT-DNN	5		**0,990**
		LSTM-DNN	5		0,920
	[23] Incecco et. al.	DNN-CTL	5	kettle, fridge, washer, microwave, dishwasher	0,932
		CTL-FT-DNN	5		0,910
		ATL-DNN	5		0,916

VI. CONCLUSION

In this study, an automated NILM system for residential use has been proposed. The system consists of an stacked structure of ANNs composed by an AANN used to extract the more relevant features useful to differentiate the energy behaviour of appliance loads during their operation and by a MLP classifier trained with these features to identify which appliances are responsible for the events detected along the aggregate power signal. Also, an additional module performs numerical integration over the results of the MLP classifier and detected events to disaggregate the loads of each appliance that form the aggregate power signal, which allows the calculation of their individual consumption in a given period.

Experimental results show that the proposed system is able to execute these tasks with high performance and are especially relevant considering that it deals with low frequency data, which makes possible to use low cost hardware for measuring and storage of processed data. Also, the study considers appliances with complex load behavior (multiple states, continuous variation), as well as similarity of power consumption in some levels of operation. Results also indicate that the use of AANN to extract sparse features from the data provides an important improvement in the performance of the proposed system.

The main limitation of approaches based on HMM is that the domain knowledge (base load, model noise) needs to be extracted a priori from the observation data and variables need to be explicitly modeled. This limitation makes these methods difficult to use. Artificial neural network approaches, like the one proposed in this study, unlike the HMMs, treat the variables as background, retrieving the specific appliance features from the aggregated power signal while extract the target.

Concerning the studies based on artificial neural networks that have been applied to NILM over the years, most of them use complex architectures such as recurrent and convolutional DNN structures with many layers that generate a large number of parameters to be processed. Also, most of them use one network per appliance to be identified. The system proposed in this study, unlike the other approaches, uses only one AANN with a single hidden layer to extract the features of all appliances, and one MLP with a single hidden layer to identify them, which makes it less computationally complex.

Besides that, it is possible to highlight other advantages innovated in this proposal: the use of few samples of consumed power and time of the up and down edges of each level of appliance operation, instead of using windows of samples with a fixed size related to the complete operating period of an appliance, which makes the system less sensitive to changes in the appliance usage pattern of different houses, a desirable characteristic for NILM systems. Finally, the use of numerical integration to calculate the individual consumption of the appliances eliminates the need to know the functions that represent these consumption patterns and provides a high approximation of the actual consumption.

As a future work, it is recommended to increase the number of appliances to be identified, to test the system with other databases and to consider scenarios where appliances go into operation at the same time.

REFERENCES

[1] A. Zoha, A. Gluhak, M. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838–16866, Dec. 2012.

[2] Y. F. Wong, Y. Ahmet Sekercioglu, T. Drummond, and V. S. Wong, "Recent approaches to non-intrusive load monitoring techniques in residential settings," in *Proc. IEEE Comput. Intell. Appl. Smart Grid (CIASG)*, Apr. 2013, pp. 73–79.

[3] International Energy Agency, *World Energy Outlook 2015*. Paris, France: OECD Publishing, 2015. [Online]. Available: https://www.oecd-ilibrary.org/content/publication/weo-2015-en

[4] G. Erbach, "Understanding energy efficiency," EPRS, Brussels, Belgium, Tech. Rep. PE 568.361, 2015. [Online]. Available: http://www.europarl.europa.eu/RegData/etudes/BRIE/2015/568361/EPRS_BRI%282015%29568361_EN.pdf

[5] M. Murray and J. Hawley, "GOT DATA? The value of energy data access to consumers," More Than Smart, Mission Data Coalition, Seattle, WA, USA, Tech. Rep. 2, Jan. 2016. [Online]. Available: http://www.missiondata.io/s/Got-Data-value-of-energy-data-access-to-consumers.pdf

[6] *The Role of ICT in Energy Efficiency Management Household Sector*, World Energy Council, London, U.K., 2018.

[7] M. Zeifman and K. Roth, "Viterbi algorithm with sparse transitions (VAST) for nonintrusive load monitoring," in *Proc. IEEE Symp. Comput. Intell. Appl. Smart Grid (CIASG)*, Apr. 2011, pp. 1–8.

[8] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, "NILMTK: An open source toolkit for non-intrusive load monitoring," in *Proc. 5th Int. Conf. Future Energy Syst.*, 2014, pp. 265–276.

[9] Y. Zheng, W. Kong, Y. Song, and D. J. Hill, "Optimal operation of electric springs for voltage regulation in distribution systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2551–2561, Apr. 2020.

[10] K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," *Energy Policy*, vol. 52, pp. 213–234, Jan. 2013.

[11] A. Ruano, A. Hernandez, J. Ureña, M. Ruano, and J. Garcia, "NILM techniques for intelligent home energy management and ambient assisted living: A review," *Energies*, vol. 12, no. 11, p. 2203, Jun. 2019.

[12] S. Dordević, M. Dimitrijević, and V. Litovski, "A non-intrusive identification of home appliances using active power and harmonic current," *Facta Universitatis,: Electron. Energetics*, vol. 30, no. 2, pp. 199–208, 2017.

[13] J. Maitre, G. Glon, S. Gaboury, B. Bouchard, and A. Bouzouane, "Efficient appliances recognition in smart homes based on active and reactive power, fast Fourier transform and decision trees," in *Proc. Workshops 29th AAAI Conf. Artif. Intell.*, 2015, pp. 24–29.

[14] S. Semwal, M. Singh, and R. S. Prasad, "Group control and identification of residential appliances using a nonintrusive method," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 23, pp. 1805–1816, 2015.

[15] Z. Wang and G. Zheng, "Residential appliances identification and monitoring by a nonintrusive method," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 80–92, Mar. 2012.

[16] F. Paganelli, F. Paradiso, S. Turchi, A. Luchetta, P. Castrogiovanni, and D. Giuli, "Appliance recognition in an OSGi-based home energy management gateway," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 2, Feb. 2015, Art. no. 937356.

[17] Y.-H. Lin and M.-S. Tsai, "A novel feature extraction method for the development of nonintrusive load monitoring system based on BP-ANN," in *Proc. Int. Symp. Comput., Commun., Control Autom. (3CA)*, May 2010, pp. 215–218.

[18] H.-H. Chang, K.-L. Lian, Y.-C. Su, and W.-J. Lee, "Power-Spectrum-Based wavelet transform for nonintrusive demand monitoring and load identification," *IEEE Trans. Ind. Appl.*, vol. 50, no. 3, pp. 2081–2089, May 2014.

[19] L. R. Morais and A. R. G. Castro, "Competitive autoassociative neural networks for electrical appliance identification for non-intrusive load monitoring," *IEEE Access*, vol. 7, pp. 111746–111755, 2019.

[20] P. P. M. Nascimento, "Applications of deep learning techniques on NILM," Ph.D. dissertation, Alberto Luiz Coimbra Inst. Graduate Stud. Res. Eng., Federal Univ. Rio de Janeiro, Rio de Janeiro, Brazil, 2016.

[21] J. Kelly, "Disaggregation of domestic smart meter energy data," Ph.D. dissertation, Imperial College Sci., Technol. Med., Dept. Comput., Univ. London, London, U.K., 2017.

[22] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for nonintrusive load monitoring," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 2604–2611.

[23] M. D'Incecco, S. Squartini, and M. Zhong, "Transfer learning for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1419–1429, Mar. 2020.

[24] L. Massidda, M. Marrocu, and S. Manca, "Non-intrusive load disaggregation by convolutional neural network and multilabel classification," *Appl. Sci.*, vol. 10, no. 4, p. 1454, Feb. 2020.

[25] M. Kaselimi, N. Doulamis, A. Voulodimos, E. Protopapadakis, and A. Doulamis, "Context aware energy disaggregation using adaptive bidirectional LSTM models," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3054–3067, Jul. 2020.

[26] A. Harell, S. Makonin, and I. V. Bajic, "Wavenilm: A causal neural network for power disaggregation from the complex power signal," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 8335–8339.

[27] O. Parson, "Unsupervised training methods for non-intrusive appliance load monitoring from smart meter data," Ph.D. dissertation, Dept. Phys. Sci. Eng., Electron. Comput. Sci., Univ. Southampton, Southampton, U.K., 2014.

[28] Y. F. Wong, T. Drummond, and Y. A. Şekercioğlu, "Real-time load disaggregation algorithm using particle-based distribution truncation with state occupancy model," *Electron. Lett.*, vol. 50, no. 9, pp. 697–699, 2014.

[29] M. J. Johnson and A. S. Willsky, "Bayesian nonparametric hidden semi-Markov models," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 673–701, 2013.

[30] M. Aiad and P. H. Lee, "Unsupervised approach for load disaggregation with devices interactions," *Energy Buildings*, vol. 116, pp. 96–103, Mar. 2016.

[31] S. Makonin, F. Popowich, I. V. Bajic, B. Gill, and L. Bartram, "Exploiting HMM sparsity to perform online real-time nonintrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2575–2585, Nov. 2016.

[32] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Improving nonintrusive load monitoring efficiency via a hybrid programing method," *IEEE Trans. Ind. Informat.*, vol. 12, no. 6, pp. 2148–2157, Dec. 2016.

[33] C. Rottondi, M. Derboni, D. Piga, and A. E. Rizzoli, "An optimisation-based energy disaggregation algorithm for low frequency smart meter data," *Energy Informat.*, vol. 2, no. S1, p. 13, Sep. 2019.

[34] S. Dash, R. Sodhi, and B. Sodhi, "An appliance load disaggregation scheme using automatic state detection enabled enhanced integer-programming," *IEEE Trans. Ind. Informat.*, early access, Feb. 24, 2020, doi: 10.1109/TII.2020.2975810.

[35] C. C. Aggarwal, *Neural Networks and Deep Learning: A Textbook*, 1st ed. New York, NY, USA: Springer, 2018.

[36] R. L. Burden, J. D. Faires, and A. M. Burden, *Neural Networks and Deep Learning: A Textbook*, 10th ed. Boston, MA, USA: Cengage Learning, 2015.

[37] 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, 2011, pp. 1–6.

[38] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Sci. Data*, vol. 2, Mar. 2015, Art. no. 150007.

[39] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[40] B. Zhao, L. Stankovic, and V. Stankovic, "On a training-less solution for non-intrusive appliance load monitoring using graph signal processing," *IEEE Access*, vol. 4, pp. 1784–1799, 2016.

[41] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajic, "AMPds: A public dataset for load disaggregation and eco-feedback research," in *Proc. IEEE Electr. Power Energy Conf.*, Aug. 2013, pp. 1–6.

[42] H. Liu, Q. Zou, and Z. Zhang, "Energy disaggregation of appliances consumptions using HAM approach," *IEEE Access*, vol. 7, pp. 185977–185990, 2019.

[43] D. Li and S. Dick, "Whole-house non-intrusive appliance load monitoring via multi-label classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2749–2755.

[44] S. M. Tabatabaei, S. Dick, and W. Xu, "Toward non-intrusive load monitoring via multi-label classification," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 26–40, Jan. 2017.

[45] S. Singh and A. Majumdar, "Non-intrusive load monitoring via multi-label sparse representation-based classification," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1799–1801, Mar. 2020.

[46] R. Jia, Y. Gao, and C. J. Spanos, "A fully unsupervised non-intrusive load monitoring framework," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2015, pp. 872–878.

[47] M. Gaur and A. Majumdar, "Disaggregating transform learning for non-intrusive load monitoring," *IEEE Access*, vol. 6, pp. 46256–46265, 2018.

[48] L. C. Jain, *Recent Advances in Artificial Neural Networks*, 1st ed. Boca Raton, FL, USA: CRC Press, 2017.

[49] C.-Y. Liou, W.-C. Cheng, J.-W. Liou, and D.-R. Liou, "Autoencoder for words," *Neurocomputing*, vol. 139, pp. 84–96, Sep. 2014.

[50] Y.-H. Lin and M.-S. Tsai, "The integration of a genetic programming-based feature optimizer with Fisher criterion and pattern recognition techniques to non-intrusive load monitoring for load identification," *Int. J. Green Energy*, vol. 12, no. 3, pp. 279–290, 2015.

[51] A. Y. Ng, *CS294A Lecture Notes on Sparse Autoencoders*. Stanford, CA, USA: Stanford Univ., 2011. [Online]. Available: https://web.stanford.edu/class/cs294a/sparseAutoencoder_2011new.pdf

[52] A. Y. Ng, J. Ngiam, C. Y. Foo, Y. Mai, C. Suen, A. Coates, A. Maas, A. Hannun, B. Huval, T. Wang, and S. Tandon, *Unsupervised Feature Learning and Deep Learning Tutorial*. Stanford, CA, USA: Stanford Univ., 2011. [Online]. Available: http://ufldl.stanford.edu/tutorial/unsupervised/Autoencoders/

[53] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," Université de Montréal, Montreal, QC, Canada, Tech. Rep. 1316, Feb. 2008.

[54] X. Sun, N. M. Nasrabadi, and T. D. Tran, "Task-driven dictionary learning for hyperspectral image classification with structured sparsity constraints," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4457–4471, Aug. 2015.

[55] M. M. A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Inf. Sci.*, vol. 345, pp. 340–354, Jun. 2016.

[56] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, 2nd ed. New York, NY, USA: Springer, 2018.

[57] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. London, U.K.: Pearson, 2015.

[58] G. James, D. Witten, T. Haste, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R* (Springer Texts in Statistics), 7th ed. New York, NY, USA: Springer, 2017.

[59] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Netw.*, vol. 6, no. 4, pp. 525–533, Jan. 1993.

[60] K. Faceli, A. C. Lorena, J. Gama, and A. C. P. L. F. Carvalho, *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina*, 1st ed. Rio de Janeiro, Brazil: LTC, 2011.

SUELENE DE JESUS DO CARMO CORRÊA received the M.S. degree in computer science from the Federal University of Pará, Belém, Brazil, in 2012, where she is currently pursuing the Ph.D. degree in electrical engineering.

She is also an Assistant Professor with the Federal Institute of Education, Science and Technology of Pará, Castanhal, Brazil. Her research interests include computational intelligence and energy systems.

ADRIANA R. G. CASTRO received the M.S. degree in electrical engineering from the Federal University of Pará, Belém, Brazil, in 1995, and the Ph.D. degree in electrical engineering from the University of Porto, Porto, Portugal, in 2004.

She is currently an Associate Professor with the Federal University of Pará. Her research interests include computational intelligence, energy systems, and control of process.