Journal of Energy Storage 61 (2023) 106784

Contents lists available at ScienceDirect

Journal of Energy Storage

journal homepage: www.elsevier.com/locate/est

Research Papers

Intelligent energy storage management trade-off system applied to Deep Learning predictions

Moisés Cordeiro-Costas [a,*], Daniel Villanueva [b], Pablo Eguía-Oller [a], Enrique Granada-Álvarez [a]

CINTECX, Universidade de Vigo. Rúa Maxwell s/n, 36310, Vigo, Spain
Universidade de Vigo, Industrial Engineering School. Rúa Maxwell s/n, 36310, Vigo, Spain

ARTICLE INFO

Keywords:
Building performance
Deep Learning (DL)
Deep Reinforcement Learning (Deep RL)
Electricity demand prediction
Intelligent Energy Management System (IEMS)
Photovoltaic production prediction

ABSTRACT

The control of the electrical power supply is one of the key bases to reach the sustainable development goals set by United Nations. The achievement of these objectives encourages a dual strategy of creation and diffusion of renewable energies and other technologies of zero emission. Thus, meet the emerging necessities require, inevitably, a significant transformation of the building sector to improve the design of the electrical infrastructure. This improvement should be linked to advanced techniques that allows the identification of complex patterns in large amount of data, such as Deep Learning ones, in order to mitigate potential uncertainties. Accurate electricity and energy supply prediction models, in combination with storage systems will be reflected directly in efficiency improvements in buildings. In this paper, a branch of Deep Learning models, known as Standard Neural Networks, are used to predict electricity consumption and photovoltaic generation with the purpose of reduce the energy wasted, by managing the storage system using Reinforcement Learning technique. Specifically, Deep Reinforcement Learning is applied using the Deep Q-Learning agent. Furthermore, the accuracy of the predicted variables is measured by means of normalized Mean Bias Error (nMBE), and normalized Root Mean Squared Error (nRMSE). The methodologies developed are validated in an existing building, the School of Mining and Energy Engineering located on the Campus of the University of Vigo.

1. Introduction

Buildings in residential and industrial sectors have been designed, in the past, for different situations, needs and ways of life than these days, without sufficient consideration of climatic conditions. These facts, represent the highest share of final energy in Europe, corresponding to 40 % of the final energy consumption and 36 % of the emissions [1]. Furthermore, energy demand in the building sector is growing, it has increased by 3 % annually since 2010 and this trend is expected to continue in coming years [2]. Responding to many emerging needs, inevitably requires a very significant improvement in the energy efficiency of built heritage. Thus, despite the increase in population and energy demand at a global level that is forecasts for 2040, an improvement in energy efficiency in buildings could reduce the emission of greenhouse gasses by >40 %, which allows to be in line with the goals of the Paris Agreement [3].

The reduction of emissions in buildings goes hand in hand with distributed generation, since it is an option that guarantees sustainable energy and is crucial to mitigate the uncertainties of the grid [4]. The continuous improvement in efficiency and costs promotes the use of solar photovoltaic (PV) generation as the most used solution to reduce consumption and emissions in building sector [5]. In fact, PV installations in buildings show the highest rate of growth in installed power in the world during the last decades and its share is increasing every year [6]. Therefore, PV installations in buildings are expected to play a key role in achieving climate targets [7].

The energy efficiency of buildings can be improved by 30 % without any structural change by optimizing the operation of loads and distributed energy [8]. The battery is recognized as a key element for real-time trade-off of energy supply and demand in buildings [1] and is projected to expand its annual growth rate in coming years [9]. The accurate predictive energy modeling of loads and production in buildings is essential to ensure the correct operation of the storage system, which will be reflected directly in energy efficiency improvements. Traditional modeling approaches as physical representations, i.e., TRNSYS, or mathematical programming methods, i.e., regressions, have been shown

* Corresponding author.
E-mail addresses: moises.cordeiro.costas@uvigo.es (M. Cordeiro-Costas), dvillanueva@uvigo.es (D. Villanueva), peguia@uvigo.es (P. Eguía-Oller), egranada@uvigo.es (E. Granada-Álvarez).

https://doi.org/10.1016/j.est.2023.106784
Received 1 June 2022; Received in revised form 26 January 2023; Accepted 27 January 2023
Available online 8 February 2023
2352-152X/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

to have some issues as computational cost or accuracy respectively, making it difficult to adapt to current needs. Thus, in a world where smart devices, such as smart meters, constantly produce high amount of data, reveals the need to work with efficient and accurate models, such as Artificial Intelligence (AI) techniques [10].

Currently, there is a great effort by the European Commission to show the need to improve the development of AI models. Thus, in the Coordinated AI Plan [11], this discipline is considered as a pillar that allowed Europe to boost its competitiveness and underpin the advancement of the areas of digitalization and green transition. The most extended branch of AI is the Machine Learning (ML), nevertheless, a more advanced category of ML models, known as Deep Learning (DL), has been gaining more presence in current research fields as autonomous driving [12], speech recognition [13] or healthcare [14]. These ones, replicate how the human brain behaves, being possible to obtain accurate predictions at a low cost by the identification of complex patterns in data. Common characteristic of DL models is the structure, from some input features, it is sought to extract a pattern in order to extrapolate the results to new data [15,16].

Accurate DL forecasting is a complex problem as this type of model depends on many parameters, i.e., bias, kernel or regularization, and hyperparameters, i.e., number of hidden layers, hidden layer activations or number of nodes per layer. The search for the appropriate DL configuration to achieve the best performance is a challenging problem and it is in continuous improvement [17]. Nevertheless, the capability of DL methods to extract and forecast patterns in building loads [18,19] and distributed generation [20,21] is widely demonstrated and with growing interest.

The configurations employed in this manuscript to forecast electricity consumption and PV production take into account a set of different parameters and hyperparameters to guarantee an adjusted forecast of these variables. Thus, the DL models are obtained by optimizing through different values of hidden layers, hidden units, activation functions, kernel initializations, and learning rate. With the purpose to ensure the interdependence of each sample, the training process is carried out through cross-validation using 10 folds. In this way, the extraction of the patterns and the generalization of the results are ensured to use these predictions as input of the energy management system.

The focus on the AI forecast allows to make accurate decisions in real time in the storage system, choosing the best option to meet energy demands in buildings. Interpretation of this data to make the decision taking with minimal human intervention can be carried out by an Intelligent Energy Management System (IEMS) [22]. With the AI approach, IEMS demonstrate a high degree of success of saving controlling and monitoring energy. The storage trade-off can be optimized in order to reduce the energy bills by maximizing the self-consumption [23]. The IEMS predictive control can be performed by model-based or model-free decision algorithms. On one hand, model-based decision algorithms compute the action taking by using mathematical approximations modeling the physical world, p.eg., dynamic programming. On the other hand, in model-free decision algorithms, the action taking is carried out with the experience received by the interaction with the world, p.eg. Reinforcement Learning (RL) [24].

The predictive control of the IEMS is maximized model-free decision algorithms since the model is in continuous exploration. RL is one of the most active areas of research in AI in coming years [25]. These are presented in recent research fields as Internet of Things [26], Energy Management [27] or Neuroscience [28]. In contrast with typical ML techniques, where models learn from some input data, in RL, the learning occurs via interacting with an environment. It is a goal-oriented learning where the model depends on the consequence of its actions. There are diversity RL models, i.e., Q-Learning, SARSA or Dyna-Q. The selection of the most appropriate model to solve the problem is function of how it is intended to learn, how the data is treated or the type of problem to be solved [29]. Deep RL is a branch of RL, where the actions, instead of being randomly selected, are chosen by DL methods based on the experience gained in selecting the states [30].

The aim of this paper is to use DL techniques to predict energy demands and renewable energy production in buildings because it has been demonstrated their accuracy in modeling complex no-linear relationships [31]. In order to improve the building energy use, using the DL predictions, an IEMS has also been considered, whereby energy trade-off is made according to the needs of the building with the Deep RL method. Usually, storage management methods have been applied separately to prediction of electricity demand [32,33] or renewable energy production [34,35]. Nevertheless, since the energy supply management is one of the pillars to obtain the sustainable development goals established by the United Nations [36], it is necessary to study synergies and establish prediction techniques coordinated with a management system to obtain an accurate and efficient building model.

The current global energy crisis makes the energy management of buildings of great interest in current studies. The characteristics of the free-decision models proposed by the advances in AI technology are in line with trends towards automatization. Thus, current research addresses this trouble by applying this type of technique [33,34]. In this manuscript, the free-decision model employed is the Deep RL. As with DL forecasting, the appropriate configuration is selected using the optimal values of a set of hyperparameters, such as exploration/exploitation rate, discount factor, number of episodes, and rewards used. In this way, the energy management of a building that includes PV production is sought to minimize the energy costs and maximize self-consumption, improving its energy efficiency.

The novelty of this paper lies in the application of AI techniques to predict electricity demand and distributed generation to trade off the battery. The proposed methodology can make predictions with high reliability, with DL, to establish the best action to be taken by the storage system, with Deep RL, to improve building performance. The AI architectures selection for building energy management are optimal through a set of values used for hyperparameters tunning process, i.e., number of hidden layers number of hidden units, activation functions, kernel initialization, and learning rate, for forecasting PV power and electricity consumption using DL techniques, and exploration/exploitation rate, discount rate, number of episodes, and rewards, in the IEMS by means of modeling with Deep RL. This allows the correct extraction of patterns from the IEMS proposed to improve the energy management of a real building, maximizing the use of energy produced in situ by PV panels, and minimizing energy costs considering a tariff with hourly discrimination.

The content of this paper is organized as follows: the techniques developed for DL prediction and Deep RL control are explained in Section 2; the studied building, and the DL and the Deep RL models specific to the real-life situation are presented in Section 3; the evaluation of the IEMS trade-off is discussed in Section 4; and the conclusions are outlined in Section 5.

2. Models and methods

This section presents the mathematical approach used for IEMS trade-off in buildings. In the first part, the methodology for DL predictions is shown. Subsequently, the validations used for error assessment of the predicted variables are presented. Finally, the methodology for energy management trade-off through Deep RL is introduced.

2.1. Deep Learning predictions

One of the most common DL models is the Standard Neural Network (SNN) due to its good fittings with tabular data [37]. The layered connection is based on the weight structure (kernel, w, and bias, b), whose values are those that represent the extraction capacity of the SNN [38]. The information given by the input variables is passed through the layers by means of activation functions in each of the nodes. This process

is known as forward propagation and is presented in *eq*. 1. Rectified Linear Unit (ReLU), linear, hyperbolic tangent (tanh) and sigmoid are the most used activation functions in DL problems [39].

$$z = \sum_{i=1}^{n} w_i \bullet g(z_i) + b_i \tag{1}$$

where z is the node equation, *i* corresponds to the previous layer neuron, *n* corresponds to the number of nodes of the layer *i*, and $g(z_i)$ is the activation function of node equation of the previous layer.

The SNN training process is divided in three stages: forward propagation, backward propagation and weights update. In forward propagation, presented above in Eq. 1, weights initialization is needed. The initial values should be randomly selected to avoid entrapment in the information extraction. They must also be as close as possible to zero, in order to obtain the optimal solution with a reduced number of iterations. On one hand, the most kernel initializations used are random normal, glorot normal and He normal [40]. These initializations depend on the activation function used, so He normal optimizes the use of ReLU activations, glorot normal, tanh activations, and random normal, sigmoid and linear activations [41,42]. On the other hand, bias initialization is not a critical process during the algorithm, so zero initialization is generally chosen.

In backpropagation, the information provided with the SNN connections to the output layer is evaluated against the target variable. This evaluation is carried out with some metric as those presented in Section 2.2. Once the distance between real and predicted variables is recognized, the uncertainty is transferred to the input layer via the chain rule [37]. Eq. 2 presents the backpropagation in the layer equation. The backpropagation equations for weights, either kernel and bias, is given in Eq. 3 and Eq. 4 respectively.

$$\partial z = \sum_{j=1}^{n} w_j \bullet \partial z_j \bullet g'(z) \tag{2}$$

where ∂z is the derivative of the node equation, *j* corresponds to the next layer neuron, *n* corresponds to the number of nodes of the layer *j*, and $g'(z)$ is the derivative of the activation function of node equation.

$$\partial w = \sum_{i=1}^{n} [\partial z \bullet g(z_i)] \Big/ N \tag{3}$$

where ∂w is the derivative of the kernel weight, and *N* is the sample length.

$$\partial b = \sum \partial z \Big/ N \tag{4}$$

where, ∂b is the derivative of the bias weight.

The training process is carried out through mini-batch gradient descent using the Adam optimizer [43]. This avoids the typical problems that arise with weight updates in the classical gradient descent optimization, such as standstill at a local minimum and low level of convergence. In this way, the weights updates performed by means of average update, known as batch learning, in each epoch, that is, times the training process go across the training sample. Eq. 5 and Eq. 6 show the weights update, kernel and bias respectively, considering mini-batch gradient descent with Adam optimizer training process.

$$w \leftarrow w - \alpha \bullet \frac{\beta_1 \bullet v_{\partial w} + (1-\beta_1) \bullet \partial w}{1-\beta_1^t} \Bigg/ \sqrt{\frac{\beta_2 \bullet s_{\partial w} + (1-\beta_2) \bullet \partial w^2}{1-\beta_2^t}} + \varepsilon \tag{5}$$

where α is the learning rate, β_1 is the first moment of the exponential decay rate, typically initialized to 0.9, β_2 is the second moment of the exponential decay rate, typically initialized to 0.999, $v_{\partial w}$ and $s_{\partial w}$ are Adam optimizer kernel weights, and ε is a small number to prevent zero division.

$$b \leftarrow b - \alpha \bullet \frac{\beta_1 \bullet v_{\partial b} + (1-\beta_1) \bullet \partial b}{1-\beta_1^t} \Bigg/ \sqrt{\frac{\beta_2 \bullet s_{\partial b} + (1-\beta_2) \bullet \partial b^2}{1-\beta_2^t}} + \varepsilon \tag{6}$$

where $v_{\partial b}$ and $s_{\partial b}$ are Adam optimizer bias weights.

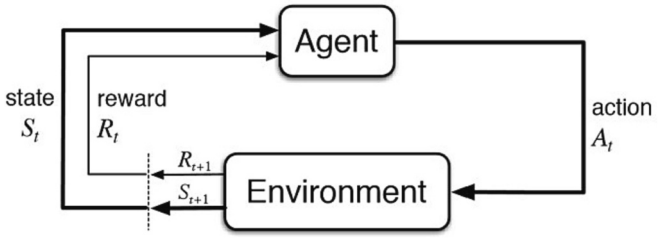

Fig. 1. Basic operation of a Reinforcement Learning problem.

2.2. Error assessment validation

The error metrics used to measure the accuracy of the DL model are normalized Mean Bias Error (nMBE), and normalized Root Mean Squared Error (nRMSE). Both, nMBE and nRMSE, are normalized metric, which make the errors comparable. One the hand, the nMBE provides a measurement of the general bias of a given variable, Eq. 7. Positive and negative values mean the underprediction or overprediction respectively.

$$nMBE = \sum_{k=1}^{N} \frac{\widehat{Y}_k - Y_k}{N} \Big/ Y_{max} \tag{7}$$

where, $\widehat{Y}_k$ is the DL predicted value at time *k*, Y_k is the objective variable at time *k*, and Y_{max} is the maximum value of the objective variable.

On the other hand, the nRMSE indicates the ability of the model to predict the overall load shape that is reflected in the dataset, Eq. 8. Its use is very common, and it is considered an excellent general purpose error metric for numerical predictions.

$$nRMSE = \sqrt{\sum_{k=1}^{N} \frac{(\widehat{Y}_k - Y_k)^2}{N}} \Big/ Y_{max} \tag{8}$$

2.3. Deep Reinforcement Learning control

RL models have a common trend, to obtain the best model that fits to an optimal policy. Nevertheless, there are several agents that can perform this optimization. The used agent depends on the specific problem to carry out. Essentially, it is needed to consider a breakdown of what is looking to reproduce. To do this, it is necessary to evaluate whether actions are continuous or discrete, whether you want to learn during the episode or at the end of the episode, or whether it is a control problem. The learning method must also be considered, there are mainly two: on-policy methods and off-policy methods [29,44]. On one hand, on-policy methods estimate the value of a policy while using it for control. On the other hand, in off-policy methods, the policy used to generate the behavior may be unrelated to the policy that is evaluated and improved.

The basis of the IEMS trade-off is to obtain the best charging and discharging periods of the storage system to maximize the potential of distributed energy generation, thus minimizing energy costs. In this way, the IEMS is a control problem with discrete actions. Furthermore, to improve the behavior throughout the process and in order to respond quickly to the needs of the building, the considered problem should be actualized in each time step, i.e., during the whole episode. Finally, because it is possible that there are random actions through the day that make consumption or production not always the same, an off-policy method is selected. The agent that best fits the needs presented is Q-learning. Furthermore, to provide intelligence to the system, reducing

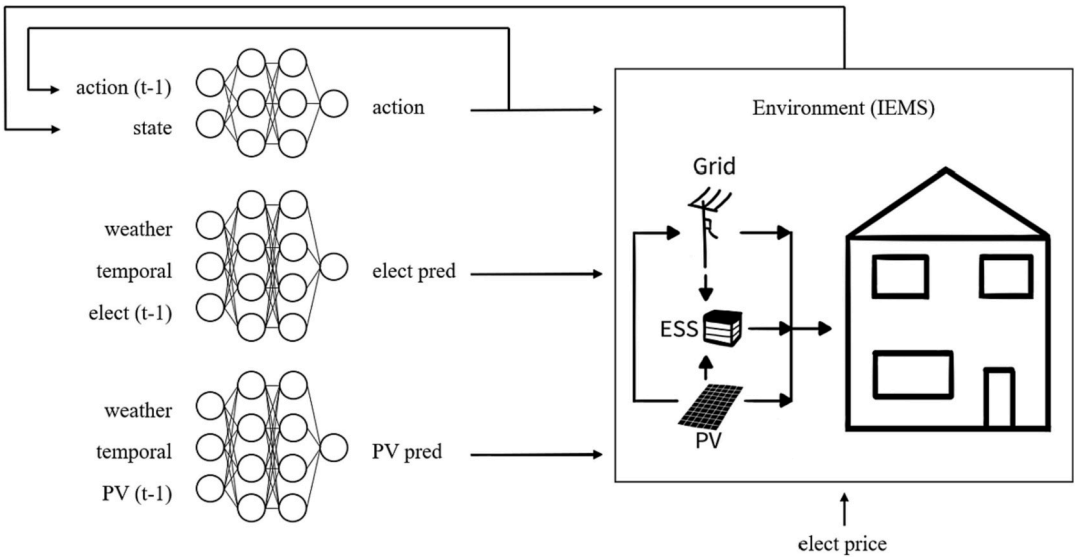

Fig. 2. Topology diagram of the studied IEMS.

the randomness in action selection, a DL method is used to select the action that best adapts to the state of the environment in each time step. The basis of the DL method is presented in the previous section.

RL is a trial-and-error method that seeks the optimal policy from the iteration for each time step in episodes. In every time step, the agent interacts with the environment and selects one action to advance to the next time step. A reward is given to the agent in every action it takes. The better the action, the higher the reward, and vice versa. The goal is to get the maximum total reward when the episode ends. The RL operation is summarized in Fig. 1.

In a RL, the agent performs actions randomly at each step. While the agent interacts with the environment, this method maps the ways according to the reward obtained, known as path reward. In Deep RL, actions are taken using DL, similar to those expressed in Section 2.1. Thus, this technique learns the most appropriate action to the state in the current conditions thanks to the ability of neural networks to generalize states. In this manuscript, Deep RL has been employed since the decision process is more robust to the variations that are generated in the electrical consumption of a building, which allows a better understanding and application by the agent to the actions that the IEMS have to carry out.

To obtain a good, accumulated reward, the agent must exploit, whether the reward is high, or explore, whether the reward is low, the different actions to improve the goal achieved. The used agent, Q-Learning, performs the search for the optimal policy through the value iteration applied to Bellman's optimality equation [45], as presented in Eq. 9.

$$Q(S,A) \leftarrow Q(S,A) + \alpha \bullet [R(S,A,S') + \gamma \bullet max_{A'} Q(S',A') - Q(S,A)] \quad (9)$$

where, $Q(\bullet)$ is the learned action-value function, S is the current state, A is the current action, α is the step-size parameter, $R(S,A,S')$ is the reward, γ is the discount rate, S' is the selected state, and A' is the action taken.

By means of the Q-Learning agent, typical in RL models, the value of $Q(\bullet)$ is estimated from the search in a table, known as Q-table. Thus, the state-action space that is represented in the Q-table must be small enough to be manageable, which only allows for a discrete range of states. The use of neural networks in taking actions in Deep RL methods, through the Deep Q-Learning agent, allows the agent to use the experience for the reproducibility of actions taken. In this way, the robustness of Deep Q-Learning agent allows to prevent the divergence of the outputs and to use a continuous state, in which the state of charge of the IEMS can be simulated effectively [46]. The update formula carried out by the Deep Q-Learning agent is presented in Eq. 10.

$$\theta \leftarrow \theta + \alpha \bullet [R(S,A,S') + \gamma \bullet max_{A'} T(S',A') - Q(S,A;\theta)] \bullet \nabla_\theta Q(S,A;\theta) \quad (10)$$

where θ are the parameters, $T(\bullet)$ represent the objective network, and $\nabla_\theta Q(\bullet)$ is the gradient of the Q-function. Because batch-learning is used in neural networks, the *eq*. 11 must be imposed.

$$total_{steps} \; mod \; C = 0, T \leftarrow Q \quad (11)$$

where $total_{steps}$ is the total number of updates applied to the Q-function in the current time, *mod* represents the concept of modify, and C is the learning step.

Proper action learning in the Deep RL algorithm is highly influenced by the imposition of the rewards. So, following Eq. 9, the update of the current state ($Q(S,A)$) is influenced by a second term in the summatory that considers the reward (R), the discount rate (γ) and a maximum ($max_a Q(S',a)$). The reward is the compensation earned by the agent in the transition from the current state to the one the agent proposes, the discount rate determines the importance of this reward in the current action compared to the immediate rewards, and the maximum is the action that is expected to give a maximum total reward based on the actions previously learnt by the agent because of the imposition of the neural networks in the selection of actions. In this way, if the rewards/punishments are very close, the agent may not understand the difference between a good target and a bad one. On the contrary, if there is a big difference between punishment and reward values, convergence to the desired objective is not ensured since the agent may not understand clearly what the objective is [29,44].

3. Case study

The aim of this study is to obtain a prediction model of PV production and electricity consumption of buildings with high accuracy from DL technique. The building trade-off is managed with a battery system applying Deep RL techniques. In this way, with the combination of these techniques, it is possible to improve the building energy use and, therefore, its energy efficiency. The architecture of the proposed system is shown in Fig. 2.

Table 1
Analysis of the sample quality.

	PV production		Self-consumption		Grid injection		Building consumption		Grid consumption	
	Valid	NaN	Valid	NaN	Valid	NaN	Valid	NaN	Valid	NaN
Feb	100.00 %	0.00 %	98.30 %	1.70 %	98.30 %	1.70 %	95.33 %	4.67 %	95.33 %	4.67 %
Mar	97.54 %	2.46 %	97.54 %	2.46 %	97.54 %	2.46 %	99.33 %	0.67 %	99.33 %	0.67 %
Apr	97.25 %	2.75 %	97.25 %	2.75 %	97.25 %	2.75 %	97.22 %	2.78 %	97.22 %	2.78 %
May	97.98 %	2.02 %	97.92 %	2.08 %	97.92 %	2.08 %	98.81 %	1.19 %	98.81 %	1.19 %
Jun	97.50 %	2.50 %	97.48 %	2.52 %	97.48 %	2.52 %	99.98 %	0.02 %	99.98 %	0.02 %
Jul	97.69 %	2.31 %	97.65 %	2.35 %	97.65 %	2.35 %	99.89 %	0.11 %	99.89 %	0.11 %
Aug	96.89 %	3.11 %	96.73 %	3.27 %	96.73 %	3.27 %	98.50 %	1.50 %	98.50 %	1.50 %
Sep	99.42 %	0.58 %	99.33 %	0.67 %	99.33 %	0.67 %	99.91 %	0.09 %	99.91 %	0.09 %
Total	97.86 %	2.14 %	97.72 %	2.28 %	97.72 %	2.28 %	98.91 %	1.09 %	98.91 %	1.09 %

3.1. Building characteristics

The validation of the models presented is carried out through the study of data obtained in the School of Mining and Energy Engineering located on the Campus of the University of Vigo. The building is divided into 4 floors: the first is the reception and the dining room, the second is the school service and the library, and the third and fourth are the offices, laboratories and classrooms. The roof has a PV installation with 296 PV modules of 410 Wp divided into 18 branches connected in series. The PV connection to the grid is made through 2 inverters of 50 kW, from which a three-phase signal of 100 kW of nominal power is obtained. The combined losses of the PV system are 29.56 %. The actual supply modality is that of individual self-consumption with surpluses receiving financial compensation.

The monitoring of electricity consumption and PV production is carried out with the direct data exchange functionality offered by the inverter system. The data collection period corresponds to 2021 and goes from February 18 at 12 a.m. to September 30 at 11:50 p.m. with a frequency of ten minutes. With the purpose of determine the quality of the data, a filtering process is performed using a data acceptance criterion. The energy balances presented in Eq. 12 and Eq. 13 are also considered in order to ensure the validity of the data, removing the rows with a decompensation >0.01 kW. Through these data analyses, the sample quality is obtained, this is shown in Table 1.

$$P_{cons} = P_{Self-cons} + P_{grid,cons} \tag{12}$$

where P_{cons} is the power consumed in the building, $P_{Self-cons}$, is the self-consumed power in the building, and $P_{grid,\ cons}$, is the power consumed from the grid.

$$P_{PV} = P_{Self-cons} + P_{grid,inj} \tag{13}$$

where P_{PV} is the PV power produced, and $P_{grid,\ inj}$, is the power injected in the grid.

As can be seen in Table 1, the quality of the sample is excellent, with total errors of 2.28 % in the worst case, in self-consumption and grid injection, and 1.09 % in the best case, in building consumption and grid consumption. The worst month monitored is April, with errors >2.75 % in all of variables, and the best is September, with errors <0.70 % in all variables.

The PV production system can supply 63.26 % of the building consumption, however, the current autarky coefficient is 41.39 %. Thus, 65.44 % of PV production is self-consumed, injecting the remaining 34.43 % into the grid. 59.56 % of the energy that the building needs comes from the grid. The storage system proposed in this paper is expected to improve the autarky, reducing the building energy costs. Furthermore, because of the prediction of electricity consumption and PV production through DL technique, it is possible to forecast future electricity consumptions and PV productions. In this way, it is possible to make a trade-off with the best charging and discharging periods of the storage system that could be selected according to the most convenient moments, p.eg., where energy costs are lower or where there is renewable production, reducing building energy costs.

Table 2
Electricity costs in the self-consumption with surpluses receiving financial compensation scenario.

	Power term (€/kW)		Energy term (€/kWh)		
	Peak hours	Valley hours	No solar incidence	Solar incidence	Grid injection compensation
Cost	3.662947	0.956834	0.203603	0.343179	0.060802

The self-consumption with surpluses receiving financial compensation is the selected electricity costs scenario. There are 2 energy branches, defined by the existence or not of solar incidence. This solar incidence goes from 7 a.m. to 4 p.m. GMT in summer schedule and from 9 a.m. to 4 p.m. GMT in winter schedule. Furthermore, compensation for injection on the grid is also considered. Peak hours and valley hours differ in the power term. Valley hours are 10 p.m. to 6 a.m. GMT in weekdays and 24 h a day on holidays and weekends. The rest of the time corresponds to peak hours. The considered power terms are 100 kW in peak hours and 75 kW in valley hours. The associated costs with this scenario are presented in Table 2.

3.2. Deep Learning predictions

The electricity consumption of the building and the PV production prediction is carried out with DL technique through temporal variables and weather data. On one hand, the temporal variables used are the hour of the day, the day of the week and the day of the year. On the other hand, the weather data used are relative humidity, temperature and solar irradiation. The weather variables used are selected from the meteorological station located on the Campus of Vigo of the University of Vigo, a few meters from the analyzed building. The quality of this data is excellent, without errors in the period studied.

To accomplish DL predictions, the dataset is divided in two groups: training and testing. On one hand, training is where the model predicts the variables. This prediction is done using a 10-fold cross-validation method with a batch of 64 and 5000 of epochs, where training is divided into a train set and a validation set. With the use of this technique, the reliability of the model generated in the train set is ensured in the validation set. The training process stops when the model's performance does not improve. The selected threshold to stop the model is 100 iterations. On the other hand, after the cross-validation training, the model is tested. Both training and testing processes allows periods without data, i.e., NaN values.

The DL models have been made in order to obtain the best configuration using a seed and considering the following hyperparameters:

- Training sample: 96 % dataset, of which 96 % corresponds to training set and 4 %, to validation set.
- Testing sample: 4 % dataset.
- Hidden layers: {1; 2; 3}.

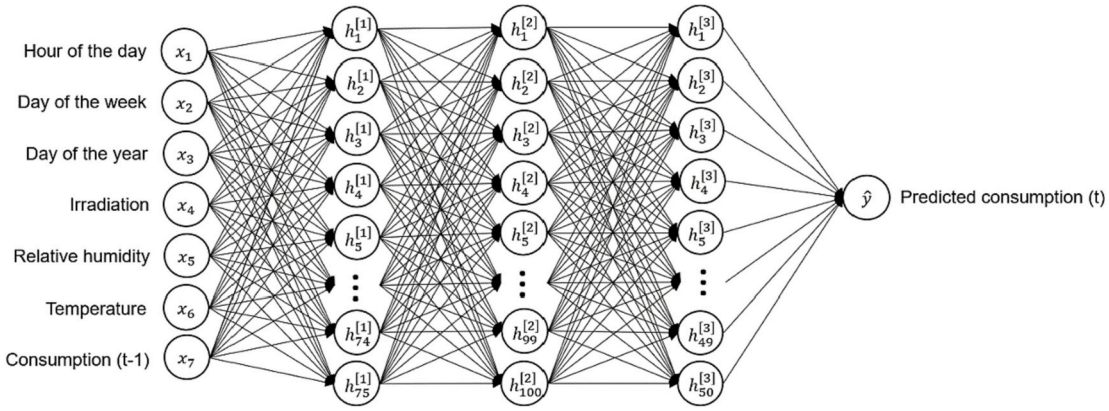

Fig. 3. Optimal DL structure to predict electricity consumption.

Table 3
Averaged k-fold errors in the optimal structure to predict electricity consumption.

	Train set	Validation set	Test set	Whole dataset
nMBE	0.40 %	−0.50 %	0.08 %	0.37 %
nRMSE	5.90 %	4.70 %	9.03 %	5.92 %

Table 4
Averaged k-fold errors in the optimal structure to predict photovoltaic production.

	Train set	Validation set	Test set	Whole dataset
nMBE	−0.21 %	0.08 %	0.04 %	−0.19 %
nRMSE	7.15 %	2.38 %	4.24 %	6.93 %

- Hidden units: {0; 25; 50; 75; 100}.
- Activations: {linear; ReLU; tanh; sigmoid}.
- Kernel initialization: {glorot normal; He normal; random normal} in function of the activation used.
- Bias initialization: zeros.
- Optimizer: Adam, whose learning rate could be {0.001; 0.005; 0.010; 0.020; 0.040; 0.080; 0.160}.

The prediction of electricity consumption is made through temporal variables, i.e., hour of the day, day of the week, and day of the year, weather data, i.e., irradiation, relative humidity and temperature, and consumption in the previous time step. By iterating the hyperparameters presented above and fitting the parameters by the DL model, the optimal structure is presented in Fig. 3.

As can be seen in Fig. 3, the optimal prediction of electricity consumption has 3 hidden layers with 75, 100, and 50 neurons. The activations are tanh, ReLU and ReLU for the first, second and third hidden layer respectively, and tanh for the output layer. So, the kernel initializations are glorot normal, He normal, He normal, and glorot normal respectively. The optimal learning rate is 0.001. Table 3 shows the averaged k-fold errors obtained in this structure with the metrics presented.

As presented in Section 2.2, the prediction deviation has been calculated dividing by the maximum value. In the evaluated building, the maximum of the electricity consumption is 146.37 kW. As can be seen in Table 3, the model is accurate, with mean nRMSE errors in the 10 k-fold <5 % in the validation process, under 6 % in the training and in the whole dataset, and over 9 % in the testing set. This indicates that the presented DL structure can model the building consumption with a deviation of 8.67 kW in the complete dataset, with mean errors of 8.63 kW in the training set, 6.88 kW in the validation set, and 13.22 kW in the test set. The nMBE errors are also reduced, with a deviation around 0.50 % from zero. The nMBE indicates that there are an underprediction in the complete sample, also in the training and testing samples. Nevertheless, the cross validated sample is overpredicted.

The prediction of PV production is made through temporal variables, i.e., hour of the day, day of the week, and day of the year, weather data, i.e., irradiation, relative humidity and temperature, and production in the previous time step. By iterating the hyperparameters presented above and fitting the parameters by the DL model, the optimal structure is presented in Fig. 4.

As can be seen in Fig. 4, the optimal prediction of PV production has 3 hidden layers with 100, 100, and 25 neurons. The activations are ReLU, ReLU and linear for the first, second and third hidden layer

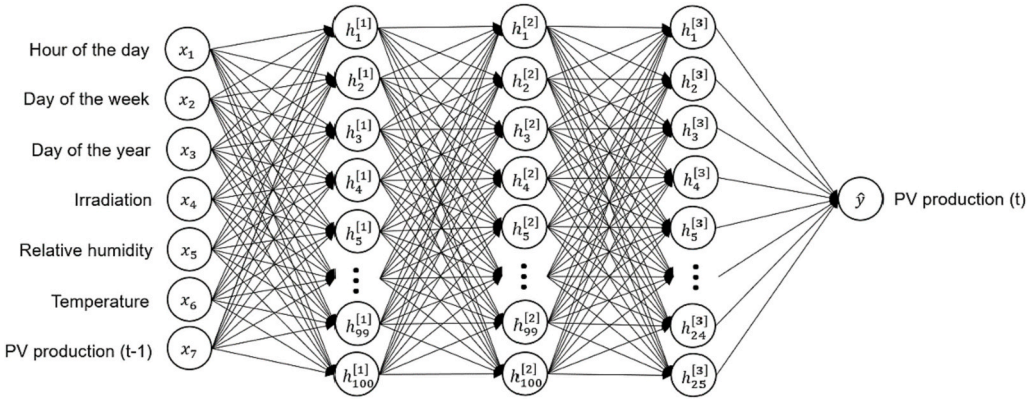

Fig. 4. Optimal DL structure to predict PV production.

Table 5
Rewards and punishments considerations in Deep Reinforcement Learning modeling regardless of solar incidence.

	Peak hours		Valley hours	
	$P_{PV} \leq P_{cons}$	$P_{PV} > P_{cons}$	$P_{PV} \leq P_{cons}$	$P_{PV} > P_{cons}$
discharge	+5	−5	−5	−5
charge	−5	+5	+5	+5

respectively, and tanh for the output layer. So, the kernel initializations are He normal, He normal, random normal, and glorot normal respectively. The optimal learning rate is 0.001. Table 4 shows the averaged k-fold errors obtained in this structure with the metrics presented.

As commented above, the prediction deviation has been normalized using equations presented in Section 2.2 dividing by the maximum. The maximum of the PV production in the studied building is 100 kW. As can be seen in Table 4, the model is accurate, with mean nRMSE errors in the k-fold around 7 % in the training and in the whole dataset, and over 2 % and 4 % in validation and test processes respectively. This indicates that this DL structure can model the PV production with a deviation of 6.93 kW in the complete dataset, with mean errors of 7.15 kW in the training set, 2.38 kW in the validation set, and 4.24 kW in the test set. The nMBE errors also are reduced with a deviation around 0.20 % respect to zero. The nMBE indicates an overprediction in the whole dataset and in the training sample. On the other hand, the test and validation samples are underpredicted.

3.3. Deep Reinforcement Learning modeling

The storage system trade-off is carried out with Deep RL, where, through the predicted variables, i.e., electricity consumption and PV production, the temporal variables, as hour of the day and day of the week, and the electricity costs, the storage system is managed. The boundary conditions utilized are given by physical limits, i.e., the charge level must be between 0 % and 100 %, and theoretical limits, since the battery should have a charge level over than 10 % to increase its life cycle. The initial battery state of charge (SoC) is 100 %. The data used to simulate the storage system is a charge and discharge power of 31 kW, a capacity of 42 kWh and an efficiency of 96 %.

The action selection in every time step is done through a DL modeling in order to maximize the consumption from PV production. The activation function in the output layer must be sigmoid, because we are looking for the most probable bivariate action. In this model, the cross-validation technique is not considered, so data is divided in training and testing samples, where the training sample corresponds to a mean week, and the testing sample, to the whole dataset. As a period of one week is considered for training, the batch size is 32. In contrast, in the testing the batch size is 64 for considering the whole dataset. The rest of the DL modeling considerations are those presented above. The Deep RL hyperparameters presented below have been selected in order to obtain the highest possible objective by using a seed.

- Agent: Deep Q-Learning.
- Exploration/exploitation trade-off: Epsilon-Greedy model, whose values could be {0.05; 0.10; 0.15; 0.20; 0.25}.
- Discount rate: {0.80; 0.85; 0.90; 0.95; 0.99}
- Episodes: {50; 100; 250; 500}
- Episode length: an average week with the mean values at each time step of the predicted variables.
- Warm-up: 2 episodes.
- States: in every time step, the combination of the period of the day (peak hours, and valley hours with solar incidence or not), the SoC, the predicted PV power production, and the predicted power consumption of the building.

Actions: discharge ($P_{ch} = 0$ & $P_{grid,\ inj} \geq 0$); *charge* ($0 < P_{ch} \leq P_{ch,\ max}$ & $P_{grid,\ inj} \geq 0$).

- Rewards/punishments: +5 or − 5 whether action is *discharge* or *charge* in the states. These values have been selected in a testing process considering ±1, ± 5, ± 10, ± 20, with the presented values (±5) being those with the greatest stability in the management of the storage system. Table 5 summarizes the considered rewards.

The rewards have been considered taking into account the following rules:

- In the case of peak hours, the discharge is prioritized unless there is a PV surplus. Thus, the reward is positive if the IEMS prioritizes the surplus of the PV resource for charging the battery instead of injection into the grid, i.e., that the action taken is *charge*. The reward is negative in the opposite case, that the IEMS prioritizes the injection to the grid instead of charging the battery, i.e., that the action taken

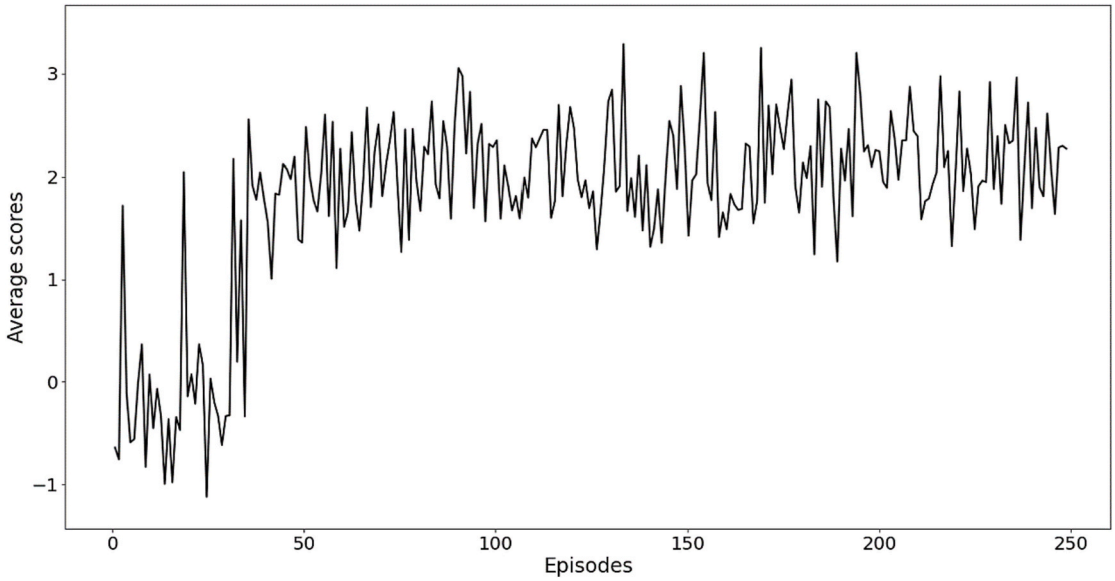

Fig. 5. Learning process of the Deep Q-learning agent through the episodes.

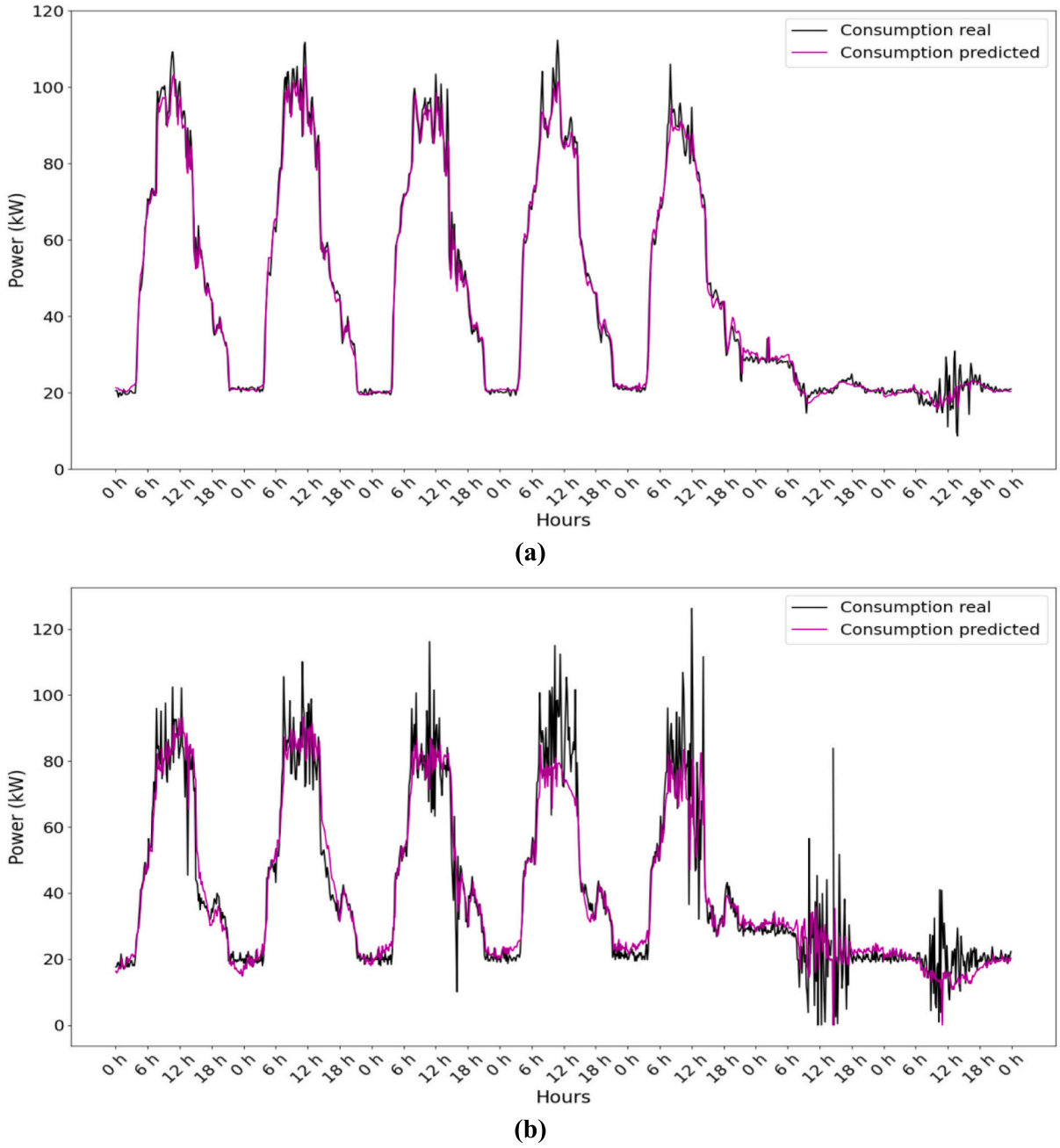

Fig. 6. Comparison between real and predicted with Deep Learning electricity power consumption, where: **(a)** Training sample; **(b)** Testing sample.

is *discharge*. If there is no PV surplus, the reward is positive if the action is *discharge* and negative if it is *charge*.

- In the case of valley hours, charging is always prioritized, regardless of the photovoltaic resource, since this period occurs mostly at night, i.e., when PV production is zero. Thus, the reward is positive if the action is *charge* and negative if it is *discharge*.

Through train and error trade-off, the agent learns the best situations to charge or discharge the storage system. The optimal action selection is obtained considering 250 episodes with an Epsilon-Greedy of 0.05, and a discount rate of 0.99. This optimal model has 3 hidden layers with 50, 25 and 25 neurons. The activations are linear, linear and ReLU for the first, second and third hidden layer, and sigmoid for the output layer. So, the kernel initializations are random normal, random normal, He normal and random normal respectively. The optimal learning rate is 0.001. In contrast to DL models, Deep RL is a trial and error method, so the performance is extracted from the interaction with the environment. The agent learning process during the episodes is shown in Fig. 5.

The Fig. 5 shows Deep RL learning behavior. Starting with no idea what action to take, the agent can achieve the optimal policy based on experience. The learning process begins with exploring what actions to take. When the reward of some action is good, the agent exploits that path. On the contrary, if the action is not good enough, the agent continues to exploit. The fact that from several episodes the mean reward is not constant is because the agent continues to explore in order to obtain a better policy with frequency given by the Epsilon-Greedy model.

4. Results

DL predictions allows an accurate and smart energy management.

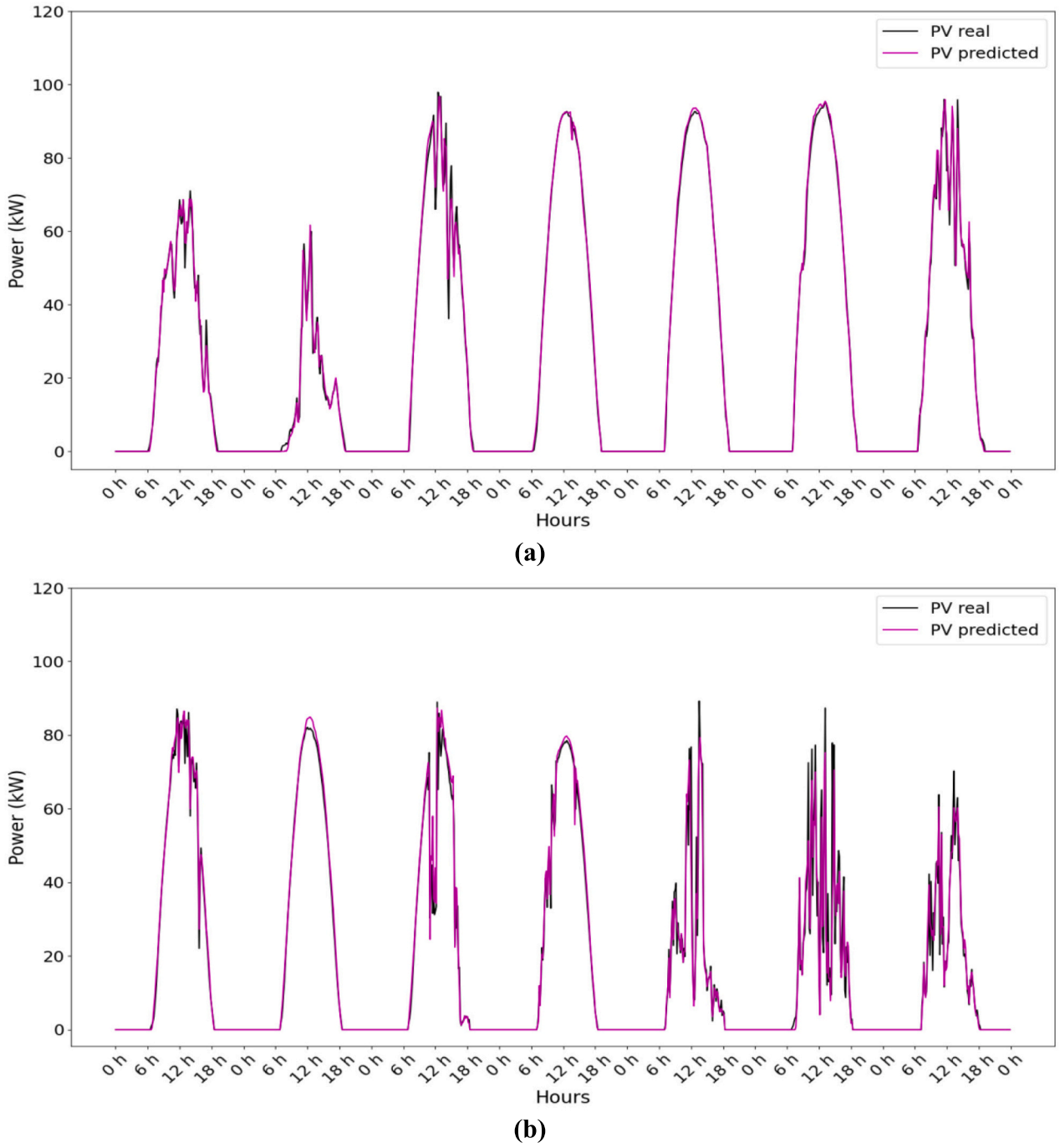

Fig. 7. Comparison between real and predicted with Deep Learning photovoltaic power generation, where: **(a)** Training sample; **(b)** Testing sample.

With these models, the IEMS can decide quickly the best moments to charge the storage system in current or even later states from previous information. Thus, there is a high level of comfort in the building while improving energy efficiency and minimizing energy costs. The comparison of the building consumption is presented in Fig. 6 and the one of the PV production, in Fig. 7.

As can be seen in Fig. 6, the DL k-fold model presents accurate predictions of electricity power consumption. The model fits very well the variations in both samples, training and testing. This can represent the variations that occur on weekdays and weekends, and even holidays. Also differentiate between day and night behavior. Analyzing this figure, together with metrics presented in Table 3, it can be exposed that the training sample predictions are more accurate than those of the testing sample. The model has a light underprediction in both cases. Therefore, it can be determined that this DL structure is capable of modeling and predicting accurately the electricity power consumption of the studied building.

Fig. 7 shows that the presented PV power production model is a good predictor since fits successfully the variation either in training and testing samples. The predicted solar PV production is high in central hours of the day and zero at night, as occurs in the real data. As previously deduced in Table 4, the training set has better results than testing set. Nevertheless, both cases have a slight variation respect to reality.

As has been discussed above, those DL models accurately fit the electricity demand and PV production. This provide intelligence to the IEMS since subsequent actions can be predicted with certainty. The

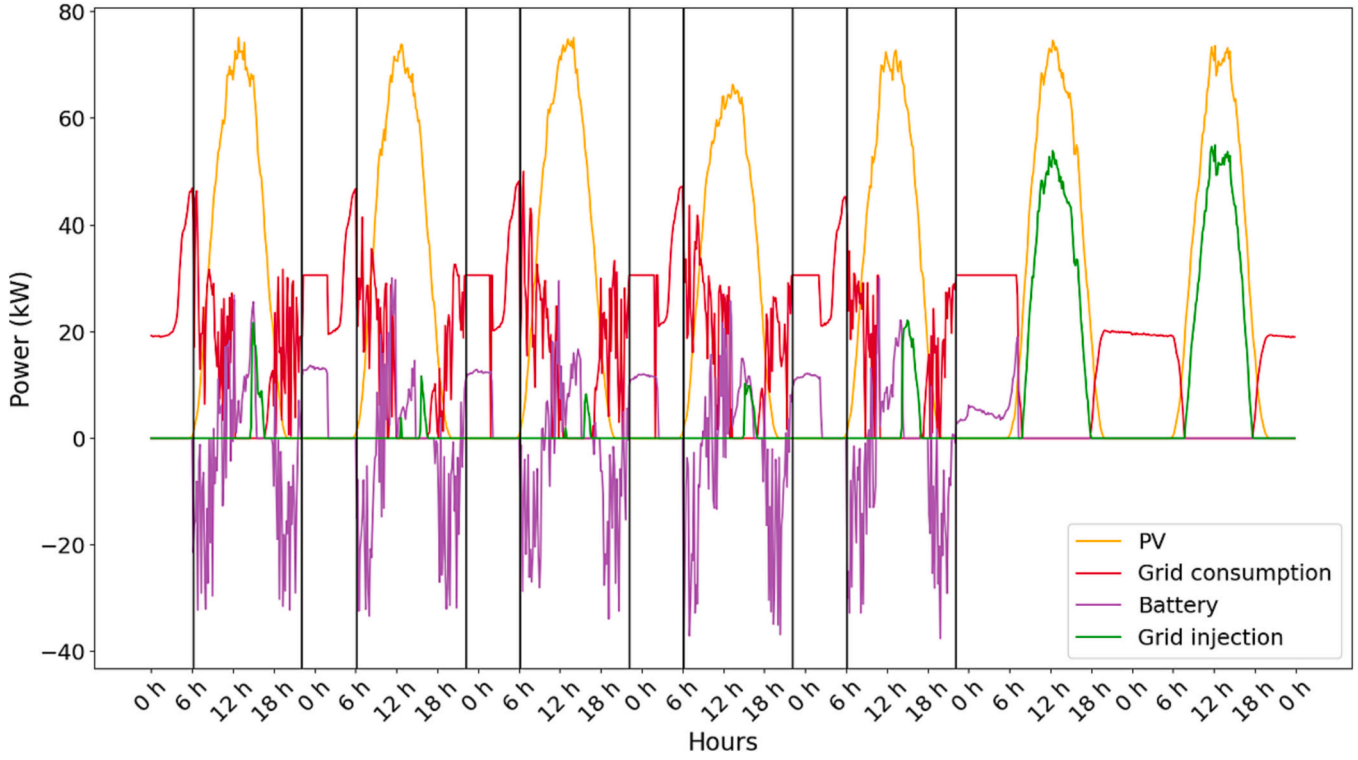

Fig. 8. Load balance obtained with the Deep RL method in an average week, where vertical lines are the changes in electricity prices. The negative values in the battery indicate the energy discharged and the positive, the charged one.

trade-off can be improved learning the optimal policy with Deep RL. In this way, it is possible to adapt quickly to energy fluctuations or future actions while learning from current states. Improving the building energy management. Fig. 8 shows the energy management using this combined methodology.

In the study of energy modeling with Deep RL, the periods marked by the vertical lines in Fig. 8 have been set, the peak hours (black lines), established from 6 a.m. to 10 p.m. As can be seen in Figs. 8, 2 periods can be distinguished, weekdays and weekends. On the one hand, on weekdays, the IEMS decreases the energy consumption of the grid (red line) when peak hours begin. At this moment, the Deep RL model sends a trigger to the battery (purple line) to support grid power and ensure the building's power needs. At the time the hours with solar incidence appear, i.e., 9 a.m. according to the selected tariff, the grid continues to deliver a little energy to support the PV production (orange line) and charge the battery meanwhile the building needs are fulfilled. Just as the model predicts that the state of charge of the battery is enough, the grid consumption reaches 0. In this way, in the subsequent hours, when the PV production is declining, the battery has sufficient charge to support the grid once again in the building necessities. When valley hours come, the battery is fully charged using the grid. On the other hand, during the weekends, due to valley hours are established and the building only has residual demands, the battery does not come into operation and remains totally charged, i.e., the building demand is completely from grid and PV. Therefore, it can be confirmed that the IEMS is able to recognize hours with lower electricity costs, i.e., nights and weekends. Fig. 9 shows the origin of the energy consumed by the building to analyze the differences exhibited by the proposed IEMS with respect to the model without AI.

Observing Fig. 9, the grid consumption (red line) has a similar behavior in the current case (a) and the case with AI (b) at the start of the day and during weekends. When the peak hours commence, i.e., 6 a.m., consumption in the current case continues the trend marked by the previous hours, nevertheless, the AI model, has an abrupt reduction. At the moment that the PV (orange line) has a certain production, i.e., around 10 kW, the consumption begins to be reduced in the system without AI, on the contrary, the IEMS has a variable consumption. The minimum grid consumption is 0 in the AI model and this value is maintained for a few hours. This reduction is not obtained in the current system, and the minimum reached is occasional. At the end of the hours of sun marked by the tariff, i.e., 4 p.m., an increasing trend is observed in the model without AI and a variable consumption in the IEMS. When the peak hours end, i.e., 10 p.m., the AI system has a higher power need, and this is maintained until the battery is fully charged. The use of solar energy is greater in the IEMS, where, contrary to what happens in the current case, practically all the energy supplied by the PV source is self-consumed (green line).

The IEMS presented allows a better use of energy and a lower dependence on the grid. Thus, there is an improvement in the autarky coefficient of the building, going from 41.39 % to 52.95 % due to an increase in the energy produced by the PV panels. The self-consumption increases 14.83 points, from 65.44 % to 80.27 %. This improvement in energy efficiency in the building also has a great impact on electricity costs. Thus, the grid consumption is reduced by 10.63 points, from 59.56 % to 48.93 % of the building demand. Table 6 summarize the presented results.

As can be seen in Table 6, the proposed IEMS improves sustainability in the building studied. Observing the grid consumption column, this improvement causes the building to increase the degree of energy efficiency from grade B, established in the range of 55–75 % of consumption, to grade A, established in the range 42–55 %. Since the energy mix in Spain has an emission of 259 g $CO_{2\ eq}$/kWh, the increase in the self-consumption reduces the equivalent carbon footprint by 8.56 kg $CO_{2\ eq}$. These contributions represent a saving of 16.67 % in the annual electricity cost of the building, reducing the electricity bills by around 25,000 €/year.

5. Conclusions

Buildings constantly generate new information through sensors or

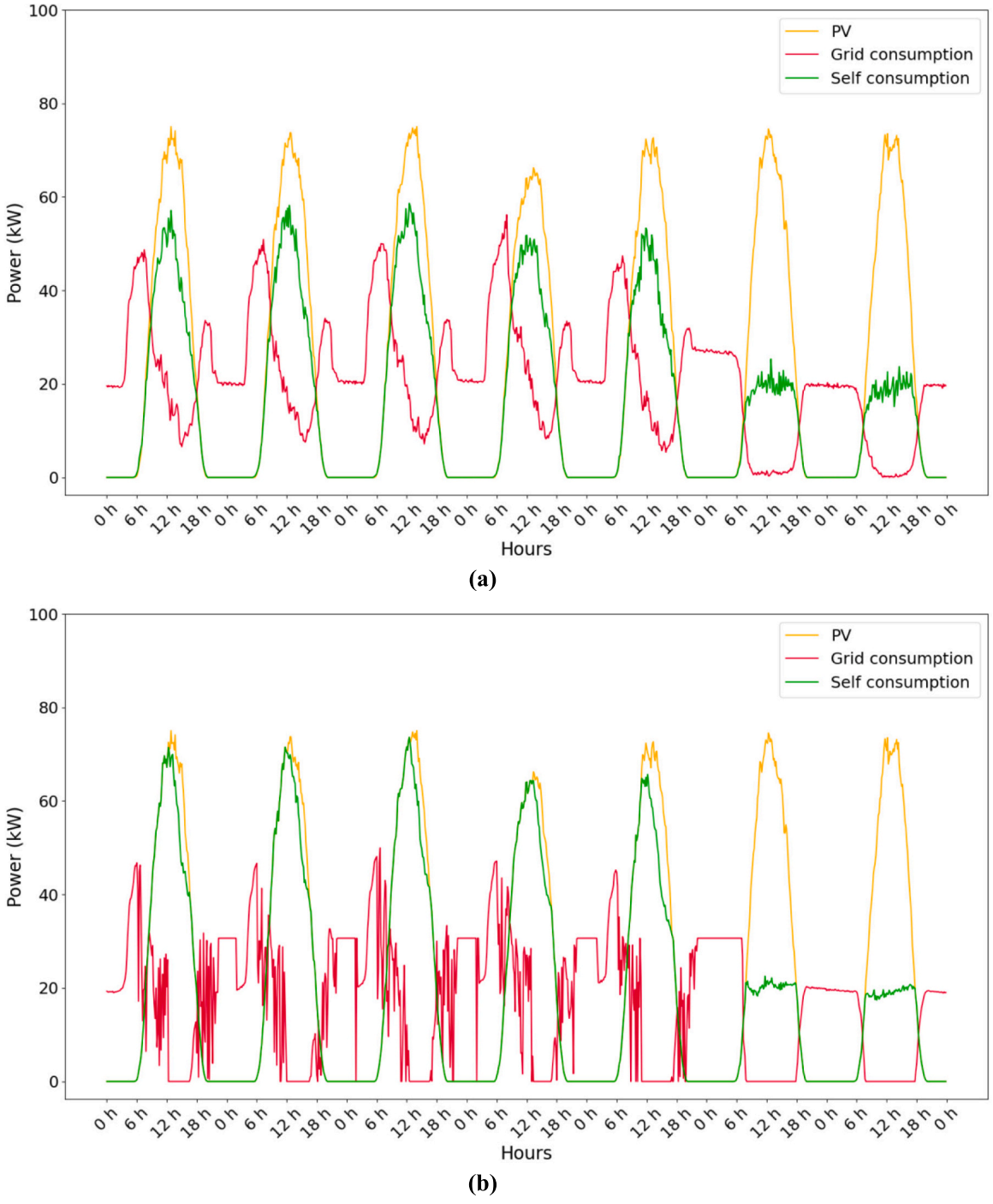

Fig. 9. Distinction of the source of consumption in the building, where: **(a)** current case; **(b)** IEMS case.

Table 6
Percentage of energy use in the studied systems.

	Autarky coefficient	Self-consumption	Grid consumption
Current system	41.39 %	65.44 %	59.56 %
Proposed IEMS	52.92 %	80.27 %	48.93 %

smart meters. To know how to manage this type of information to increase the building performance is key to achieving the emission reduction targets set by the United Nations. Therefore, the development of new techniques with methods that allow to obtain accurate results efficiently with constant new information, such as AI, is of vital importance. This paper presents a methodology to increase the building performance by means of an IEMS. The energy trade-off is carried out with Deep RL, with the Deep Q-Learning agent. This method, based on predictions of distributed energy and consumption with DL, selects the optimal action in each period. This combination of techniques allows energy management that considers actual and future demands. Furthermore, as the time increases, the information also increases, allowing the presented method to learn and fit more accurately to possible energy fluctuations.

Results shows that the proposed techniques for IEMS can reduce energy costs by improving energy efficiency of buildings. On one hand,

DL allows accurate modeling of electricity consumption and PV production, with average k-fold errors computed with nRMSE around 6 % and 7 % respectively. The prediction deviation obtained with nMBE is also reduced in both cases, being 0.5 % in electricity consumption and 0.2 % in PV production. The presented DL models have a slight underprediction in electricity consumption and a slight overprediction in PV production.

On the other hand, Deep RL can take advantage of this predictions to manage the storage system, selecting the optimal actions to increase the building performance. With the combination of AI techniques, the studied building with IEMS increases its autarky coefficient, improving the self-consumption by 22.65 %, thus reducing the grid dependence by 17.85 %. Therefore, this IEMS supposes an increase in energy efficiency of the building, scaling from grade B to grade A, a reduction in the carbon footprint of 8.56 kg $CO_{2\ eq}$ and a savings on the electricity bills for a value of around 25,000 €/year.

Author statement

All the authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript.

Funding

This work was supported by the Ministry of Science, Innovation and Universities of the Spanish Government (PV Smart project TED2021-130677B-I00), and the Universidade de Vigo, Spain (grant 00VI 131H 6410211). Funding for open access charge: Universidade de Vigo/CISUG, Spain.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported by the Ministry of Science, Innovation and Universities of the Spanish Government (PV Smart project TED2021-130677B-I00), and the Universidade de Vigo, Spain (grant 00VI 131H 6410211). Funding for open access charge: Universidade de Vigo/CISUG, Spain.

References

[1] European Commission, A European Long-term Strategic Vision for a Prosperous, Modern, Competitive and Climate Neutral Economy, EU, Brussels, Belgium, 2018.
[2] International Energy Agency, Appliances and Equipment, EU, Paris, France, 2021.
[3] Energy Efficiency 2018. Analysis and Outlooks to 2040, EU, Paris, France, 2018.
[4] Solar Power Europe, EU Market Outlook for Solar Power 2019-2023, EU, Brussels, Belgium, 2019.
[5] S. Quoilin, K. Kawadias, A. Mercier, I. Pappone, A. Zucked, Quantifying self-consumption linked to solar home battery systems: statistical analysis and economic assessment, Appl. Energy 182 (2016) 58–67.
[6] J.A. Ballesteros-Gallardo, A. Arcos-Vargas, F. Núñez, Optimal design model for a residential PV storage system an application to the Spanish case, Sustainability 13 (2021) 575.
[7] European Renewable Energies Federation, European Policy Advisory Paper, EU, Belgium, Brussels, 2020.
[8] L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang, X. Guan, A review of deep reinforcement learning for smart building energy management, IEEE Internet Things J. 8 (15) (2021) 12046–12063, 9426901.
[9] M. Ali, K. Prakash, M.A. Hossain, H.R. Pota, Intelligent energy management: evolving developments, current challenges, and research directions for sustainable future, J. Clean. Prod. 314 (2021), 127904.
[10] L. Langer, T. Volling, An optimal home energy management system for modulating heat pumps and photovoltaic systems, Appl. Energy 278 (2020), 115661.
[11] European Commission, Coordinated Plan on Artificial Intelligence 2021 Review, EU, Brussels, Belgium, 2021.
[12] L. Chen, X. Hu, T. Xu, H. Kuang, Q. Li, Turn signal detection during nighttime by CNN detector and perceptual hashing tracking, IEEE Trans. Intell. Transp. Syst. 18 (12) (2017) 3303–3314.
[13] C. Donahue, B. Li, R. Prabhavalkar, Exploring speech enhancement with generative adversarial networks for robust speech recognition, ICASSP 5024–5028 (2018), 8462581.
[14] M.D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J.C. Folk, M. Niemeijer, Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning, Investig. Ophthalmol. Vis. Sci. 57 (13) (2016) 5200–5206.
[15] M.Z. Alom, T.M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, M. Hasan, B.C. Van Essen, A.A.S. Awwal, V.K. Asari, A state-of-the-art survey on deep learning theory and architectures, Electronics 8 (3) (2019) 292.
[16] Q. Zhang, L.T. Yang, Z. Chen, P. Li, A survey on deep learning for big data, Inf. Fusion 42 (2018) 146–157.
[17] M. Cordeiro-Costas, D. Villanueva, P. Eguía-Oller, Optimization of the electrical demand of an existing building with storage management through machine learning techniques, Appl. Sci. 11 (17) (2021) 7991.
[18] A. Rahman, V. Srikumar, A.D. Smith, Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks, Appl. Energy 212 (2018) 372–385.
[19] M. Comesaña-Martínez, L. Febrero-Garrido, F. Troncoso-Pastoriza, J. Martínez-Torres, Prediction of building's termal performance using LSTM and MLP neural networks, Appl. Sci. 10 (21) (2020) 1–16, 7439.
[20] J. López-Gómez, A. Ogando-Martínez, F. Troncoso-Pastoriza, L. Febrero-Garrido, E. Granada-Álvarez, J.M. Pérez-Canosa, Photovoltaic power prediction using artificial neural networks and numerical weather data, Sustainability 12 (24) (2020) 1–19, 10295.
[21] A. Dairi, F. Harrou, Y. Sun, S. Khadraoui, Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach, Appl. Sci. 10 (23) (2020) 1–20, 8400.
[22] M.S. Ibrahim, W. Dong, Q. Yang, Machine learning driven smart electric power systems: current trends and new perspectives, Appl. Energy 272 (2020), 115237.
[23] R. Gopinath, M. Kumar, C. Prakash Chandra Joshua, K. Srinivas, Energy management using non-intrusive load monitoring techniques – state-of-the-art and future research directions, Sustain. Cities Soc. 62 (2020), 102411.
[24] L. Langer, T. Volling, An optimal home energy management system for modulating heat pumps and photovoltaic systems, Appl. Energy 278 (2020), 115661.
[25] Z. Wang, T. Hong, Reinforcement learning for building controls: the opportunities and challenges, Appl. Energy 269 (2020), 115036.
[26] M. Min, L. Xiao, Y. Chen, P. Cheng, D. Wu, W. Zhuang, Learning-based computation offloading for IoT devices with energy harvesting, IEEE Trans. Veh. Technol. 68 (2) (2019) 1930–1941, 8598893.
[27] R. Xiong, J. Cao, Q. Yu, Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle, Appl. Energy 211 (2018) 538–548.
[28] Q.J.M. Huys, T.V. Maia, M.J. Frank, Computational psychiatry as a bridge from neuroscience to clinical applications, Nat. Neurosci. 19 (3) (2016) 404–413.
[29] N.C. Luong, D.T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, D.I. Kim, Applications of deep reinforcement learning in communications and networking: a survey, IEEE Commun.Surv.Tutor. 21 (4) (2019) 3133–3174, 8714026.
[30] T.T. Nguyen, N.D. Nguyen, S. Nahavandi, Deep reinforcement learning for multiagent systems: a review of challenges, solutions, and applications, <sb:contribution><sb:title>IEEE Trans. </sb:title></sb:contribution><sb:host><sb:issue><sb:series><sb:title>Cybern.</sb:title></sb:series></sb:issue></sb:host> 50 (9) (2020) 3826–3839, 9043893.
[31] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M.P. Reyes, M.L. Shyu, S.C. Chen, A survey on deep learning: algorithms, techniques, and applications, ACM Comput. Surv. 51 (5) (2018) 92.
[32] J.R. Vázquez-Canteli, Z. Nagy, Reinforcement learning for demand response: a review of algorithms and modeling techniques, Appl. Energy 235 (2019) 1072–1089.
[33] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, T. Jiang, Deep reinforcement learning for smart home energy management, IEEE Internet Things J. 7 (4) (2020) 2751–2762, 8919976.
[34] C. Sun, F. Sun, S.J. Moura, Nonlinear predictive energy management of residential buildings with photovoltaics & batteries, J. Power Sources 325 (2016) 723–731.
[35] I. Kim, Markov Chain Monte Carlo and acceptance-rejection algorithms for synthesising short-term variations in the generation output of the photovoltaic system, IET Renew.Power Gener. 11 (6) (2017) 878–888.
[36] European Commission, Report From the Commission to the European Parliament and the Council: On Progress of Clean Energy Competitiveness, EU, Brussels, Belgium, 2020.
[37] Y. Lecun, Y. Bemgio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[38] J. Schmidhunber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117.
[39] V. Sze, Y.-H. Chen, T.-J. Yang, J.S. Emer, Efficient processing of deep neural networks: a tutorial and survey, Proc. IEEE 105 (12) (2017) 2295–2329, 8114708.

[40] H. Li, M. Krček, G. Perin, A comparison of weight initializers in deep learning-based side-channel analysis, Lect. Notes Comput. Sci 12418 (2020) 126–143.
[41] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on imagenet classification, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1026–1034.
[42] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, J. Mach. Learn. Res. 9 (2010) 249–256.
[43] Y. Feng, Y. Tu, Phases of learning dynamics in artificial neural networks in the absence or presence of mislabeled data, <sb:contribution><sb:title>Mach. Learn.</sb:title></sb:contribution><sb:host><sb:issue><sb:series><sb:title>Sci. Technol.</sb:title></sb:series></sb:issue></sb:host> 2 (4) (2021).
[44] K. Arulkumaran, M.P. Deisenroth, M. Brundage, A.A. Bharath, Deep reinforcement learning: a brief survey, IEEE Signal Process. Mag. 34 (16) (2017) 26–38, 8103164.
[45] E. Mocanu, D.C. Mocanu, P.H. Nguyen, A. Liotta, M.E. Webber, M. Gibescu, J. G. Slootweg, On-line building energy optimization using deep reinfocement learning, IEEE Trans.Smart Grid 10 (4) (2019) 3698–3708, 8356086.
[46] S. Ohnishi, E. Uchibe, Y. Yamaguchi, K. Nakanishi, Y. Yasui, S. Ishii, Constrained deep Q-learning gradually approaching ordinary Q-learning, Front. Neurorobot. 13 (2019) 103.