



**QUEEN'S  
UNIVERSITY  
BELFAST**

## Effective identification of distributed energy resources using smart meters net-demand data

Moreno Jaramillo, A. F., Lopez-Lorente, J., Lavery, D. M., Martinez-del-Rincon, J., Morrow, D. J., & Foley, A. M. (2022). Effective identification of distributed energy resources using smart meters net-demand data. *IET Smart Grid*, 5(2), 120-135. Advance online publication. <https://doi.org/10.1049/stg2.12056>

**Published in:**  
IET Smart Grid

**Document Version:**  
Publisher's PDF, also known as Version of record

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

### **Publisher rights**

Copyright 2022 the authors.

This is an open access article published under a Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited

### **General rights**

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

### **Open Access**

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

## ORIGINAL RESEARCH

# Effective identification of distributed energy resources using smart meter net-demand data

Andres F. Moreno Jaramillo<sup>1</sup>  | Javier Lopez-Lorente<sup>2</sup>  | David M. Lavery<sup>1</sup>  |  
Jesus Martinez-del-Rincon<sup>1</sup>  | D. John Morrow<sup>1</sup> | Aoife M. Foley<sup>3</sup> 

<sup>1</sup>School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, UK

<sup>2</sup>PV Technology Laboratory, FOSS Research Centre for Sustainable Energy, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus

<sup>3</sup>School of Mechanical and Aerospace Engineering, Queen's University Belfast, Belfast, UK

## Correspondence

Andres F. Moreno Jaramillo, School of EEECS, Queen's University Belfast, Ashby Building, 125 Stranmillis Road, Belfast, Northern Ireland BT9 5AH, UK.

Email: [amorenajaramillo01@qub.ac.uk](mailto:amorenajaramillo01@qub.ac.uk)

## Funding information

Engineering and Physical Sciences Research Council, Grant/Award Numbers: EP/L001063/1, EP/M002837/1; Department for the Economy, Grant/Award Number: USI 110

## Abstract

International policies and targets to globally reduce carbon dioxide emissions have contributed to increasing penetration of distributed energy resources (DER) in low-voltage distribution networks. The growth of technologies such as rooftop photovoltaic (PV) systems and electric vehicles (EV) has, to date, not been rigorously monitored and record keeping is deficient. Non-intrusive load monitoring (NILM) methods contribute to the effective integration of clean technologies within existing distribution networks. In this study, a novel NILM method is developed for the identification of DER electrical signatures from smart meter net-demand data. Electrical profiles of EV and PV systems are allocated within aggregated measurements including conventional electrical appliances. Data from several households in the United States are used to train and test classification and regression models. The usage of conventional machine learning techniques provides the proposed algorithm with fast processing times and low system complexity, key factors needed to differentiate highly variable DER power profiles from other loads. The results confirm the effectiveness of the proposed methodology to individually classify DER with performance metrics of 96% for EV and 99% for PV. This demonstrates the potential of the proposed method as an embedded function of smart meters to increase observability in distribution networks.

## KEYWORDS

demand-side management, distributed energy resources, non-intrusive load monitoring, smart grids, supervised machine learning methods

## 1 | INTRODUCTION

The adoption of distributed energy resources (DER) such as distributed generation technologies (i.e. photovoltaic [PV] systems, battery energy storage systems and small-scale wind turbines) and low-carbon loads (i.e. electric vehicles [EV], electric heating and air conditioning units) has increased in the last decades [1]. This has been motivated by worldwide climate-led policies to reduce CO<sub>2</sub> equivalent (CO<sub>2</sub>e) emissions [2], such as the Kyoto Protocol (1997), the Paris Agreement (2015) and the United Nations Sustainable Development Goals (2015).

Rooftop PV systems and EV represent the DER with the highest presence in low-voltage distribution networks. Authorities' and public acceptance of PV systems has constantly

grown during the last decade based on benefits such as lower generation costs, increased electrical system flexibility and their contribution to a 'greener' power generation system [3]. For example, the rooftop PV capacity grew by 25% in Germany in 2020 with respect to 2019 [4]; in Spain, residential PV installations increased by 9% with respect to the previous year with 113 MW<sub>p</sub> of PV installed in the same year [5]; in the United States, 2020 saw the installation of 17.2 GW<sub>p</sub> of residential PV systems [6]. Similarly, the electrification of the transport sector is advancing rapidly with a clear growth of battery-driven vehicles for private and public transport. Countries in Europe, the USA and China are leading this transition and a significant growth of EV occurred in 2020, despite the negative impacts of the COVID-19 pandemic in

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *IET Smart Grid* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

several economic sectors. For example, in 2020, 75% of cars sold in Norway were EV while in the United Kingdom, over 108,000 battery EV (BEV) and 67,717 plug-in EV (PHEV) were sold, tripling the EV sales share with respect to 2019. In China, around 1.16 million EVs were sold during 2020. Simultaneously, several nations have set targets to ban sales of internal combustion engines (ICE) during the next decades, for example, Norway in 2025, and Iceland, the Republic of Ireland, Israel, the Netherlands, and the UK by 2030, among others [7].

In the low-voltage side of distribution networks, utility companies are facing an increasing installation of PV distributed generation. Additionally, with higher numbers of EV owners and limited control over residential charging points, there are unintended consequences of clustering in particular neighbourhoods [8]. The problem is compounded with a deficient record keeping of existing EV and PV systems and the fact that there is generally no real-time monitoring of these technologies. So, utility companies are aware of these problems, but they lack desirable information such as the electrical network connection of DER and what their capacities are [9]. Small-scale distribution generation does not only contribute to reducing power demand from a household but also to supplying power during peak times when generation is higher than the electric load. This causes issues in the network such as reverse power flow, sharper load peaks, system imbalances, voltage and frequency oscillations, power quality issues, and transformer ageing among others in low-voltage distribution networks [10, 11]. The large power and energy demanded by EV charging infrastructure in secondary distribution networks can also cause technical constraints in the electrical system. This is due to the connection of EVs in domestic electrical infrastructure originally designed to supply power to lower loads for which the electrical network was planned. As a result, issues such as large peak load demand, system imbalances, higher harmonic distortions, just to mention but a few, have increased in the low-voltage side of the electrical network [12].

Several measures have been proposed internationally in the industrial and academic environments for these technical challenges. Curtailment, static var compensators, robust optimization and solid-state transformers are some of the measures proposed to reduce the negative impact of distributed generation in low-voltage networks [13, 14]. Demand-side management alternatives have been also proposed for EVs considering a dual purpose to act as loads and as a resource for power system backup. Measures such as variable pricing schemes, vehicle to grid (V2G), or vehicle to home (V2H) have been also assessed [15]. Nevertheless, none of the presented measures has served utility companies to implement a simple and cost-effective solution replicable globally under the mass deployment of DER in distribution networks.

Non-intrusive load monitoring (NILM) techniques contribute to smart grids in intelligent load identification, bill purposes and demand-side management [16]. These methods have been widely studied since the first NILM algorithm was proposed by Hart [17]. The purpose of these techniques is to distinguish loads from aggregated measurements.

In [18], the authors previously presented a state of the art covering load modelling and NILM methods for DER electrical profile identification/forecasting. Relevant research studies were filtered finding the majority of NILM focussed on conventional loads rather than DER electrical profiles. From over 127 research studies analysed, 4 were focussed on NILM for EV and 5 for PV electrical profiles disaggregation, none of them combining both. Therefore, this paper introduces an innovative NILM method for the identification of DER from aggregated measurements using smart metre net-demand data. The method is developed upon conventional machine learning algorithms such as k-nearest neighbour (kNN), random forest (RF), and an artificial neural network (ANN) known as multilayer perceptron. The novelty of the proposed algorithm consists of implementing a NILM method to identify the presence and thus the electrical network connection point of DERs, and also to quantify the capacity of EV and PV electrical signatures in the customer side. The classification and regression models are evaluated using data from the ‘Pecan Street – Dataport’ public dataset [19].

The main contribution of this research consists of evaluating the proposed NILM method for the identification of the characteristic electrical profiles produced by EV and PV systems in the residential sector. This helps to increase the observability of low-voltage distribution networks, a sector presently lacking observability and in need of advanced grid monitoring [20].

This study builds on preliminary research in [21] and [22]. In [21], a regression model based on RF and kNNs was proposed. The method was based on the disaggregation of a PV system generation profile from aggregated household measurements of house C recorded in the Smart star project dataset (denoted as *Smart\** project) [23]. In [22], a NILM classification model based on support vector machines (SVM) and statistical features of electrical measurements was proposed. Data acquired with an OpenPMU instrument [24] was used to classify the presence of a rooftop PV system from the aggregated load of a residential dwelling in the UK. The main differences and new contributions in this research are: (i) the inclusion of both classification and regression approaches; (ii) the identification of EV and PV electrical profiles using the same method; (iii) the addition of ANN as the classification and regression model; (iv) the inclusion of statistical variables from windows of the active power as the input of the classification and regression models; (v) a deeper evaluation of the NILM method performance; (vi) the usage of the Pecan Street household loads; and (vii) the training and validation on a larger dataset comprising 1 month of data from 18 different households.

The remaining study is divided into seven sections. In Section 2, relevant research studies are discussed and NILM methods focussed on DER load identification are summarised. Section 3 describes the proposed methodology and introduces the main concepts of the NILM experimental method. Section 4 reports the results of the classification and regression model for DER identification. In Section 5, a sensitivity analysis is presented focussing on the evaluated metrics and

processing times. The paper continues in Section 6 with the benchmarking and validation of the model against other literature and datasets. A discussion is included in Section 7 to provide a comprehensive assessment of strengths and limitations of the proposed NILM method. Finally, Section 8 presents the conclusions and lines of future work available to this study.

## 2 | WORK TO DATE

NILM techniques based on machine learning methods proposed to date can be classified as supervised or unsupervised based on the information provided to the model during training and testing.

Supervised NILM methods use labelled data; thus, the class of an expected output is known. One of the main disadvantages of these techniques consists of acquiring labelled data, which requires monitoring not only aggregated loads but also specific ones. In contrast, unsupervised methods using cluster techniques, such as k-means, deep learning, Hidden Markov Models (HMM) and variations thereof do not require labelling. The disadvantage of these methods is their higher complexity compared with supervised techniques [18].

NILM techniques have been widely studied for the identification of conventional loads yet limited work has considered these approaches applied to DER identification. However, the increasing penetration of DER in low-voltage networks has changed this trend and NILM for DER identification is gaining relevance in the scientific literature. In [25], the authors presented an event-based NILM method to classify EV and air conditioning units among aggregated measurements from a household. The method was based on statistical features from active power records reaching recall and precision metrics of 93% and 88%, respectively. In [26], a NILM regression method was proposed to separate appliances power contribution from individual houses in the USA. Among the loads, power consumption from EV and conventional electrical appliances was disaggregated using an unsupervised technique. As a result, an

average root-mean-squared error of 1.101 was achieved for 20 households. Regarding solar generation systems, the authors previously proposed a classification approach to identify the contribution of a 3.5-kW peak (kWp) PV system [22]. SVMs and a principal component analysis were implemented as a supervised NILM method to separate PV generation from conventional loads, obtaining an  $F_1$  score of 96%. In [27], a regression model was developed to estimate PV power generation and PV module orientation. Based on a deep learning approach, the authors used a large dataset of over 1000 customers to identify the size of PV generation systems. The unsupervised approach contributed to achieving a mean average percentage error of 2.09% in the disaggregation of PV power generation. The mentioned and recent research studies focussed on the disaggregation of EV, and PV are summarised in Table 1.

Overall, it can be observed that the literature has focussed on assessing the penetration of either EV or PV and the study of both DER combined is seldom addressed. An initial investigation for the classification was done in the IEEE European Low-Voltage Test Feeder focussing on aggregated measurements in the low-voltage side of the distribution network [32]. However, the lack of suitable monitoring on low-voltage distribution electrical systems could make the implementation of such approaches difficult in the near future [30]. Therefore, the work in this study improves upon previous studies by addressing multiple DER (i.e. EV and PV) and by approaching the classification and regression problem from smart metre net-load demand data.

## 3 | PRINCIPLE AND METHOD

Aggregated measurements of an electrical system sum up individual contributions from each load connected to the circuit. This can be represented as in Equation (1).

$$P(t) = \sum_{i=1}^j P_i(t) \quad \forall t \in T \quad (1)$$

**TABLE 1** Comparison of existing non-intrusive load monitoring methods for DER electrical profile identification from smart metres net-demand data

DER	Method	Classification	Regression	Supervised	Unsupervised	Ref.
EV	Modified ICA	x			x	[12]
EV	MAD-SW	x		x		[25]
EV	RF - PCA	x		x		[28]
PV	SVM-SW	x		x		[22]
PV	Agglomerative clustering—SOM	x			x	[29]
PV	Analytical	x	x	x		[30]
PV	CSSS		x	x		[31]
EV	Conditional hidden semi-Markov model		x		x	[26]
PV	Deep neural network (DNN)		x		x	[27]

Abbreviations: CSSS, contextually supervised source separation; ICA, independent component analysis; MAD, mean absolute deviation; PCA, principal component analysis; SOM, self-organising maps; SVM, support vector machines; SW, sliding window.

where  $P(t)$  represents the net load in the time  $t$ ,  $P_i$  the power consumption of the  $i$ th load in a house with  $j$  loads, and  $T$  the simulation horizon. The value of  $T$  can be computed as in Equation (2).

$$T = \text{Days of data} * 1440 \frac{\text{min}}{\text{day}} * \text{houses} \quad (2)$$

The proposed algorithm is based on statistical features computed from windows of aggregated measurements, which are used as inputs of the chosen machine learning techniques to identify DER electrical profiles. The NILM method consists of six stages as in the flowchart illustrated in Figure 1.

### 3.1 | Data acquisition and interpretation

This research has been developed using the ‘Pecan Street – Dataport’ dataset [19]. Pecan Street is one of the largest public

databases available for research studies. It includes data from the electrical, water and transport sector with more than a thousand residential dwellings and businesses. Volunteers from mainly four locations in the USA contributed to developing this dataset used for research and product testing purposes. Additionally, it includes information from about 250 houses with PV generation and 65 load profiles from EV.

In this research study, aggregated loads from 18 houses located in Austin, Texas are used. These houses were selected within the dataset among those with PV and EV because they have different EV profiles, which add higher variability to the input dataset. The chosen profiles included EV owners with vehicles such as Chevrolet Volt, Mitsubishi i-MiEV Nissan Leaf, Tesla Model S. These vehicles draw active power between 3.5 and 10 kW at each of their respective houses. Among the selected load profiles, 10 had installed rooftop PV systems on their premises with a PV generation capacity in the range of between 3 and 12 kW peak. Active power measurements are recorded every minute for a period of 1 month in 2016.

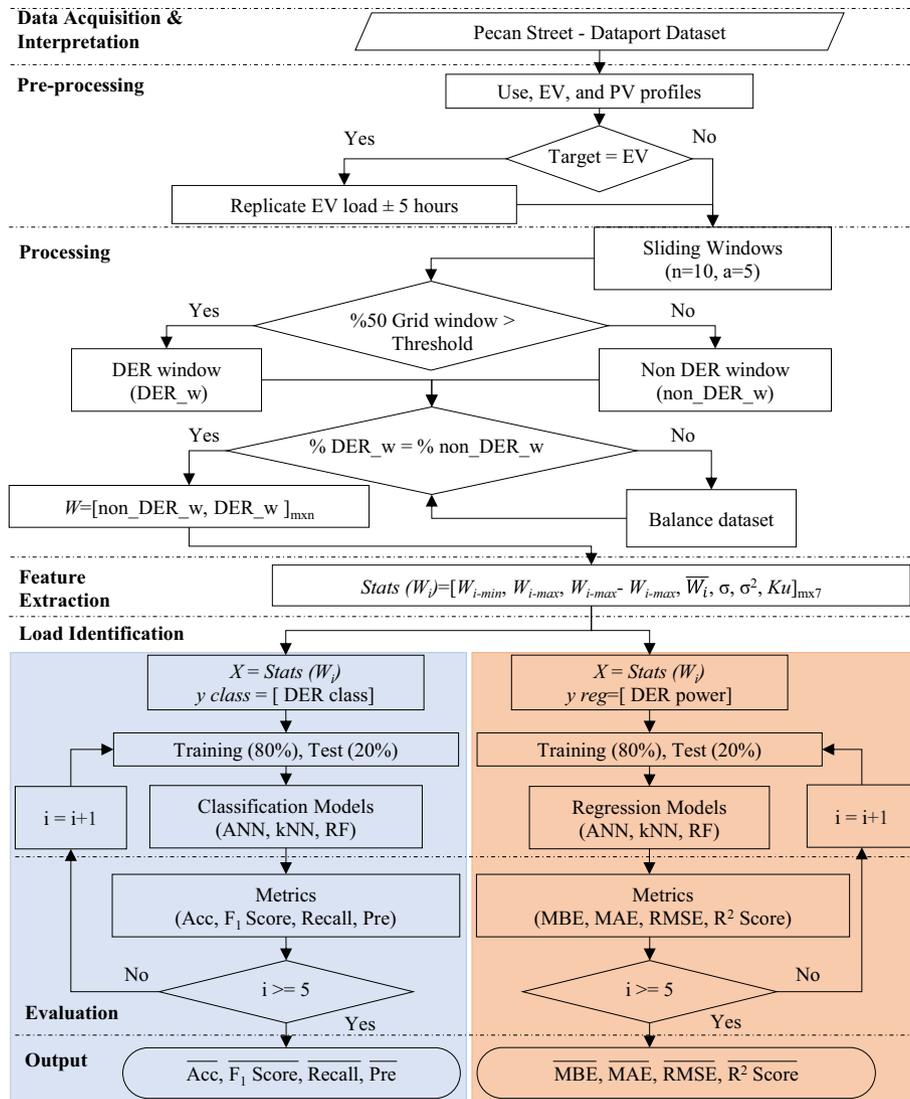


FIGURE 1 Flowchart of the proposed non-intrusive load monitoring method

### 3.2 | Data pre-processing

The Pecan Street dataset is filtered to select DER electrical profiles (PV and EV) and aggregated household load from the properties studied. This data is required for the training and testing stages. Pecan Street also includes individual load profiles from conventional devices (i.e. kitchen appliances, electronics, lighting etc.). In practice, the utility implementing this method is likely to only have the aggregated household load, so these individual load profiles are not used in this study since they are already contained within the aggregate data. As a result, four electrical profiles are obtained as illustrated in Figure 2 for house 9647.

It was found that EV owners charge their vehicle for about 3 h each day (between 10 and 21 kWh depending on the amount of energy the car needs), representing one-eighth of the existing data. Therefore, it is sensible to simulate additional EV charging patterns by adding the EV profile to the data at other times of day (5 h in advance and 5 h ahead) whenever EV is the objective of the proposed NILM. This provides richer and more diverse information to be used as input of the classification and regression algorithms. In contrast, the PV profiles are largely cyclical lasting several hours per day and with no generation at night times. Thus, PV generation profiles are used as originally reported in the dataset.

### 3.3 | Data processing

The proposed method utilises sliding windows, which is a technique used for the identification of events in small periods of time within a dataset. This permits the division of the dataset into several parts to be trained and tested, enabling real-time identification [33] instead of having to wait the full day to get the load profile. The sliding-window technique relies on two main factors: the size of the window ( $n$ ) and the shifting factor ( $a$ ) between consecutive windows. Each window is formed as in Equation (3).

$$W_i = [P(i * a), \dots, P(i * a + n - 1)] \quad (3)$$

where

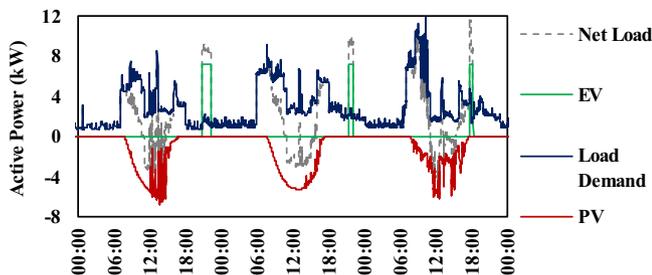


FIGURE 2 House 9647 electrical profiles from load demand, photovoltaic generation, electric vehicles power consumption, and the net-load demand from the 9th to the 11th of January 2016

$$0 \leq i \leq \frac{T}{a} - n = m \quad (4)$$

Consequently, values on each window are analysed to determine if the window includes relevant information from a particular DER or mainly load demand from conventional loads, so it is used in training as a positive or negative example. To do so, a threshold value of 0.1 kW is used to filter noisy measurements from EV load demand or PV generation and to label each window as ‘DER’ or ‘non\_DER’, respectively. Therefore, if the recorded DER power is lower than the set threshold value for more than half of the window size, it is labelled as not having the targeted DER. In contrast, windows presenting a higher influence of DER in the house's net load are set as containing the DER. The labelling of each window is defined in Python as in Figure 3.

The labelling of windows into classes facilitates the identification of how unbalanced the training data is. Subsequently, the number of ‘DER’ and ‘non\_DER’ windows are matched by randomly discarding those from the predominant class. This enables the creation of a balanced dataset with both classes equally represented and contributes to reducing the probability of overfitting machine learning algorithms to one of the classes.

### 3.4 | Feature extraction

The time series created as windows of  $n$  samples represents a single characteristic (aggregated net load) in the time domain of the aggregated household measurements. Thus, statistical variables are extracted from each window ( $W_i$ ) to reduce noise in the net load signal, optimise computational times, and provide machine learning models with several time-domain features [22, 34]. Namely, minimum ( $W_{i-\min}$ ), maximum ( $W_{i-\max}$ ), the maximum minus the minimum of  $W_i$  ( $W_{i-\max} - W_{i-\min}$ ), mean ( $\overline{W}_i$ ), variance ( $\sigma^2$ ), standard deviation ( $\sigma$ ) and kurtosis ( $K_\mu$ ) are obtained from each window. The mathematical expressions to derive these features are defined as follows [22, 34].

$$\overline{W}_i = \frac{1}{n} \sum_{j=0}^{n-1} W_i(j), \text{ for } 0 \leq i \leq m \quad (5)$$

$$\sigma^2(W_i) = \frac{1}{n-1} \sum_{j=0}^{n-1} (W_i(j) - \overline{W}_i)^2 \quad (6)$$

```
for item in Wi:
    if item > Threshold:
        window_class += 1
if window_class >= n/2:
    Wi_class = 'DER'
else:
    Wi_class = 'non_DER'
```

FIGURE 3 Python code to label the  $i$ th window as ‘distributed energy resources (DER)’ or ‘non\_DER’

$$\sigma(W_i) = \sqrt{\sigma^2(W_i)} \quad (7)$$

$$K_u(W_i) = \frac{E[W_i - \overline{W_i}]^4}{\sigma^4} \quad (8)$$

The input matrix ( $X$ ), used for both classification and regression models, is then formed with the described features as it is expressed in (9).

$$\begin{aligned} \text{Stats}(W_i) &= [W_{i-\min}, W_{i-\max}, W_{i-\max} - W_{i-\min}, \overline{W_i}, \sigma, \sigma^2, K_u] \\ &\quad \text{for } 0 \leq i \leq m \\ X &= \text{Stats}(W_i)_{m \times 7} \end{aligned} \quad (9)$$

### 3.5 | Load identification

Processed data are used to train and test machine learning algorithms to obtain classification and regression models. In classification problems, the expected output ( $y_{\text{class}}$ ) is defined as a DER class (EV/PV or not EV/not PV). In contrast, regression models use active power as expected values ( $y_{\text{reg}}$ ). Since the proposed NILM method identifies only one DER at the time, two different models are trained, one for PV and one for EV prediction. Considering those models can be either regressors or classifiers, this results in a total of 4 different models being trained.

Subsequently, processed data is divided into training and test sets. In this case, cross-validation is implemented to avoid overfitting the system and to provide a realistic measure of the proposed method's performance. This technique also contributes to making efficient usage of the available data. A k-fold of 5 is set to split the processed data into 5 random smaller sets, from which 4 parts are used to train the model and 1 is used to test its performance. This process is repeated 5 times until each fold is used as a test set. To avoid overlapping windows in both training and test sets, the input data  $X$  and the expected output vector  $y$  have been created using the windows obtained from the processed dataset randomly. Performance metrics are provided as the average value of the results achieved on each fold. Each test is completed using three conventional machine learning algorithms: kNN, RF, and an ANN based on a multilayer perceptron.

**Neighbour-based methods (kNN)** establish the expected values based on the correlation of samples (called neighbours) surrounding the unknown value [21]. Since the method does not need to know the input data or characteristics, it is adaptable to predict complex loads, such as EV and PV systems from aggregated measurements [32]. The proposed method uses uniform weights to predict new samples; thus, known values are equally evaluated to compute a prediction [35].

**RF** defines its predictions based on an agreement between estimators (called trees). Each estimator, which can be seen as an upside-down tree, works separately to make a prediction considering training records [21]. The final answer is then

computed as the average of all predictions in the forest [33]. RF uses criterions such as Gini impurity and entropy to measure the quality of a sample classification on each node of the three [36]. Gini impurity is a metric based on erroneous classifications, which provides 1.0 for a misclassification and 0.0 for a perfect prediction. Entropy measures how well organised is a dataset [28].

A **multilayer perceptron (ANN)** with only one hidden layer is also proposed for the DER profile identification. ANNs are used in several research environments, such as computer vision [37]. These methods have been considered to present the highest performance for NILM methods [38]. In this work, a supervised multi-layer perceptron (MLP) is implemented with an input of dimension  $n$  (window size), a hidden layer of 100 neurons (default value of the scikit-learn library in Python), and output 1 (prediction). This method is trained using backpropagation and the activation functions are rectified linear units and identity for hidden and output layers, respectively. The theoretical background for ANN can be consulted in [35, 36], and [39].

### 3.6 | Evaluation metrics

The performance of classification and regression models is evaluated with different metrics. In classification approaches, Boolean values are compared to provide an efficacy of a method. True and false predictions of positive (i.e. a window containing DER) and negative values (i.e. a window without DER) are used as the base to calculate the performance of the NILM method. The used metrics are accuracy ( $A_{CC}$ ), recall ( $Re$ ), precision ( $Pr$ ), and  $F_1$  score, which are the most common metrics used among NILM studies. These metrics are defined as defined in Equations (10) to (13) [18].

$$A_{CC} = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

$$Pr = \frac{TP}{TP + FP} \quad (11)$$

$$Re = \frac{TP}{TP + FN} \quad (12)$$

$$F_1 \text{ Score} = 2 \frac{Pr \times Re}{Pr + Re} \quad (13)$$

where  $TP$ ,  $TN$ ,  $FP$ , and  $FN$  are true positive, true negative, false positive and false negative predictions, respectively.

Regression methods require metrics focussed on error, which is measured by comparing expected values with predicted results. The performance of the regression model is evaluated using the mean bias error (MBE), mean average error (MAE), root-mean-squared error (RMSE), and coefficient of determination ( $R^2$  score). Definitions of the implemented metrics for regression models are shown in Equations (14) to (17) [21, 40].

$$MBE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i) \quad (14)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (15)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (16)$$

$$R^2 \text{ Score} = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (17)$$

where  $y_i$  is the  $i$ th expected value,  $\hat{y}_i$  is the  $i$ th prediction,  $\bar{y}$  is the median value of the expected values vector ( $y_{\text{test}}$ ), and  $m$  is the size of  $y_{\text{test}}$ .

## 4 | EXPERIMENTAL SETUP AND RESULTS

### 4.1 | Description of dataset and initial conditions

One month of data, from selected houses with EV and PV electrical profiles from the Pecan Street dataset, has been used to train and test the machine learning models. This short time of data is prolonged when EV is selected as a target by making the shifting EV load profiles described in Section 3.2. This reduces dependency on large datasets for EV to test the proposed method and enables us to compare achieved results with the existing literature.

While smart metres frequently report in sampling rates between 1 min and 1 h [41, 42], recent research studies repeatedly use a variable window size between 1 s and 10 min. In this research study, window sizes of  $n = 5, 10, 15$  min and  $\alpha = 1$  min are initially evaluated matching most of the literature. This serves as an initial approach to analyse the impact of parameters' variation on each machine learning algorithm. Thus, the number of maximum iterations ( $\text{max}_i$ ), neighbours ( $k$ ), and estimators ( $r$ ) have been selected for ANN, kNN and RF using grid-search within the ranges provided in Table 2. The variation of these parameters allows the evaluation of their

**TABLE 2** Grid-search values for the parameters of each machine learning algorithm

Parameter value set	ANN ( $\text{max}_i$ )	kNN ( $k$ )	RF ( $r$ )
1	100	2	50
2	200	5	100
3	300	10	150
4	400	15	200
5	500	20	250

effect on the performance and processing times of the proposed method, a trade-off needed to be included for the real-time identification approach.

The proposed NILM method was tested using Python 3.8.8 and the machine learning library scikit-learn (version 0.24.1) on an Intel Core i7-8700 @ 3.20 GHz PC with a 16-GB DDR4 RAM. All experiments were run in the same machine to get comparable computational costs.

### 4.2 | Approach for DER classification

Results were clustered considering the target (EV or PV) and the machine learning technique in use. The selected classification metrics, Equations (10) to (13), provide the best score when results are closer to 1. When the metric score is over 0.5, it indicates that the model is able to recognise the input data to provide a prediction, rather than making a guess. The results are presented using box plots to provide a complete picture of the initial 900 simulations ( $90 \times 5$  per DER). Each box can be interpreted in four percentiles graphically denoted by error bars (each one representing a percentile or 25% of the results) and the box itself, where 50% of the values fall within. The line in the centre of the box indicates the median value of all results for each experiment. Outliers are plotted outside of the error bars.

Overall, RF exhibited better performance for the classification of both EV and PV than kNN and ANN as shown in Figure 4. If the results achieved between recall and precision are compared, in general, it can be said that the models tend to provide a higher number of false negative predictions. In the first scenario, defined for the classification of EV load profile, the reported metrics are between 75% and 97%, with an average result of 89%. Machine learning methods exhibited a consistent performance with ANN metrics presenting low variability but the lowest performance of the three methods. The classification of PV generation profiles exhibited better performance in average but higher deviation in performance metrics than those for EV. Evaluation metrics were obtained in the range of 78% and 99%, with a mean value of 92%. As it can be seen in Figure 4, with an overall  $F_1$  score of 94%, RF results outperformed the ANN and kNN machine learning methods, which reached an average performance of about 89% and 91% for the same metric.

### 4.3 | Approach for DER regression

The regression metrics MBE, MAE, RMSE, and  $R^2$  score are used to evaluate the performance of the proposed NILM method. The best possible scores for these metrics are 0 in the case of MBE, MAE, and RMSE while the  $R^2$  score indicates that a prediction is close to the expected value when the metric is 1.

Overall, consistent error-based metrics were achieved to predict EV load profiles presenting higher errors than those for PV. Two factors contributing to this are the negative power consumption observed in the smart metre when the PV generation is higher than the load demand in the house, and the

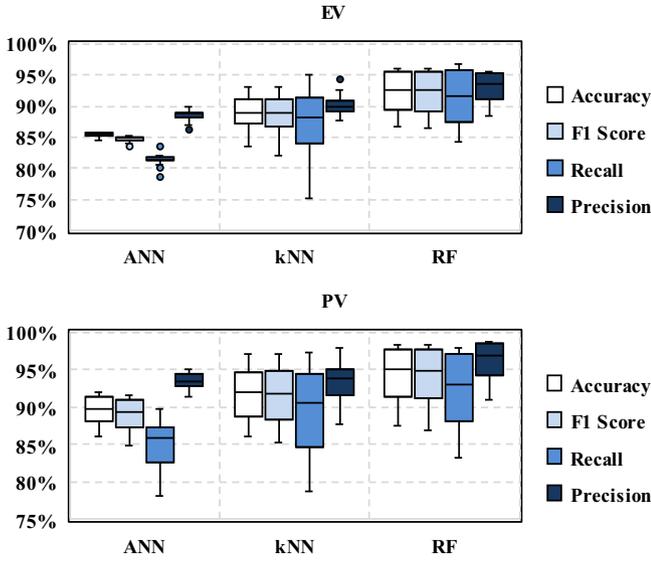


FIGURE 4 Classification metrics achieved per distributed energy resources

second one is related to the larger data available to train the model for PV than EV. Achieved results for the proposed NILM method as a regression model are illustrated in Figure 5.

The MBE indicates the overall bias of the model to estimate unknown samples. While positive values indicate the regression model over forecast new samples, negative ones indicate it is likely the model will underestimate them. In our case, the identification of EV projected a negative MBE, which represents a large bias to underestimate new samples. This can be caused by the similar characteristics of EV with other larger loads in the houses and the effect of high variability in PV generation profiles affecting the training stage of the identification models. In the case of PV, the metric was still negative but considerably close to 0. This means the model is likely to generate new predictions with low bias and these estimations might be slightly lower than the expected values.

The MAE reveals the average deviation of the prediction concerning expected values. The proposed method provided a larger error to identify EV than PV systems. In the case of EV identification, this metric presented an average value of 0.42 kW using RF, which represents 12% and 4% error for identification of the 3.5- and 10-kW EV load profiles. This deviation increased in the case of a kNN and ANN as regressors, in which case maximum average deviations of 16% and 24% were obtained, respectively, for 3.5-kW EV. In the case of PV systems, lower values were obtained for this metric indicating better performance to identify this DER. An average MAE of 0.39, 0.29 and 0.23 kW was reported for PV identification based on ANN, kNN and RF. This represents a deviation between 2% and 13% for the identification of the PV systems ranging between 3 and 12 kWp.

The RMSE and the  $R^2$  score are metrics related to how well the model can predict new samples. Significant error values are penalised by RMSE (in kW), which means the larger the

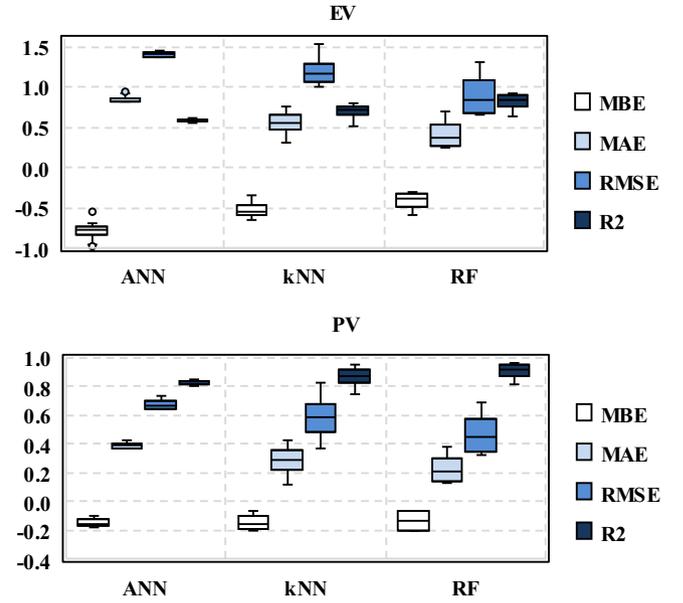


FIGURE 5 Regression metrics achieved per distributed energy resources

deviation of the residuals from the actual value, the larger the RMSE. Electric vehicle identification produced the largest RMSE values. Average RMSE values of 1.41, 1.20 and 0.90 kW were obtained for ANN, kNN and RF, correspondingly. This represents a deviation regarding EV power consumption in the selected houses between 9% and 40%. In the case of PV systems identification, better RMSE values were obtained with 0.68 kW for ANN, 0.58 kW for kNN as regression models while a score of 0.47 kW was obtained for RF. This represents a deviation regarding PV systems in the input dataset between 23% and 4%.

The  $R^2$  score provides a measure of how well the model makes new predictions with the best possible score of 1.0. This metric exhibited average scores around 0.59, 0.70 and 0.82 for the EV identification using ANN, kNN and RF, respectively. Therefore, even though the model works relatively well for the identification of EV, there is still room for improvement. Higher values of  $R^2$  score were reported for PV systems with overall scores of 0.83 for Ann, 0.87 for kNN, and 0.92 for RF.

## 5 | SENSITIVITY ANALYSIS

In this section, the variation of several parameters is analysed to evaluate their effect on the performance and processing times of the proposed NILM method, namely, the DER, window width ( $n$ ), and the machine learning parameter value set. The results are reported below considering each machine learning algorithm and the DER evaluated.

### 5.1 | Performance

Evaluation metrics for the classification approach present the model as a successful method to identify either EV or PV

systems. Since the  $F_1$  score provides a balanced representation of accuracy, recall, and precision, results for the  $F_1$  score are provided as a function of the DER and the proposed parameters of each machine learning technique. This is  $\max_i$ ,  $k$ , and  $r$  for ANN, kNN, and RF, respectively, as defined in Table 2. As an example, experimental results for a window width of  $n = 10$  are exhibited in Figure 6.

For this value of  $n$ , the  $F_1$  score oscillated between 85% and 91%. In general, the lowest performance was achieved for the identification of EV (dark blue) for the three algorithms. This lower performance can be understood from the negative effect caused by the high variability of PV generation and the mixed EV load demand when a small-capacity charger is used as it was shown in Figure 2. In contrast, disaggregation of PV systems exhibited better performance than their counterpart (Red). This is due to a large number of windows that are available for PV in training—since the PV generation lasts in average 10–12 h versus the shorter burst of few hours for EV—and the excess of power is delivered back to the grid, which is seen as a negative signal easier to identify by the algorithm during training. Looking at the performance of machine learning algorithms, kNN continuously reduced its performance for a  $k$  higher than five neighbours. Contrarily, ANN and RF slightly improve the performance of the NILM method with the increase in their parameters. Therefore,  $\max_i = 500$ ,  $k = 5$  and  $r = 250$  are selected as best parameters of each algorithm.

Once machine learning main variables are fixed, additional simulations are performed changing the window size within values in the range of  $n = [5, 10, 15, 30, 45, 60]$  (min). The classification results using these parameters are summarised per DER, window size, and machine learning algorithm as it is presented in Table 3. Despite overall classification scores were close within different window sizes, in general, it can be said that larger windows contributed to an increase in performance metrics with RF providing the best performance among the chosen machine learning techniques.

Consistently for the regression approach, the machine learning techniques provided the lower error for the above-mentioned parameters. As it was previously observed in the classification approach, predictions of the regression models were similar for the three evaluated window sizes. In general, all machine learning algorithms provided their best

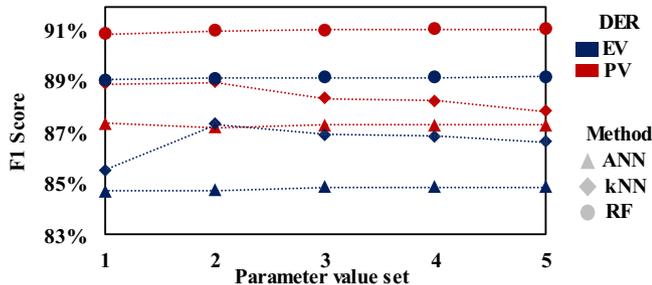


FIGURE 6 Performance of the proposed non-intrusive load monitoring classification algorithm based on evaluated target and selected machine learning technique

performance for 60-min windows. However, ANN did not present a constant outcome for the identification of EV load profiles. A summary of all 900 simulations ran for regression models are provided in Table 4 as a function of the scenario, the window width ( $n$ ), and the machine learning technique.

## 5.2 | Processing times

Processing times of the evaluated NILM method were in the order of the microseconds per processed window for the majority of the simulations. Each machine learning algorithm exhibited similar computational times to process data from EV and PV windows. Due to the implementation of statistical variables as inputs of classification and regression models, relatively constant times were observed between different sizes of the window. This is because the window size does not change the dimension of the input vector ( $X$ ) to train and test the algorithms, designed to have seven features.

Among the implemented machine learning methods, kNN demonstrated to be the fastest one for training and testing purposes providing results under 3 and 17  $\mu$ s per window, respectively. This method presented test processing times 12 and 5 times higher than training times for classification and regression approaches. In contrast, ANN and RF required longer computational times per window to train the model. In the case of ANN, the training process took up 1 ms for both classification and regression models. This approach projected the fastest times to make predictions requiring less than 3  $\mu$ s per window, this is less than 0.3% of the time to train the model. RF presented computational times of up to 500  $\mu$ s per window during the fitting stage for classification purposes. This time was improved for the classification of EV and PV windows, when up

TABLE 3 Overall  $F_1$  score of each machine learning algorithm as classifier in the proposed non-intrusive load monitoring method

Target	Window ( $n$ )	Artificial		
		neural network (%) ( $\max_i = 500$ )	kNN (%) ( $k = 5$ )	Random forest (%) ( $r = 250$ )
Electric vehicles	5	0.84	0.85	0.86
	10	0.85	0.87	0.89
	15	0.85	0.89	0.91
	30	0.85	0.91	0.94
	45	0.85	0.93	0.96
	60	0.85	0.93	0.96
Photovoltaic	5	0.85	0.86	0.87
	10	0.87	0.89	0.91
	15	0.88	0.91	0.93
	30	0.90	0.94	0.97
	45	0.91	0.96	0.98
	60	0.92	0.96	0.98

**TABLE 4** Overall performance of each machine learning algorithm as regression model in the proposed non-intrusive load monitoring

Target	Window ( $n$ )	ANN ( $\max_i = 500$ )				kNN (%) ( $k = 5$ )				RF (%) ( $r = 250$ )			
		MBE (kW)	MAE (kW)	RMSE (kW)	$R^2$	MBE (kW)	MAE (kW)	RMSE (kW)	$R^2$	MBE (kW)	MAE (kW)	RMSE (kW)	$R^2$
EV	5	-0.76	0.85	1.42	0.58	-0.63	0.73	1.41	0.58	-0.58	0.68	1.30	0.65
	10	-0.77	0.86	1.40	0.59	-0.53	0.62	1.29	0.65	-0.48	0.53	1.07	0.76
	15	-0.74	0.83	1.38	0.60	-0.48	0.55	1.20	0.70	-0.41	0.44	0.93	0.82
	30	-0.83	0.82	1.37	0.61	-0.47	0.43	1.06	0.77	-0.34	0.31	0.74	0.88
	45	-0.71	0.87	1.41	0.59	-0.44	0.39	1.01	0.79	-0.30	0.27	0.68	0.91
	60	-0.83	0.90	1.43	0.57	-0.45	0.38	1.01	0.79	-0.32	0.25	0.65	0.91
PV	5	-0.17	0.43	0.73	0.80	-0.19	0.41	0.76	0.78	-0.21	0.38	0.69	0.82
	10	-0.17	0.40	0.70	0.81	-0.19	0.35	0.67	0.83	-0.20	0.30	0.58	0.88
	15	-0.18	0.39	0.68	0.83	-0.18	0.30	0.61	0.86	-0.17	0.25	0.50	0.90
	30	-0.13	0.37	0.65	0.84	-0.14	0.22	0.49	0.91	-0.11	0.18	0.40	0.94
	45	-0.14	0.37	0.64	0.85	-0.08	0.18	0.43	0.93	-0.07	0.14	0.35	0.95
	60	-0.10	0.38	0.65	0.84	-0.10	0.16	0.40	0.94	-0.06	0.13	0.32	0.96

Abbreviations: ANN, artificial neural network; EV, electric vehicles; kNN, k-nearest neighbour; MAE, mean average error; MBE, mean bias error; PV, photovoltaic; RF, random forest; RMSE, root-mean-squared error.

to 60  $\mu$ s per window were required to predict a DER class from a particular window. However, the regression model based on RF provided the largest times during training, reaching up to 2 ms per window. Similarly, this method presented the largest computational time to disaggregate EV and PV power from windows of aggregated measurements with predictions projected in up to 100  $\mu$ s. Given that all those prediction times are in  $\mu$ s or less and that the minimum windows duration is 5 min, these results show how all the proposed methods can operate in real time, yielding a prediction for the window before the next load window is collected.

Computational times for the classification and regression model performance evaluation are illustrated per DER and machine learning algorithm in Figure 7.

### 5.3 | Sensitivity to algorithms' input features

In this section, we compare the effectiveness of the statistical features being used as input for our machine learning algorithm versus using the raw input load values. Thus, using machine learning model parameters defined in Section 5.1, classification and regression metrics for two sets of inputs are compared. Namely, windows of active power and statistical variables (see Equation 9) were used as inputs of the NILM method. For the window of the active power, the input matrix  $X$  is defined as

$$X_i = W_i \quad (18)$$

being  $W_i$  defined as per Equation (3).

Windows of 60 min were chosen as a case base to analyse the main difference in performance metrics of the proposed

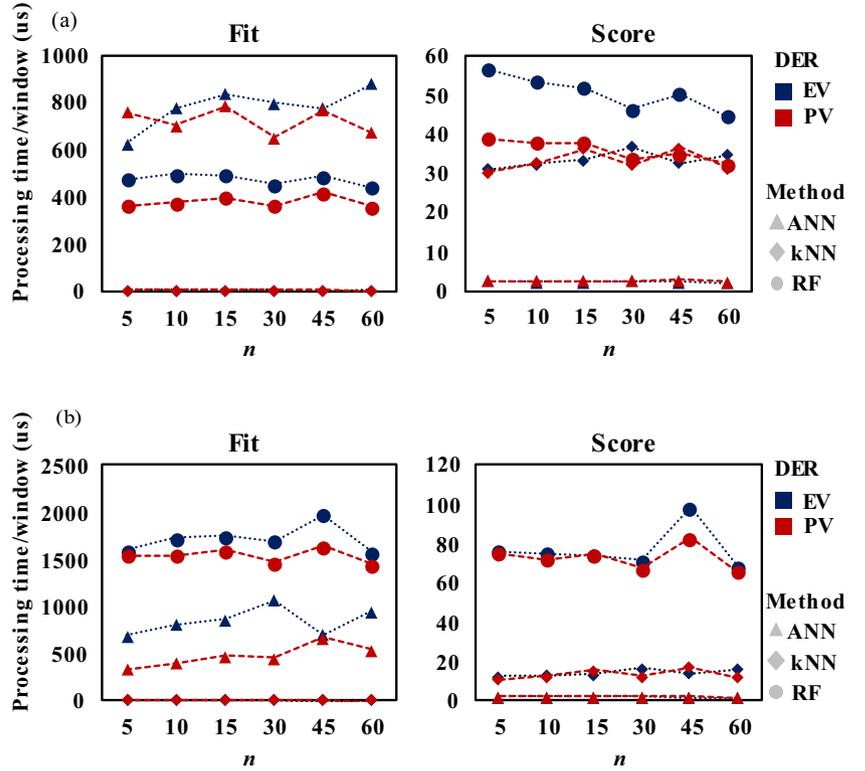
method. This window size was selected because, in general, the three machine learning techniques have presented their best performance for this value of  $n$  as described in Section 5.1. Achieved results are presented in Table 5.

Overall, the statistical variables as inputs of the NILM method did not generate a significant improvement of ANN performance metrics. In fact, a reduction in its performance was achieved for  $n = 60$ . Metrics such as MAE, RMSE and  $R^2$  score were reduced by 5%, 1% and 2% for EV, and 3%, 4% and 1% for PV. The MBE was also further located from 0 in about 23% for EV, but this metric was improved in almost 50% for PV. Nevertheless, kNN and RF projected a considerable increase of metrics in the classification as well as a substantial reduction of the deviation from EV and PV electrical profiles projections. While an improvement of about 4% was achieved for the  $F_1$  score for both DER using kNN, the other metrics presented improvements over 10% for EV and 20% for PV. RF also exhibited a significant increase in metrics such as the  $F_1$  score with 3% and 2% improvement for EV and PV identification, respectively. The  $R^2$  score was outperformed in about 4% for PV but a considerable increase of 15% was achieved during EV identification.

## 6 | BENCHMARKING AND VALIDATION

The performance of the proposed NILM technique is compared to other methods available in the literature. The developed NILM model is also tested in other publicly available datasets from the literature to evaluate its adaptability to identify DER from electrical profiles different from the ones used for training purposes.

**FIGURE 7** Processing times per window of each machine learning technique evaluated in the proposed non-intrusive load monitoring method for classification and regression purposes. (a) Processing times per distributed energy resources (DER) in the classification. (b) Processing times per DER in regression



**TABLE 5** Effect of statistical variables as inputs of the non-intrusive load monitoring method

Target	Metric	Windows ( $W_i$ )			Stats ( $W_i$ )		
		ANN	kNN	RF	ANN	kNN	RF
EV	F <sub>1</sub> score	0.85	0.89	0.93	0.85	0.93	0.96
	MBE	-0.68	-0.53	-0.50	-0.83	-0.45	-0.32
	MAE	0.86	0.52	0.50	0.90	0.38	0.25
	RMSE	1.42	1.17	1.00	1.43	1.01	0.65
	R <sup>2</sup>	0.58	0.72	0.79	0.57	0.79	0.91
PV	F <sub>1</sub> score	0.89	0.92	0.96	0.92	0.96	0.98
	MBE	-0.20	-0.14	-0.17	-0.10	-0.10	-0.06
	MAE	0.37	0.22	0.22	0.38	0.16	0.13
	RMSE	0.63	0.50	0.44	0.65	0.40	0.32
	R <sup>2</sup>	0.85	0.91	0.93	0.84	0.94	0.96

Abbreviations: ANN, artificial neural network; EV, electric vehicles; kNN, k-nearest neighbour; MAE, mean average error; MBE, mean bias error; PV, photovoltaic; RF, random forest; RMSE, root-mean-squared error.

### 6.1 | Comparison with existing literature

Relevant research studies regarding the classification and the identification of both EV and PV systems have been compared with the proposed NILM method. To provide a comprehensive analysis, achieved metrics are presented for the proposed method as the classification and regression model using

$n = 60$  min and  $r = 250$  for RF. This is because as it was illustrated in Figure 6 and Table 4, RF outperformed both ANN and kNN for the evaluated scenarios for classification and regression approaches. However, since the experimental settings and dataset are different from those in other studies, this is not a direct comparison but a reference of how well the method performed regarding other methods.

Most of the relevant references used in this research study to compare achieved results are focussed on the DER active power from the Pecan Street Dataset, which contains both EV and PV of several houses. The exceptions are [22] (OpenPMU [24] dataset with electric current as feature) and [21] (*Smart\** Project dataset). The main characteristics and presented performance metrics of NILM methods developed for the classification of EV or PV systems are summarised in Table 6.

As it can be seen, other authors have previously proposed relatively similar NILM model parameter configurations in terms of window size (3–10 min), step size (1 min), and performance. In our case, the largest window width is proposed to improve NILM method performance using only one month to train and test the algorithm. As a result, the system can effectively identify EV and PV classes from household aggregated loads with outstanding metrics overcoming previous research studies. In the case of PV identification, outstanding results were achieved in all metrics providing similar or better performance in comparison with existing methods.

Similar results were achieved for the implementation of the proposed method as a regression model for EV and PV disaggregation. RF metrics were consistently used to compare

Ref.	Parameters			Houses	Acc (%)	Pr (%)	Re (%)	F <sub>1</sub> (%)
	<i>n</i>	<i>a</i>	Data size					
Electric vehicles (EV) as target								
[12]	10 min	1 min	-	11	-	-	-	94
[28]	10 min	5 min	2 years	6	93	92	93	93
[25]	3 min5 min	1 min	15 days	1	-	88	93	90
Proposed	60 min	1 min	30 days	18	96	96	97	96
Photovoltaic (PV) as target								
[22]	1 s	0.5 s	5 days	1	96	99	93	96
[29]	Week, month, year	-	4 years	*	99	-	-	-
Proposed	60 min	1 min	30 days	18	99	99	99	99

\*1085 houses, 269 PV, 104 EV.

achieved results with other sources. As it was proposed in the classification approach, results for the identification of EV and PV using 30 days of data from the Pecan Street dataset are compared with the existing literature. These results are presented and contrasted with other sources in Table 7.

Regarding methods proposed in the current literature, the identification of EV power consumption from aggregated measurements provided a better RMSE score than the one presented in [26]. In this source, the achieved normalised RMSE represented a deviation of 28% with respect to the maximum power of the selected EV. In our case, the reached RMSE oscillated between 26% and 9% for EV load profiles within the range of 3.5 and 10 kW. In addition, the performance of the system to disaggregate PV power generation is consistent with the literature. The identification of PV generation profiles projected an RSME deviation between 16% and 4% for installed capacities from 3.5 and 12 kWp. This is consistent with the achieved MAE, which exhibits errors representing up to 8% of the original PV signal. Thus, results presented by authors in [27] and [30] are improved in this case. The  $R^2$  score also proves the proposed method to be more accurate than previous results presented in [21].

## 6.2 | Model generalisation: comparison to other datasets

To evaluate the adaptability of the proposed method and remove any bias from the input data, external datasets have been used as testing sets, namely, 3 days of house 4336 from Pecan Street and 3 days of house C from the *Smart\** Project [23]. Considering the results achieved during the cross-validation stage, the model parameters have been set to  $n = 60$ ,  $\max_i = 500$ ,  $k = 5$ , and  $r = 250$ . The training set was selected as 30 days of the 18 houses used in sections 4 and 5. Additionally, the adequation of the external data to create the test set was performed as described in Section III. However, an initial step was completed for House C of *Smart\** Project to convert the electric current of the PV system into power,

**TABLE 6** Proposed non-intrusive load monitoring method for distributed energy resources identification using classification models versus the existing literature

which is described in [21]. The NILM method was evaluated using ANN, kNN, and RF. Achieved results are shown in Table 8.

Similar to previous results, RF provided the best metrics for the classification and the identification of EV and PV profiles from unknown testing sets. In the evaluated scenarios, the proposed NILM model performed well to obtain classes from net load demand measurements. However, a slight reduction in performance metrics was observed in general, which is expected. In the case of PV power disaggregation, the performance of the system provided a considerable deviation between predictions and expected values. It can be said that the errors are relatively small if one analyses metrics such as MAE, MSE, and RMSE. However, the  $R^2$  score (which should be closer to 1 in the best cases) indicates a low performance in the case of kNN and an average performance of ANN and RF. Furthermore, poor metrics were obtained for EV identification. This is consistent with the difficulty to identify the EV load profile using only active power, which can be similar to large loads or the peak demand of other loads at certain times, for example, washing machines.

## 7 | DISCUSSION

The implementation of the balanced data step in the development of the proposed NILM method contributed to improving overall performance for the DER identification in processing times and performance. The method performed up to 8 times faster for the classification and about 3 times faster for regression purposes when using balanced data as input of the machine learning algorithms. The discrimination of windows into containing or not containing the targeted DER led to training and test sets, which were smaller in size and also richer in information. This is because the dominant window class is reduced to make them match in size to the ones with less appearance in time. Thus, the data passed to the machine learning algorithms contain the same percentage of windows with the influence of DER power consumption/generation as

**TABLE 7** Proposed non-intrusive load monitoring method for distributed energy resources identification using regression models versus existing literature

Ref.	Method	Model parameters			Metrics
		Window ( $n$ )	Step ( $a$ )	Input data size	
EV as target					
[26]	k-means + MNLR + CHSMM	60 min	60 min	1–3 months (20 houses)	RMSE 1.101 NRMSE 0.28
Proposed	SW + RF + statistics	60 min	1 min	30 days (18 houses)	MAE 0.23 RMSE 0.60 $R^2$ 0.92
PV as target					
[30]	Analytical	15 min	-	1 year (52 houses)	RMSE 20%–50%
[27]	DNN	1–60 min	1 min	1 year (1300 houses)	MAE 0.15
[21]	kNN – RF	10 min	1 min	3 months (1 house)	MAE 0.12 $R^2$ 0.92
Proposed	SW + RF + statistics	60 min	1 min	30 days (18 houses)	MAE 0.11 RMSE 0.29 $R^2$ 0.97

Abbreviations: ANN, artificial neural network; CHSMM, conditional hidden semi-Markov model; DNN, deep neural network; EV, electric vehicles; kNN, k-nearest neighbour; MAE, mean average error; MBE, mean bias error; MNLR, multinomial logistic regression; PV, photovoltaic; RF, random forest; RMSE, root-mean-squared error; SW, sliding windows.

**TABLE 8** Performance of the proposed model for the identification of EV and PV profiles from external datasets

Distributed energy resources Method	EV			PV		
	ANN	kNN	RF	ANN	kNN	RF
House 4336 (Pecan street)						
$F_1$ score	0.71	0.67	0.75	0.87	0.84	0.90
MAE	1.07	0.65	0.63	0.21	0.21	0.19
RMSE	1.50	1.56	1.38	0.32	0.4	0.33
$R^2$	0.37	0.31	0.47	0.72	0.59	0.71
House C (smart $\times$ project data used in [21])						
$F_1$ Score	-	-	-	0.92	0.89	0.95
MAE	-	-	-	0.19	0.28	0.2
RMSE	-	-	-	0.32	0.56	0.36
$R^2$	-	-	-	0.84	0.54	0.81

Abbreviations: ANN, artificial neural network; CHSMM, conditional hidden semi-Markov model; DNN, deep neural network; EV, electric vehicles; kNN, k-nearest neighbour; MAE, mean average error; MBE, mean bias error; MNLR, multinomial logistic regression; PV, photovoltaic; RF, random forest; RMSE, root-mean-squared error; SW, sliding windows.

those with only power consumption from conventional loads. For example, in the case of EV, there are more windows classified as not containing an EV than the ones having electrical information of the car in the charging stage. Reducing the number of windows having only information of power consumption from conventional loads helps training the machine learning algorithm to a more general case. This is confirmed

with the classification metrics exhibiting relatively low variation among them for fixed parameters setup (i.e.  $Acc$ ,  $Re$ ,  $Pr$ , and  $F_1$  score for a particular scenario). Additionally, the extraction of statistical features from windows of aggregated measurements contributed to reducing the negative impact caused by an increased number of samples per window. This is because the larger size of the input matrix  $X$  negatively impacts processing times of the implemented machine learning techniques. Thus, keeping constant the dimension of  $X$  contributed to increasing the performance metrics of the NILM method with relatively close processing times for different window sizes.

In terms of proposed machine learning techniques, the three algorithms provided consistent outcomes. First, kNN proved to be the fastest method during the training. This method yielded outstanding  $F_1$  scores of 93% and 97% for the classification of EV and PV electrical signatures. In the case of the ANN, remarkable results were achieved for the implemented multilayer perceptron ANN, providing slightly lower results than kNN but the shortest processing times to make predictions. An  $F_1$  score of 85% and 92% was achieved with this method to classify EV and PV from aggregated measurements. RF demonstrated the effectiveness to understand the non-linearities of EV and PV profiles overcoming the other two methods. In the classification, the  $F_1$  score was improved to 96% and 98% for EV and PV, respectively. Similar tendencies were observed in regression approaches, with RF overcoming kNN and this one surpassing ANN metrics.

Overall, results considering the same database for training and testing demonstrated how RF presents the best performance. This makes this method applicable to houses where

historical data is available. Although this performance is maintained when different datasets are used for training and test purposes, ANN also proved to be adaptable to this scenario. In this case, lower metrics were obtained for EV and PV but in general, the proposed method provides good predictions for classification and regression methods to disaggregate the DER electrical signature.

Although a different setup was used in this manuscript, the balanced data lead to accomplishing consistent metrics along with classification and regression approaches. As a result, the proposed method provided a fair improvement of scores projected in other research studies, with most of these outcomes being outperformed. Comparison of the findings with those of other studies confirms the beneficial performance of the proposed method as a classification and regression model.

The main limitation of the proposed model is based on the resolution of the measured data (1 min). This makes it difficult to implement advanced feature projection techniques based on transient characteristics of the electrical system under analysis. Since both EV and PV systems are inverter-based technologies, harmonic distortion on electric systems is a potential feature to be selected as input of a NILM method. This could contribute to improve the performance of the proposed method to identify DER from unknown data and more specifically, for regression purposes.

## 8 | CONCLUSION

The objective of this study was to propose an online NILM method for the classification and power disaggregation of EV and PV systems using net demand readings from smart metres. This study explored sliding windows and balanced data implementation as inputs of the NILM method. Results from 900 simulations were provided using box plots to summarise, as best as possible, all carried out simulations and make simpler the comparison with other methods. Additionally, comprehensive sensitivity analysis provided an assessment of parameters' effect on the proposed NILM model. This contributed to selecting the best parameters of the evaluated machine learning technique, which were used to make a comparison with previous research studies regarding EV and PV profile identification. Lastly, a model generalisation evaluation was included to allow testing the model in external datasets.

The proposed NILM method presents an effective technique to increase the information provided from smart metres to utility companies and distribution network operators. The better access to information improves the observability in the secondary, low-voltage distribution networks by enabling access to real-time information of DER presence. Similarly, the proposed NILM method contributes to an effective transition from conventional electric power systems to dynamic smart grids, which is a feature of the green energy transition and decarbonisation of the power generation and transport sectors. The implementation of the proposed NILM type algorithms

on smart metres themselves could provide utility companies with real-time power generation/consumption from DER. Overall, these types of methods could be used as input of advanced control systems to mitigate the negative impacts of uncoordinated DER growth in low-voltage distribution networks, which could increase the system's flexibility, reliability, and efficiency. This approach also provides opportunities to integrate other DER such as electric heating to support the electrification of the heating sector.

Future work could be completed using datasets with higher reporting rates (i.e. faster measurements of voltage and current to compute active, reactive power, and other electrical characteristics) allowing the implementation of the transient analysis. Furthermore, a more advanced setup for ANN could be used to improve the performance of the method, without a considerable compromise of processing times as it was observed in Section 5.2. Additionally, the Pecan Street dataset does not include solar radiation correlated to the PV generation. However, knowing the location of each household, simulations could be made to obtain this feature. This and other environmental factors such as ambient temperature could be used as input of the algorithm to improve the performance of the proposed model for PV identification. In the case of EV systems, inner variables from the vehicle itself (e.g. battery state of charge, harmonic content etc.) could be used as an additional input to increase the performance of regression models. Therefore, the improvement of the methods not only requires the development of strategies in the academic world but also the participation of industry to provide real and updated datasets to achieve sensible solutions.

## ACKNOWLEDGEMENTS

This work was supported by the DfE US-Ireland 110 'CRENCE', EPSRC 'CAPRICA' (EP/M002837/1), and EPSRC 'i-Give' (EP/L001063/1) projects. The authors would also like to thank Pecan Street Inc. for making freely available the energy data from DER systems for this research work.

## CONFLICT OF INTERESTS

The authors declare no potential conflict of interest.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analysed in this study. The input data used in this study are available in The Pecan Street - Dataport Dataset at <https://www.pecanstreet.org/>.

## ORCID

*Andres F. Moreno Jaramillo*  <https://orcid.org/0000-0003-1952-8465>

*Javier Lopez-Lorente*  <https://orcid.org/0000-0003-0032-1149>

*David M. Lavery*  <https://orcid.org/0000-0002-5697-0546>

*Jesus Martinez-del-Rincon*  <https://orcid.org/0000-0002-9574-4138>

*Aoife M. Foley*  <https://orcid.org/0000-0001-6491-2592>

## REFERENCES

1. Raouf Mohamed, A.A., et al.: Distributed battery energy storage systems operation framework for grid power levelling in the distribution networks. *IET Smart Grid*. 4, 582–598 (2021). <https://doi.org/10.1049/stg2.12040>
2. Hereher, M., El Kenawy, A.M.: Exploring the potential of solar, tidal, and wind energy resources in Oman using an integrated climatic-socioeconomic approach. *Renew. Energy*. 161, 662–675 (2020). <https://doi.org/10.1016/j.renene.2020.07.144>
3. Mahmoud, K., Lehtonen, M.: Comprehensive analytical expressions for assessing and maximizing technical benefits of photovoltaics to distribution systems. *IEEE Trans. Smart Grid*. Early Acce. 12, 1–12 (2021). <https://doi.org/10.1109/TSG.2021.3097508>
4. Solarwirtschaft, B.: Solar boom on private rooftops. *BSW Sol*, 1–5 (2021). <https://www.solarwirtschaft.de/en/2021/02/02/solar-boom-on-private-rooftops/>
5. UNEF Annual Report: Photovoltaic solar energy, opportunity for sustainability (2021). <https://www.unef.es/en/recursos-informes?idMultiMediaCategoria=18>
6. Feldman, D., Margolis, R.: H2 2020 solar industry update. *Natl. Renew. Energy Lab*. 1–72 (2021). <https://www.nrel.gov/docs/fy21osti/79758.pdf>
7. EIA: Global EV outlook 2021. EIA (2021). <https://www.eia.org/reports/global-ev-outlook-2021/policies-to-promote-electric-vehicle-deployment>
8. Akhtar, Z., et al.: Comparison of point-of-load versus midfeeder compensation in LV distribution networks with high penetration of solar photovoltaic generation and electric vehicle charging stations. *IET Smart Grid*. 2, 283–292 (2019). <https://doi.org/10.1049/iet-stg.2018.0193>
9. Mohamed, A.A.R., Morrow, D.J., Best, R.: The deployment of low carbon technologies in modern distribution networks, *Proc. 2019 IEEE PES Innov. Smart Grid Technol. Eur. ISGT-Europe*, 2 (2019). <https://doi.org/10.1109/ISGTEurope.2019.8905643>
10. Lopez-Lorente, J., Liu, X.A., Morrow, D.J.: Effect in the aggregated demand of solar-plus-storage prosumers in the residential sector. In: *CIREN - Open Access Proceedings*, pp. 24–26. IET, Berlin (2020). <https://digital-library.theiet.org/content/journals/10.1049/oap-cired.2021.0011>
11. Tévar-Bartolomé, G., et al.: Network impact of increasing distributed PV hosting: a utility-scale case study. *Sol. Energy*. 217, 173–186 (2021). <https://doi.org/10.1016/j.solener.2021.01.066>
12. Munshi, A.A., Mohamed, Y.A.R.I.: Unsupervised nonintrusive extraction of electrical vehicle charging load patterns. *IEEE Trans. Ind. Informatics*. 15, 266–279 (2019). <https://doi.org/10.1109/TII.2018.2806936>
13. Xu, X., et al.: Enhancing photovoltaic hosting capacity—a stochastic approach to optimal planning of static var compensator devices in distribution networks. *Appl. Energy*. 238, 952–962 (2019). <https://doi.org/10.1016/j.apenergy.2019.01.135>
14. Hooshyar, H., et al.: PMU-assisted overcurrent protection for distribution feeders employing Solid State Transformers. *Sustain. Energy Grids Netw.* 10, 26–34 (2017). <https://doi.org/10.1016/j.segan.2017.02.003>
15. Huty, T.D., et al.: Peer-to-peer electricity trading as an enabler of increased PV and EV ownership. *Energy Convers. Manag.* 245, 114634 (2021). <https://doi.org/10.1016/j.enconman.2021.114634>
16. Dinesh, C., et al.: Residential appliance identification based on spectral information of low frequency smart meter measurements. *IEEE Trans. Smart Grid*. 7, 2781–2792 (2016). <https://doi.org/10.1109/TSG.2015.2484258>
17. Hart, G.W.: Nonintrusive appliance load monitoring. *Proc. IEEE*. 80, 1870–1891 (1992). <https://doi.org/10.1109/5.192069>
18. Moreno Jaramillo, A.F., et al.: Load modelling and non-intrusive load monitoring to integrate distributed energy resources in low and medium voltage networks. *Renew. Energy*. 179, 445–466 (2021). <https://doi.org/10.1016/j.renene.2021.07.056>
19. Street Inc.: Pecan: Pecan Street Dataport, (2021). <https://www.pecanstreet.org/>
20. Johansson, P., Vendel, M., Nuur, C.: Integrating distributed energy resources in electricity distribution systems: an explorative study of challenges facing DSOs in Sweden. *Util. Policy*. 67, 101117 (2020). <https://doi.org/10.1016/j.jup.2020.101117>
21. Moreno Jaramillo, A.F., et al.: Photovoltaic power disaggregation using a non-intrusive load monitoring regression model. In: *IEEE Innovation Smart Grid Technology European*, pp. 1–6. Finland (2021). <https://doi.org/10.1109/ISGTEurope52324.2021.9640183>
22. Moreno Jaramillo, A.F., et al.: Non-intrusive load monitoring algorithm for PV identification in the residential sector. *IEEE*, pp. 1–6. Ireland (2020). <https://doi.org/10.1109/ISSC49989.2020.9180192>
23. Barker, S., et al.: Smart\*: an open data set and tools for enabling research in sustainable homes, p. 6 (2012). <https://lass.cs.umass.edu/papers/pdf/sustkdd12-smart.pdf>
24. Lavery, D.M., et al.: The OpenPMU project: challenges and perspectives. In: *IEEE Power Energy Society Generation Meeting*, pp. 21–25. IEEE: Canada (2013). <https://doi.org/10.1109/PESMG.2013.6672390>
25. Rehman, A.U., et al.: Event-detection algorithms for low sampling nonintrusive load monitoring systems based on low complexity statistical features. *IEEE Trans. Instrum. Meas.* 69, 751–759 (2020). <https://doi.org/10.1109/TIM.2019.2904351>
26. Ji, Y., Buechler, E., Rajagopal, R.: Data-driven load modeling and forecasting of residential appliances. *IEEE Trans. Smart Grid*. 11, 2652–2661 (2020). <https://doi.org/10.1109/tsg.2019.2959770>
27. Mason, K., et al.: A deep neural network approach for behind-the-meter residential PV size, tilt and azimuth estimation. *Sol. Energy*. 196, 260–269 (2020). <https://doi.org/10.1016/j.solener.2019.11.100>
28. Moreno Jaramillo, A.F., Lavery, D.M., Martinez del Rincon, J., Hastings, J., Morrow, D.J.: Croatia: Supervised non-intrusive load monitoring algorithm for electric vehicle identification. In: *IEEE International Instrumental Measurement Technology Conference*, pp. 1–6 (2020). IEEE. <https://doi.org/10.1109/I2MTC43012.2020.9128529>
29. Donaldson, D.L., Jayaweera, D.: Effective solar prosumer identification using net smart meter data. *Int. J. Electr. Power Energy Syst.* 118, 105823 (2020). <https://doi.org/10.1016/j.ijepes.2020.105823>
30. Tabone, M., Kiliccote, S., Kara, E.C.: Disaggregating solar generation behind individual meters in real time. In: *5th Conference System Built Environment Association for Computing Machinery (ACM)*, pp. 43–52. Shenzhen, China (2018). <https://doi.org/10.1145/3276774.3276776>
31. Vrettos, E., et al.: Estimating PV power from aggregate power measurements within the distribution grid. *Renew. Sustain. Energy*. 11, 1–18 (2019). <https://doi.org/10.1063/1.5094161>
32. Moreno Jaramillo, A.F., et al.: Identification of distributed energy resources in low voltage distribution networks. In: *IEEE Innovation Smart Grid Technology European*, pp. 1–6. IEEE (2021). <https://doi.org/10.1109/ISGTEurope52324.2021.9639971>
33. Suprême, H., et al.: Development of new Predictors based on the concept of center of power for transient and dynamic instability detection. *IEEE Trans. Smart Grid*. 9, 3605–3615 (2018). <https://doi.org/10.1109/TSG.2016.2636816>
34. Raurale, S., McAllister, J., Del Rincon, J.M.: Emg acquisition and hand pose classification for bionic hands from randomly-placed sensors. *IEEE Int. Conf. Acoust. Speech Signal Process.* 1105–1109 (2018). <https://doi.org/10.1109/ICASSP.2018.8462409>
35. Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning: data mining, inference, and prediction (2009). <https://link.springer.com/book/10.1007/978-0-387-84858-7>
36. Géron, A.: Hands-on machine learning with Scikit-Learn, Keras & TensorFlow, 2nd ed. O'Reilly Media, Inc, United States of America (2019). <https://doi.org/10.1201/9780367816377>
37. Kamenou, E., et al.: Multi-level deep learning vehicle re-identification using ranked-based loss functions. *Proc. Int. Conf. Pattern Recognit.* 9099–9106 (2020). <https://doi.org/10.1109/ICPR48806.2021.9412415>

38. D’Incecco, M., Squartini, S., Zhong, M.: Transfer learning for non-intrusive load monitoring. *IEEE Trans. Smart Grid.* 11, 1419–1429 (2020). <https://doi.org/10.1109/TSG.2019.2938068>
39. Hornik, K.: Approximation capabilities of multilayer feedforward networks. *Neural Netw.* 4, 251–257 (1991). [https://doi.org/10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T)
40. López Lorente, J., Liu, X., Morrow, D.J.: Worldwide evaluation and correction of irradiance measurements from personal weather stations under all-sky conditions. *Sol. Energy.* 207, 925–936 (2020). <https://doi.org/10.1016/j.solener.2020.06.073>
41. Farokhi, F.: Review of results on smart-meter privacy by data manipulation, demand shaping, and load scheduling. *IET Smart Grid.* 3, 605–613 (2020). <https://doi.org/10.1049/iet-stg.2020.0129>
42. Wang, Y., et al.: Review of smart meter data analytics: applications, methodologies, and challenges. *IEEE Trans. Smart Grid.* 10, 3125–3148 (2019). <https://doi.org/10.1109/TSG.2018.2818167>

**How to cite this article:** Moreno Jaramillo, A.F, et al.: Effective identification of distributed energy resources using smart meter net-demand data. *IET Smart Grid.* 1–16 (2022). <https://doi.org/10.1049/stg2.12056>