












Article

# Solar Thermal Collector Output Temperature Prediction by Hybrid Intelligent Model for Smartgrid and Smartbuildings Applications and Optimization

José-Luis Casteleiro-Roca <sup>1,†</sup>, Pablo Chamoso <sup>2,3,†</sup>, Esteban Jove <sup>1,†</sup>,  
Alfonso González-Briones <sup>2,3,4,†</sup>, Héctor Quintián <sup>1,\*,†</sup>, María-Isabel Fernández-Ibáñez <sup>1,†</sup>,  
Rafael Alejandro Vega Vega <sup>1,†</sup>, Andrés-José Piñón Pazos <sup>1,†</sup>,  
José Antonio López Vázquez <sup>1,†</sup>, Santiago Torres-Álvarez <sup>5,†</sup>, Tiago Pinto <sup>6,†</sup>  
and Jose Luis Calvo-Rolle <sup>1,†</sup>

<sup>1</sup> CTC, Department of Industrial Engineering, CITIC, University of A Coruña, 15405 Ferrol, A Coruña, Spain; jose.luis.casteleiro@udc.es (J.-L.C.-R.); esteban.jove@udc.es (E.J.); isabel.fibanez@udc.es (M.-I.F.-I.); rafael.alejandrov.vega.vega@udc.es (R.A.V.V.); andres.pinon@udc.es (A.-J.P.P.); jose.lopez@udc.es (J.A.L.V.); jlcalvo@udc.es (J.L.C.-R.)

<sup>2</sup> BISITE Research Group, University of Salamanca, Edificio I+D+i, 37007 Salamanca, Spain; chamoso@usal.es (P.C.); alfonsogb@usal.es (A.G.-B.)

<sup>3</sup> Air Institute, IoT Digital Innovation Hub, Calle Segunda 4, 37188 Salamanca, Spain

<sup>4</sup> Research Group on Agent-Based, Social and Interdisciplinary Applications (GRASIA), Complutense University of Madrid, 28040 Madrid, Spain

<sup>5</sup> Department of Computer Science and System Engineering, University of La Laguna, Avda. Astrof. Francisco Sánchez s/n, 38200 S/C de Tenerife, Spain; storres@ull.edu.es

<sup>6</sup> RGECAD-Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Polytechnic of Porto (P.PORTO), P-4200-465 Porto, Portugal; tcp@isep.ipp.pt

\* Correspondence: hector.quintian@udc.es; Tel.: +34-881-013-117

† These authors contributed equally to this work.

Received: 27 May 2020; Accepted: 2 July 2020; Published: 5 July 2020



**Abstract:** Currently, there is great interest in reducing the consumption of fossil fuels (and other non-renewable energy sources) in order to preserve the environment; smart buildings are commonly proposed for this purpose as they are capable of producing their own energy and using it optimally. However, at times, solar energy is not able to supply the energy demand fully; it is mandatory to know the quantity of energy needed to optimize the system. This research focuses on the prediction of output temperature from a solar thermal collector. The aim is to measure solar thermal energy and optimize the energy system of a house (or building). The dataset used in this research has been taken from a real installation in a bio-climate house located on the Sotavento Experimental Wind Farm, in north-west Spain. A hybrid intelligent model has been developed by combining clustering and regression methods such as neural networks, polynomial regression, and support vector machines. The main findings show that, by dividing the dataset into small clusters on the basis of similarity in behavior, it is possible to create more accurate models. Moreover, combining different regression methods for each cluster provides better results than when a global model of the whole dataset is used. In temperature prediction, mean absolute error was lower than 4 °C.

**Keywords:** clustering; prediction; regression; solar thermal collector; hybrid model

## 1. Introduction

In recent years, preserving the environment has become a great concern. One of the reasons for this trend is environmental deterioration caused by human action. The governments of most countries

became aware of the problem; many laws, directives, and treaties are made to try to lessen or even eliminate the problem. Changing the trend is a big challenge for various reasons, such as the countries' economic models, the current configuration of the cities, and people's habits.

Many researchers have proposed solutions to the above-mentioned problems. An aspect that is becoming increasingly important is changing cities. In fact, some authors pointed to the importance of cities and their key role in fighting against climate change [1]. In this regard, within the smart city concept there are some key building-related terms, such as smart buildings.

The definition of the term smart city is under constant review [2,3]. Taking this fact into account, a smart city could be described as a city that uses the potential of technology and innovation, along with other resources that help make effective the use of this potential, promotes sustainable development, and ultimately, improves the quality of life of its citizens [4–6]. Smart cities are made up of a group of many elements, tangible or intangible [7]. One of those elements are smart buildings; they follow the same paradigm as smart cities, facilitating the achievement of their final objective [8].

Domotics or home automation, is an aspect of smart buildings. This term refers to the control of the different systems that a building includes: the air conditioning system, the lighting systems, the air purification system, the entertainment system, etc. [9]. In addition to home automation, smart buildings have the following smart city related features: sustainability, efficiency, security, etc. If possible, all these features should be part of a smart building from conception [10–12]. To accomplish all these objectives, it will be essential to employ multiple technologies [13–17].

To optimize aspects such as efficiency and sustainability, in addition to preventing unwanted loss, it is necessary to consume no more than the strictly required energy [16]. Given a house with the capacity to generate energy, it is necessary to know, beforehand, the amount of energy that will be generated and the amount of energy that will be consumed. This prediction would help ensure optimal comfort conditions. If the generation capacity is not able to meet the energy needs, the house's energy system must be supplemented by an external source [17]. For optimal energy generation, consumption, and purchase, it is essential to be able to predict all the variables satisfactorily [18].

In the state of the art, there are studies that address the topic of predicting the generation of solar energy [19,20]. In [21], the authors propose a hybrid model that combines machine-learning methods with a theta statistical method for a more accurate prediction of future solar power generation from renewable energy plants. Reference [22] presented a solar power modeling method using artificial neural networks (ANNs), specifically, a general regression neural network (GRNN) and a feed forward back propagation (FFBP), using as inputs: maximum temperature, minimum temperature, mean temperature, and radiance. The main objective of [23] was to determine time horizon, having the highest representative for generated electricity prediction of small scale solar power system applications; finally it addressed solar power prediction in a 5 min time horizon.

There are multiple prediction techniques which have been proven effective in prediction tasks. The different methods range from the simplest to the most complex ones, which frequently coincide with traditional and advanced methods, respectively. When systems are complex and contain, for example, very strong non-linearities, the methodologies that offer the best results are often based on intelligent techniques. Different studies in the medical field [24–29] show the use of intelligent techniques to predict patient response variables. In research on energy systems, [30–34], hybrid systems obtain better results than traditional techniques. In the field of anomaly detection, several studies have used this type of hybrid intelligent system [35–39].

This research presents a case study that involves a bioclimatic house. Specifically, the study focuses on the ability to accurately predict the amount of energy generated by a solar thermal installation. This ability would make it possible to purchase only the amount of external energy required to meet the energy demand, thereby achieving optimal comfort, but spending only what is necessary.

An intelligent hybrid topology has been used to create the model. First, clustering [40–42] has been performed, with the objective of obtaining data groups with similar behavioral characteristics. Then,

regression techniques [43–46] have been applied over each of the obtained groups. The performance of the developed model is very good in terms of all the operational aspects.

The next section describes a case study where the facility under study is detailed. The model approach is described in the next section, and the different stages of the process of model development are described; i.e., data processing, clustering, and regression. All the results are presented, and finally the conclusions are drawn and future lines of research are outlined.

## 2. Case Study

This research is based on the thermal installations from a bioclimatic house. The main aim of the Sotavento Galicia Foundation is to study new ways of using renewable energies. For this reason, Sotavento Experimental Wind Farm built this bioclimatic house and implemented different energy systems in it.

### *Sotavento Bioclimatic House*

This bioclimatic house (Figure 1) is located in Xermade council (Lugo), in Galicia (Spain). There are different renewable energy extractors installed around/on such a house, such as solar, wind or geothermal ones, and the main aim is to demonstrate that a house could be more environmentally-friendly.



**Figure 1.** Sotavento bioclimatic house.

Only the thermal installation is taken into account in this research. In Figure 2 all the thermal systems of the house are shown. They are divided into three different parts to isolate the generation, the accumulation, and the consumption. The generation includes solar (1), biomass (2), and geothermal (3) energies; the geothermal system is divided in the horizontal ground collector and the heat pump. The accumulation has two isolate tanks, one for the storage of solar energy (4), and another for domestic hot water (DHW) (5) and the heater system. This part of the installation also includes the preheating of domestic hot water (8) that is made using the solar tank to heat the water before going inside the DHW tank. Finally, the consumption element is divided into the heater system (6) and DHW (7).

The bioclimatic house also distributes the electrical energy generated from renewable energy sources (wind and photovoltaic), but this feature is not considered in this paper as it is not part of the research.

Figure 3 shows the layout of the described thermal solar installation. This energy is generated in the solar panel collectors (2 strings of 4 panels each one), shown on the left side of the figure; 2.5 m<sup>2</sup> each panel, with a total of 20 m<sup>2</sup>, model *SchücoSol S.2*. Solar energy accumulation is represent by the tank on the right, with a capacity of 1000 L. The schematic includes all the valves and pumps that the system needs to work.

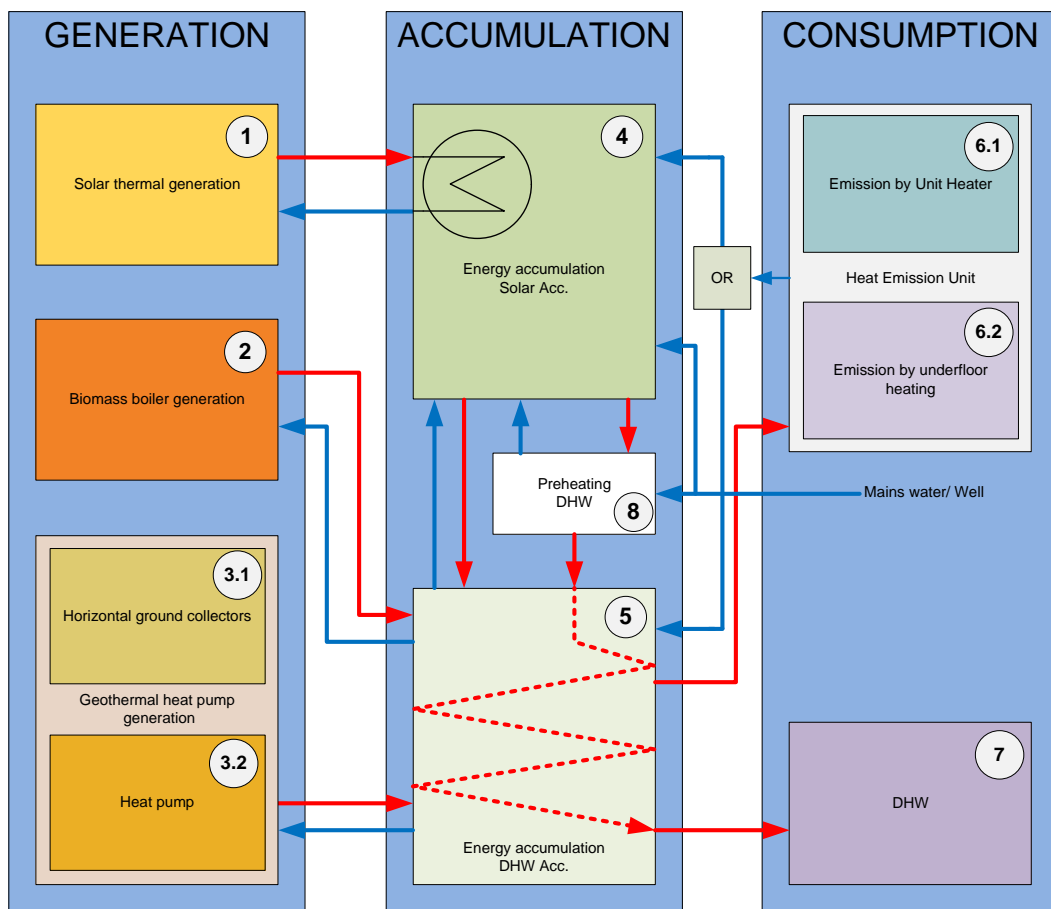


Figure 2. Thermal energy installation schematic.

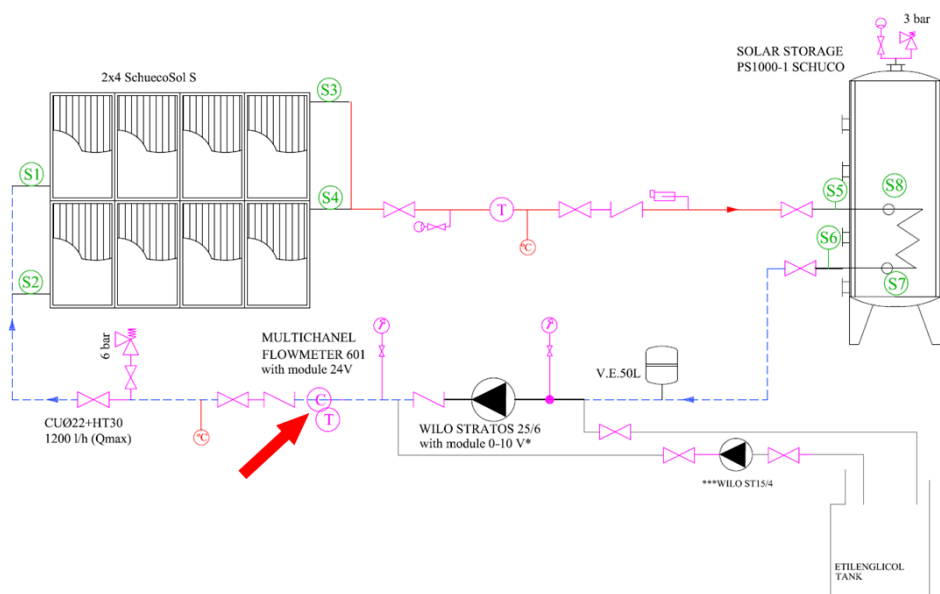


Figure 3. Solar thermal energy layout.

This research focuses on thermal solar generation; thus, only the temperature sensors S1, S2, S3, and S4 are used; they are RTD (PT1000) temperature sensors. However, the input and output

temperatures are not the only variables taken into account; others include the flow-meter (marked by the arrow in Figure 3), a *Multical*<sup>®</sup> 403, and the solar radiation sensor (a PYR-P sensor), deployed outside of the house

### 3. Model Approach

The basic model of the proposal is shown in Figure 4, where the output is set as the output temperature of the lower panel (S4). For this research, the output of the upper panel is not used. The inputs of the model are the input temperatures (lower and upper panels, S1 and S2), the flow of the etilenglicol used in the collectors, and the solar radiation.

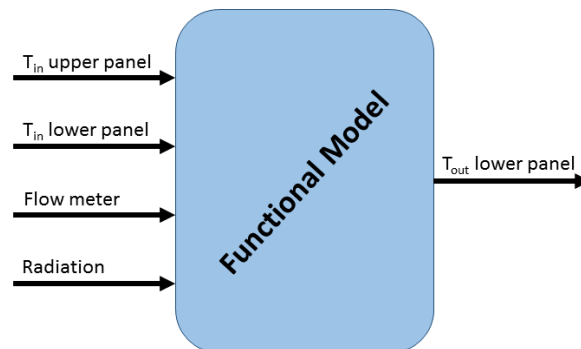


Figure 4. General schema of the functional model.

In this research, a hybrid intelligent model has been chosen to increase the accuracy of the output prediction. This type of model divides the dataset into different subsets (or clusters), and then, a regression technique is used to predict the output. Figure 5 represents the internal layout of the hybrid model; each cluster model used an intelligent technique selected from artificial neural network, polynomial regression support vector regression.

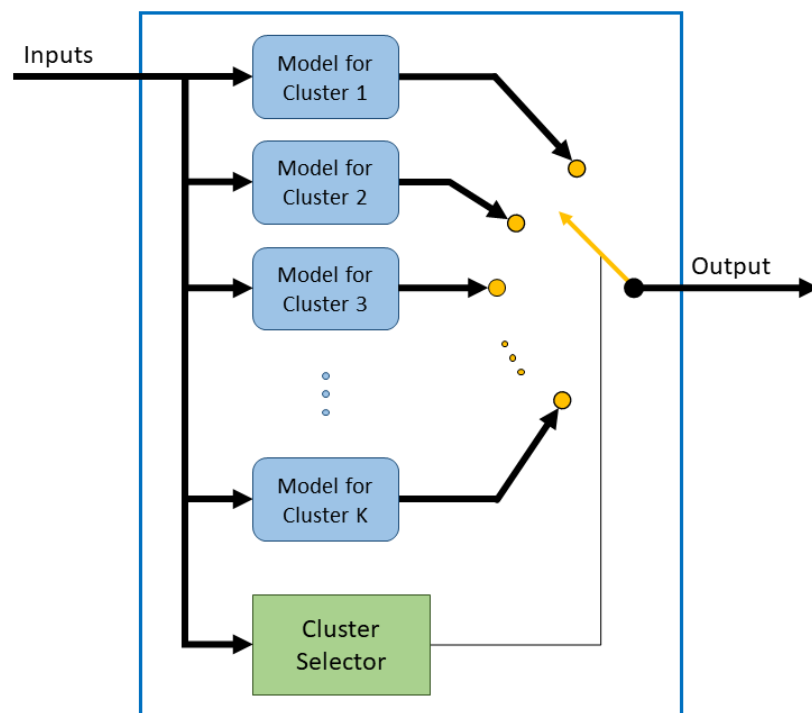


Figure 5. Internal schematic to achieve the hybrid model.

The diagram in Figure 6 represents the process followed to train the hybrid intelligent model. Firstly, different clusters were created, and then, a regression phase is made for each cluster. This regression was trained using *k-fold* cross validation that divides each data cluster into *k* groups and trains *k* models with *k-1* groups and uses the other groups for testing. After all the *k* models were tested, all the cluster data were used for testing, and the error of the specific regression technique is calculated with all the data. The *k-fold* cross validation is shown in Figure 7.

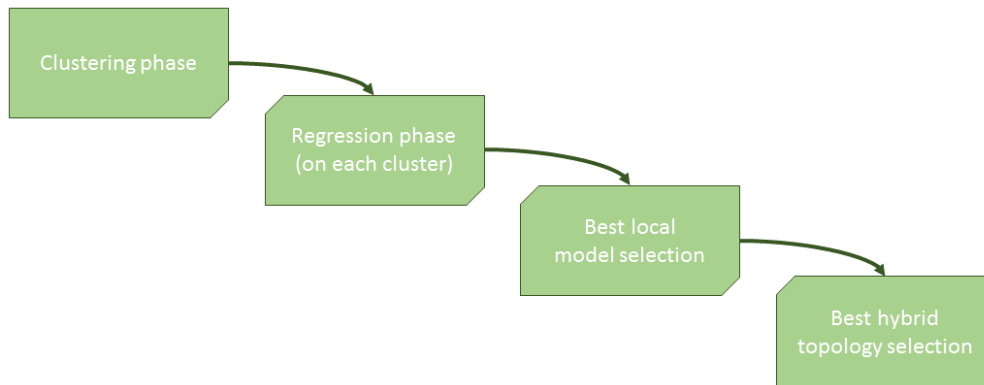


Figure 6. Flowchart of the hybrid model creation phases.

In the third step of Figure 6, the best regression algorithm is selected for each cluster. As different regression algorithms were used, it is necessary to choose the best one based on the error achieved in the training phase (described below). Moreover, some of the regression techniques tested have several tune parameters, and all of these configurations were considered as different algorithms.

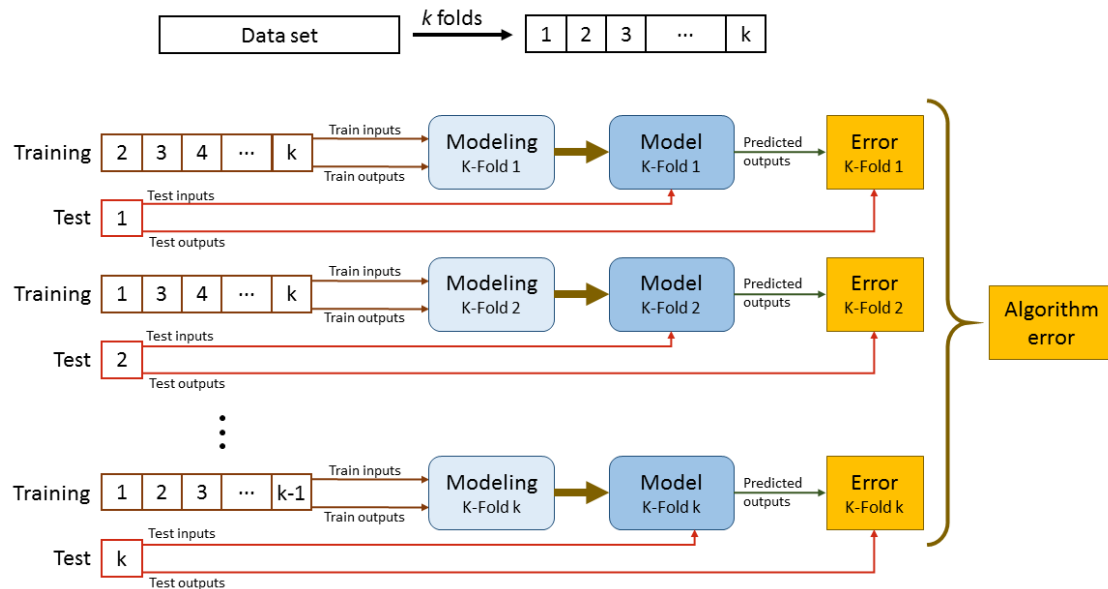


Figure 7. K-fold training and test data selection.

The last step to creating the hybrid model was the selection of the best hybrid configuration. For this purpose, a different dataset was used, isolated from the beginning of the regression phase in order to test all the hybrid configurations. The best one was chosen on the basis of the error achieved with this validation dataset.

### 3.1. Data Processing

The dataset was collected from a monitoring system that takes samples every 2 s; however, the mean value of all collected samples is stored every 10 min.

The entire dataset has been preprocessed before starting the training phase. Firstly, the incorrect samples were removed, which correspond to error in the communication process between the data acquisition system and sensors. Then, only the samples recorded when the panels were working have been considered (the radiation sensor value and the flow meter used in this step to avoid the samples when the system was not operating). The original dataset had 52,689 samples, but the number reduced to 26,665 samples after preprocessing.

The data were normalized to obtain the new sample values in the range 0 to 1. The Max-Min Scaler [47,48], Equation (1), is used to change all the samples ( $Data_j$ ) to new values ( $Data_{j_{new}}$ ).

$$Data_{j_{new}} = \frac{Data_j - \min(Data)}{\max(Data) - \min(Data)}. \quad (1)$$

To select the best hybrid topology, 5% of the samples in each cluster have been isolated from the dataset. The isolated data have been used at a later stage to verify the performance of all the hybrid configurations and to select the best hybrid topology. Moreover, to validate the obtained model, 367 samples that represent 5 operation days were isolated. These samples were not used in the hybrid model creation process. Instead, they were used only in the final step to have a realistic prediction error measurement for the hybrid intelligent model.

### 3.2. K-Means Algorithm

K-means algorithm was chosen to perform clustering and create different groups in the dataset. Different centroids were created to defined the clusters in the hyperspace; the samples are assigned to each cluster depending on the distance to these centroids [40–42]; the most common distance used is the Euclidean distance. The algorithm divides the data in the number of clusters ( $K$ ) defined by the user.

Once the final centroids have been defined, the computational cost needed by the k-means algorithm to assign new samples to each cluster is very small. However, the training time depends on the desired number of clusters and the number of training samples. The aim of the training is achieved when the final centroids [49] do not change. The training procedure can be explained as follows:

- Randomly choose the first set of centroids from the whole dataset. Since the centroids are defined as the center of the clusters, at the beginning there are no clusters.
- The cluster samples are defined with the samples that are nearest to each centroid.
- Once the clusters are defined, the centroids are swapped to the center of each cluster.

The procedure is repeated (the last two steps) until the centroids are in the same position at least to times during the training procedure. It is necessary to store the centroids to use the k-means algorithm with new samples.

### 3.3. Artificial Neural Networks

The artificial neural networks (ANN) algorithm is an artificial intelligent technique used for regression or classification. This algorithm was developed using the biological neuron model to create the basic unit component, the artificial neuron. The algorithm is called ANN because it has some artificial neurons inside, connected similarly to the biological ones.

Each neuron's input has a weight factor that allows for a different reaction to each input. Moreover, the neuron has an activation function that calculates its output using the inputs. An ANN model is able to generalize from the learning cases during the training phase [43,44]. The ANN can be used to perform complex functions thanks to its different activation functions.

The output of the activation function is called the excitation level [45], and it is normally in the range 0 to 1, or  $-1$  to 1. The configuration of the ANN includes the number of neurons, its activation functions, and its organization. The neurons are organized in layers; all the neurons that have the same inputs and outputs are in the same layer.

The multilayer perceptron is a basic feed-forward topology; the signal goes in the same way from the inputs to the outputs. The input and the output layers are directly connected to the inputs and outputs of the model; the hidden layers are the other layers that are only connected internally. In regression, the linear activation function is commonly used for the output neuron, while in the other neurons the tan-sigmoid function could be used.

### 3.4. Polynomial Regression

The polynomial regression algorithm is defined as the summation of several linear functions. Different degrees for the inputs defined the basis functions, and the maximum degree is called the degree of the polynomial.

Equations (2) and (3) show two different degree polynomials for a two inputs model. Each basis function has its own coefficient ( $c_*$ ) that is adjusted in the training phase.

$$f(x) = c_0 + c_1 \cdot x_1 + c_2 \cdot x_2. \quad (2)$$

$$f(x) = c_0 + c_1 \cdot x_1 + c_2 \cdot x_2 + c_3 \cdot x_1 \cdot x_2 + c_4 \cdot x_1^2 + c_5 \cdot x_2^2. \quad (3)$$

### 3.5. Support Vector Machines for Regression

The support vector machine is a machine learning algorithm used for classification problems. When this algorithm is used regression purposes, it is called support vector regression (SVR). This technique uses a nonlinear transformation to create a high-dimensional representation of the data; then, in the case of SVR, the algorithm performs a linear regression with the new mapping data.

This paper uses a modification of the SVR algorithm that is called least square SVR (LS-SVR) [46]. The LS-SVR's performance is similar to that the original SVR algorithm [50]; it is only necessary to adjust two internal parameters: weight vector ( $\gamma$ ) and kernel width ( $\sigma$ ). Moreover, the LS-SVR includes an optimization function that automatically tunes these parameters.

## 4. Results

This section is divided into three different parts with the aim of presenting all the results of this research. Firstly, the clustering results show the selected hybrid topologies, with the clusters and the samples in each one. Then, different regression results were represented. Since three differently configured algorithms were used, only some of the results are shown. This part includes the best regression technique, with its error measurement for each cluster. Finally, the validation of the model is described, along with the best hybrid topology and the final model error values.

### 4.1. Clustering Results

As the best number of clusters was not known beforehand, the k-means clustering technique was used to divide the dataset several times. In Table 1 it can be seen that nine different hybrid topologies were created, dividing the dataset into 2, 3, 4, 5, 6, 7, 8, 9, and 10 clusters. It is also shown that the first column corresponds to the global model (no clusters).



**Table 1.** Number of samples in each created cluster.

	Global	Hybrid Model (Local Models)								
		2	3	4	5	6	7	8	9	10
CI-1	24983	6875	5701	653	652	619	616	615	613	611
CI-2		18109	6988	5063	4287	2405	2316	1817	1463	1314
CI-3			12295	7001	4928	2687	2349	2306	1899	1446
CI-4				12267	6498	4251	2646	2333	2195	2088
CI-5					8617	6499	2668	2633	2271	2244
CI-6						8522	6361	4379	2643	2246
CI-7							8029	4757	3636	2635
CI-8								6144	4117	2708
CI-9									6146	4171
CI-10										5520

The training of the k-means algorithm was made with random initial centroids, and for all the configurations the training was repeated 20 times to avoid local minimum. Moreover, the training phase includes a final condition to avoid clusters with less than 15 samples; however, in this research, the smallest cluster has 611 samples.

#### 4.2. Modeling Results

As there are three different regression techniques, the modeling results are divided into three parts, each showing the results of a different algorithm.

##### 4.2.1. Artificial Neural Networks

All the tested ANNs have the same configuration; the input layer has four neurons (as many as the model's inputs), the internal layer has a varying number of neurons inside (this parameter is configurable), and the output layer has one neuron (as the model has only one output). The output layer neuron has a linear activation function, and the rest of the neurons in the ANN use a tan-sigmoid as their activation function. As it was said, the internal neuron number was varied to achieve the optimal one; 15 different models have been tested, each one with different neurons in the hidden layer.

Table 2 shows the error distribution through the clusters. In this case, it presents the mean absolute error (MAE) calculated for ANN with seven neurons in the hidden layer. There are a total of 15 MAE tables for artificial neural networks, as 15 different configurations have been tested. The error values have been calculated using 10 k-fold cross validation; this implies 10 different models must be trained before the error is calculated. Moreover, four different error measurements have been calculated: MAE, MSE (mean squared error), MAPE (mean absolute percentage error), and NMSE (normalized mean squared error).

**Table 2.** Mean absolute error (MAE) using an artificial neural network (ANN) with seven neurons in the hidden layer.

	Global	Hybrid Model (Local Models)								
		2	3	4	5	6	7	8	9	10
CI-1	0.0296	0.0079	0.0063	0.0196	0.0196	0.0189	0.0199	0.0199	0.0201	0.0189
CI-2		0.0358	0.0194	0.0050	0.0170	0.0054	0.0204	0.0219	0.0246	0.0163
CI-3			0.0413	0.0198	0.0043	0.0037	0.0051	0.0151	0.0206	0.0191
CI-4				0.0413	0.0332	0.0160	0.0037	0.0052	0.0201	0.0195
CI-5					0.0410	0.0343	0.1693	0.0034	0.0051	0.0239
CI-6						0.0405	0.0358	0.0271	0.0035	0.0050
CI-7							0.0405	0.0348	0.0304	0.0035
CI-8								0.0462	0.0325	0.0248
CI-9									0.0466	0.0398
CI-10										0.0465

#### 4.2.2. Polynomial Regression

Two different configurations have been trained with the Polynomial Regression algorithm; the first and the second degree polynomials have been used. As an example of this training, Table 3 shows the MAE obtained using second-order polynomial degree for each cluster. As explained, 10 k-fold cross validation is used to achieve the error measurement.

**Table 3.** MAE for second-degree polynomial regression algorithm.

	Global	Hybrid Model (Local Models)								
		2	3	4	5	6	7	8	9	10
CI-1	0.0314	0.0091	0.0064	0.0253	0.0471	0.0204	0.0210	0.0208	0.0213	0.0206
CI-2		0.0428	0.0210	0.0072	0.0172	0.0058	0.0283	0.0214	0.0181	0.0231
CI-3			0.0513	0.0208	0.0045	0.0084	0.0056	0.0160	0.0211	0.0245
CI-4				0.0467	0.0362	0.0181	0.0100	0.0056	0.0217	0.0228
CI-5					0.0480	0.0370	0.0198	0.0038	0.0056	0.0255
CI-6						0.0535	0.0396	0.0280	0.0040	0.0054
CI-7							0.0608	0.0934	0.0746	0.0040
CI-8								0.0514	0.0341	0.0449
CI-9									0.0539	0.0442
CI-10										0.0641

#### 4.2.3. Support Vector Machines for Regression

An error measurement for least square support vector regression is shown in Table 4. In this case, the algorithm has only one configuration because the least square modification uses an auto-tune function to adjust the internal parameters. Following the same training process as in the other algorithms, 10 k-fold cross validation was used.



**Table 6.** Mean squared error (MSE) for each individual hybrid model; all values  $\times 10^{-4}$ .

	Global	Hybrid Model (Local Models)								
		2	3	4	5	6	7	8	9	10
CI-1	16.8972	3.1161	1.0321	5.9668	6.2402	5.6794	6.3455	6.3268	6.1555	5.9393
CI-2		21.8021	9.8955	0.4348	6.9023	0.6041	9.0305	8.0789	9.4405	12.1097
CI-3			26.6203	9.7590	0.3851	0.3093	0.5586	7.6646	7.6268	8.0194
CI-4				26.2704	19.9626	7.4397	0.2725	0.5744	10.5137	10.7156
CI-5					26.4265	21.2670	7.0769	0.2332	0.5387	11.0445
CI-6						26.1993	22.3688	14.7125	0.2308	0.4686
CI-7							25.9529	19.6236	15.8058	0.2294
CI-8								31.8514	18.4699	10.7052
CI-9									32.7775	25.8231
CI-10										33.1709

**Table 7.** Configuration for each individual hybrid model.

	Global	Hybrid Model (Local Models)								
		2	3	4	5	6	7	8	9	10
CI-1	ANN-14	ANN-5	ANN-14	ANN-9	ANN-4	ANN-10	ANN-12	ANN-6	ANN-9	ANN-9
CI-2		ANN-13	ANN-14	ANN-14	ANN-5	ANN-6	LS-SVR	LS-SVR	LS-SVR	LS-SVR
CI-3			ANN-14	ANN-10	ANN-12	Poly-1	ANN-14	ANN-4	ANN-8	ANN-4
CI-4				ANN-15	ANN-14	ANN-3	Poly-1	ANN-12	ANN-6	ANN-10
CI-5					ANN-14	ANN-8	ANN-3	ANN-6	ANN-12	LS-SVR
CI-6						ANN-15	ANN-9	ANN-8	ANN-6	ANN-10
CI-7							ANN-14	ANN-10	LS-SVR	ANN-6
CI-8								ANN-13	ANN-11	LS-SVR
CI-9									ANN-14	ANN-11
CI-10										ANN-10

4.3. Validation Results

With the aim of selecting the best hybrid configuration (the optimal clusters number), a test has been performed using the testing dataset. This data were created with the 5% of the data of each cluster. This data were used as new input data for the nine different hybrid models, and also for the global model. Inside the hybrid model, each new sample was assigned to its local model using the euclidean distance to each cluster centroid, and the output is predicted to calculate the model error. Table 8 shows different error values to compare the hybrid configurations.

**Table 8.** Error values for the different hybrid configurations.

	Global	Hybrid Model (Local Models)								
		2	3	4	5	6	7	8	9	10
MSE	0.0016	0.0016	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0014	0.0014
MAE	0.0272	0.0270	0.0258	0.0257	0.0258	0.0259	0.0259	0.0255	0.0247	0.0250
NMSE	0.1116	0.1095	0.1038	0.1023	0.1036	0.1030	0.1047	0.1005	0.0977	0.1001

The best hybrid configuration is the one that divides the model internally into nine local models. In Table 7, it is possible to see the different algorithms and their configurations used in the final hybrid

model, including artificial neural networks with 6, 8, 9, 11, 12, and 14 neurons, and least squared support vector regression.

Moreover, to test the final hybrid configuration, five different subsets were tested. Each subset represents the data collected over the whole day, chosen randomly from the initial dataset, and isolated from the whole process described before. Figure 8 shows the variation of the real lower solar panel output temperature (blue continuous line) and the variation predicted by the model (green dashed line). The following error values have been calculated for these validation days; these values are not normalized to test the real operation of the model, but the normalized ones are included in *italics*.

- Mean squared error – MSE = 30.5010 (*0.0014*);
- Mean absolute error – MAE = 3.8027 (*0.0255*);
- Normalized mean squared error – NMSE = 0.1144;
- Mean absolute percentage error – MAPE = 14.3964 (*11.3191*).

In order to validate the innovative feature of the hybrid model, several ANN, polynomial, and SVR models have best experimented on using the whole dataset (global model). The results of this combination of models are presented in Table 9.

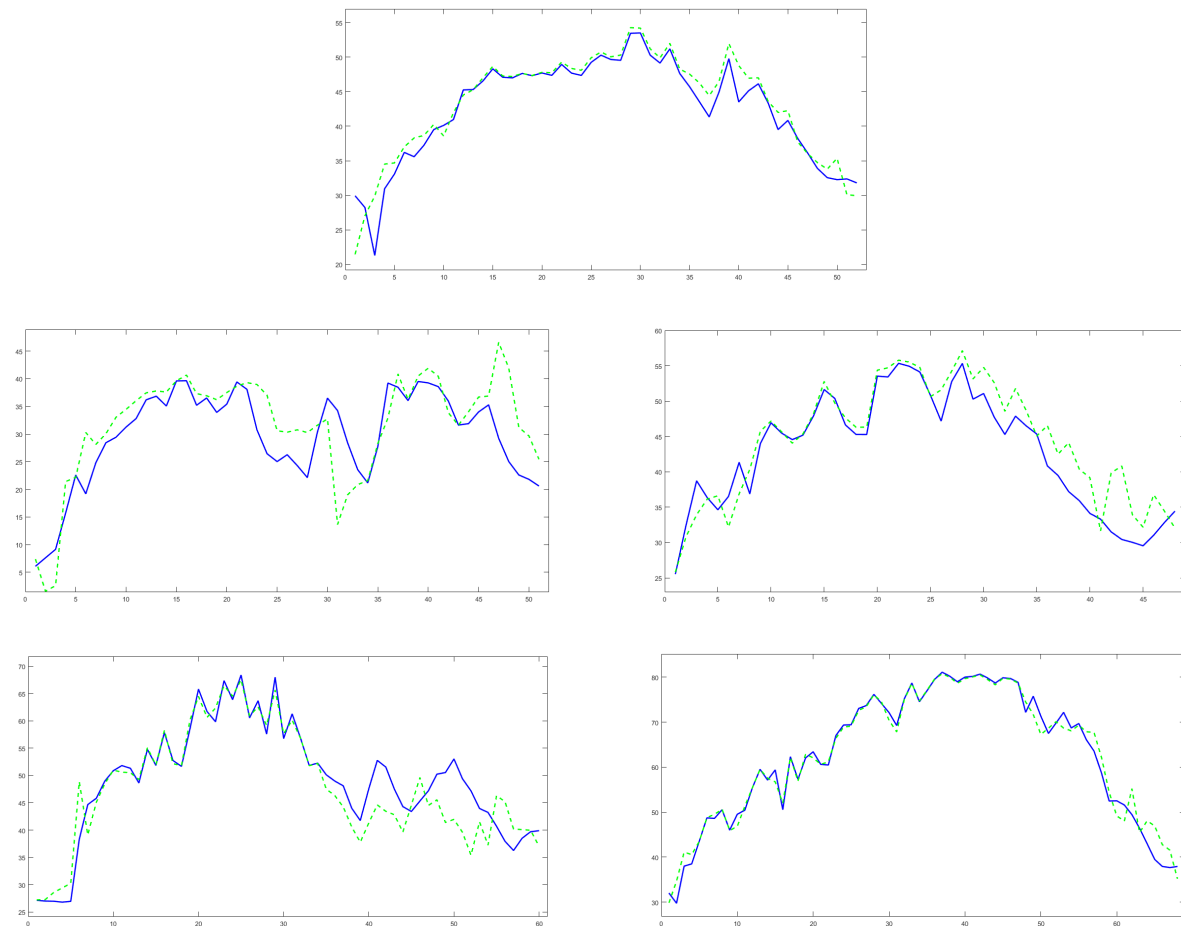


Figure 8. Validation days used to test the final hybrid model.

**Table 9.** Hybrid intelligent approach vs. global models with different regression algorithms.

	Hybrid Model with 9 Clusters	Global Models								
		LS-SVR	Poly-1	Poly-2	ANN-3	ANN-5	ANN-6	ANN-9	ANN-11	ANN-13
MSE	0.0014	0.0018	0.0024	0.0025	0.0022	0.0020	0.0020	0.0020	0.0018	0.0017
MAE	0.0247	0.0283	0.0324	0.0314	0.0320	0.0304	0.0298	0.0288	0.0282	0.0279
NMSE	0.0977	0.1161	0.1489	0.1571	0.1415	0.1279	0.1240	0.1277	0.1117	0.1073

## 5. Conclusions and Future Works

The hybrid intelligent model described in this research predicts the output temperature of a solar panel, taking into account the input temperature, the flow through the panel, and the solar radiation. This type of model could be used to measure, for example, the thermal energy absorbed by a solar collector without using thermal energy measurement equipment.

The model has been created with a real dataset recorder over two years to ensure that all climatology conditions are included in the dataset. Moreover, different subsets were separated from the beginning of the modeling process to validate and test the final model. The testing was performed with the 5% of the samples that had been isolated from each cluster; this test has made it possible to select the best hybrid configuration that has nine local models with artificial neural networks and support vector machines for regression.

The validation dataset has been isolated from the rest of the dataset, at the very beginning of data processing, and it therefore does not consider the clusters. The validation test has been performed with new data and the model has been used in real time; each sample is used as input and internally assigned to a local model to calculate the output. The performance values obtained in this test represent a prediction with less than 4 °C in MAE or less than 14.5% in MAPE (0.0255 and 11.3191 with normalized values). The obtained MSE was 30.5010, and the NMSE was 0.1144. These results demonstrate that the approach predicts more accurate values in comparison to global models.

Regarding future lines of research, it would be interesting to consider increasing the predicted horizon in order to predict the signal values in a future time. Moreover, it may be possible to create new models for the rest of the systems; this research only focused on solar thermal energy, but the bioclimatic house has many systems that could be studied.

**Author Contributions:** Data curation, E.J. and R.A.V.V.; methodology, J.A.L.V. and T.P.; project administration, M.-I.F.-I. and J.L.C.-R. (Jose Luis Calvo-Rolle); software, J.-L.C.-R. (José-Luis Casteleiro-Roca) and A.G.-B.; validation, H.Q. and S.T.-Á.; writing—original draft, P.C. and A.-J.P.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Rosenzweig, C.; Solecki, W.; Hammer, S.A.; Mehrotra, S. Cities lead the way in climate-change action. *Nature* **2010**, *467*, 909–911. [[CrossRef](#)] [[PubMed](#)]
- Dameri, R.P.; Cocchia, A. Smart city and digital city: Twenty years of terminology evolution. In Proceedings of the X Conference of the Italian Chapter of AIS, ITAIS, Milan, Italy, 14 December 2013; pp. 1–8.
- Vega Vega, R.A.; Chamoso-Santos, P.; Briones, A.G.; Casteleiro-Roca, J.L.; Jove, E.; del Carmen Meizoso-López, M.; Rodríguez-Gómez, B.A.; Quintián, H.; Herrero, Á.; Matsui, K.; et al. Intrusion Detection with Unsupervised Techniques for Network Management Protocols over Smart Grids. *Appl. Sci.* **2020**, *10*, 2276. [[CrossRef](#)]
- Dameri, R.P. Smart city definition, goals and performance. In *Smart City Implementation*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 1–22.
- Caragliu, A.; Del Bo, C.; Nijkamp, P. Smart cities in Europe. *J. Urban Technol.* **2011**, *18*, 65–82. [[CrossRef](#)]

6. Quintián, H.; Casteleiro-Roca, J.L.; Perez-Castelo, F.J.; Calvo-Rolle, J.L.; Corchado, E. Hybrid intelligent model for fault detection of a lithium iron phosphate power cell used in electric vehicles. In *International Conference on Hybrid Artificial Intelligence Systems*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 751–762.
7. Hollands, R.G. Will the real smart city please stand up? Intelligent, progressive or entrepreneurial? *City* **2008**, *12*, 303–320. [[CrossRef](#)]
8. Amin, M. Smart Grid. *Public Util. Fortn.* **2015**, 24–32.
9. Montgomery, L.; Magazine, E.H. *Home Automation: A Complete Guide to Buying, Owning and Enjoying a Home Automation System*; EH Publishing, Inc.: Framingham, MA, USA, 2014.
10. Sinopoli, J.M. *Smart Buildings Systems for Architects, Owners and Builders*; Butterworth-Heinemann: Oxford, UK, 2009.
11. Panteli, C.; Kylili, A.; Fokaides, P.A. Building information modelling applications in smart buildings: From design to commissioning and beyond A critical review. *J. Clean. Prod.* **2020**, *265*, 121766. [[CrossRef](#)]
12. Montero-Sousa, J.A.; Fernandez-Serantes, L.A.; Casteleiro-Roca, J.L.; Vilar-Martnez, X.M.; Calvo-Rolle, J.L. Energy storage management for generation-distribution facilities. *DYNA* **2017**, *92*, 140–141.
13. Jia, M.; Komeily, A.; Wang, Y.; Srinivasan, R.S. Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications. *Autom. Constr.* **2019**, *101*, 111–126. [[CrossRef](#)]
14. Jove, E.; Casteleiro-Roca, J.L.; Quintián, H.; Méndez-Pérez, J.A.; Calvo-Rolle, J.L. A fault detection system based on unsupervised techniques for industrial control loops. *Expert Syst.* **2019**, *36*, e12395, doi:10.1111/exsy.12395. [[CrossRef](#)]
15. Crespo-Turrado, C.; Casteleiro-Roca, J.L.; Sánchez-Lasheras, F.; López-Vázquez, J.A.; De Cos Juez, F.J.; Pérez Castelo, F.J.; Calvo-Rolle, J.L.; Corchado, E. Comparative Study of Imputation Algorithms Applied to the Prediction of Student Performance. *Log. J. IGPL* **2020**, *28*, 58–70. [[CrossRef](#)]
16. Casteleiro-Roca, J.L.; Gómez-González, J.F.; Calvo-Rolle, J.L.; Jove, E.; Quintián, H.; Gonzalez Diaz, B.; Mendez Perez, J.A. Short-Term Energy Demand Forecast in Hotels Using Hybrid Intelligent Modeling. *Sensors* **2019**, *19*, 2485, doi:10.3390/s19112485. [[CrossRef](#)] [[PubMed](#)]
17. Baruque, B.; Porras, S.; Jove, E.; Calvo-Rolle, J.L. Geothermal heat exchanger energy prediction based on time series and monitoring sensors optimization. *Energy* **2019**, *171*, 49–60. [[CrossRef](#)]
18. Jove, E.; Casteleiro-Roca, J.L.; Quintián, H.; Simić, D.; Méndez-Pérez, J.A.; Luis Calvo-Rolle, J. Anomaly detection based on one-class intelligent techniques over a control level plant. *Log. J. IGPL* **2020**, 1–17, in press. [[CrossRef](#)]
19. Basurto, N.; Arroyo, Á.; Vega, R.V.; Quintián, H.; Calvo-Rolle, J.L.; Herrero, Á. A Hybrid Intelligent System to forecast solar energy production. *Comput. Electr. Eng.* **2019**, *78*, 373–387, doi:10.1016/j.compeleceng.2019.07.023. [[CrossRef](#)]
20. Quintián, H.; Calvo-Rolle, J.L.; Corchado, E. A Hybrid Regression System Based on Local Models for Solar Energy Prediction. *Informatica* **2014**, *25*, 265–282. [[CrossRef](#)]
21. AlKandari, M.; Ahmad, I. Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. *Appl. Comput. Inform.* **2019**, 1–12, in press. doi:10.1016/j.aci.2019.11.002. [[CrossRef](#)]
22. Saberian, A.; Hizam, H.; Mohd Radzi, M.A.; Kadir, Z.; Mirzaei, M. Modelling and Prediction of Photovoltaic Power Output Using Artificial Neural Networks. *Int. J. Photoenergy* **2014**, *2014*, 1–10, doi:10.1155/2014/469701. [[CrossRef](#)]
23. İzgi, E.; Öztopal, A.; Yerli, B.; Kaymak, M.K.; Şahin, A.D. Short–mid-term solar power prediction by using artificial neural networks. *Sol. Energy* **2012**, *86*, 725–733, doi:10.1016/j.solener.2011.11.013. [[CrossRef](#)]
24. Casteleiro-Roca, J.L.; Perez, J.A.M.; Piñón-Pazos, A.J.; Calvo-Rolle, J.L.; Corchado, E. Intelligent Model for Electromyogram (EMG) Signal Prediction During Anesthesia. *J. -Mult.-Valued Log. Soft Comput.* **2019**, *32*, 205–220
25. Jove, E.; Gonzalez-Cava, J.M.; Casteleiro-Roca, J.L.; Méndez-Pérez, J.A.; Antonio Reboso-Morales, J.; Javier Pérez-Castelo, F.; Javier de Cos Juez, F.; Luis Calvo-Rolle, J. Modelling the hypnotic patient response in general anaesthesia using intelligent models. *Log. J. IGPL* **2019**, *27*, 189–201. [[CrossRef](#)]
26. Marrero, A.; Méndez, J.; Reboso, J.; Martín, I.; Calvo, J. Adaptive fuzzy modeling of the hypnotic process in anesthesia. *J. Clin. Monit. Comput.* **2017**, *31*, 319–330. [[CrossRef](#)] [[PubMed](#)]

27. Gonzalez-Cava, J.M.; Arnay, R.; Pérez, J.A.M.; León, A.; Martín, M.; Jove-Perez, E.; Calvo-Rolle, J.L.; Casteleiro-Roca, J.L.; de Cos Juez, F.J. A machine learning based system for analgesic drug delivery. In Proceedings of the International Joint Conference SOCO'17-CISIS'17-ICEUTE'17, León, Spain, 6–8 September 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 461–470.
28. Casteleiro-Roca, J.L.; Gomes, M.; Méndez-Pérez, J.A.; Alaiz-Moretón, H.; del Carmen Meizoso-López, M.; Rodríguez-Gómez, B.A.; Calvo-Rolle, J.L. Electromyogram prediction during anesthesia by using a hybrid intelligent model. *J. Ambient. Intell. Humaniz. Comput.* **2019**, 1–10, in press. [[CrossRef](#)]
29. Gutiérrez, C.G.; Rodríguez, M.L.S.; Díaz, R.Á.F.; Rolle, J.L.C.; Gutiérrez, N.R.; de Cos Juez, F.J. Rapid tomographic reconstruction through GPU-based adaptive optics. *Log. J. IGPL* **2019**, *27*, 214–226. [[CrossRef](#)]
30. Jove, E.; Aláiz-Moretón, H.; Casteleiro-Roca, J.L.; Corchado, E.; Calvo-Rolle, J.L. Modeling of bicomponent mixing system used in the manufacture of wind generator blades. In *International Conference on Intelligent Data Engineering and Automated Learning*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 275–285.
31. Casteleiro-Roca, J.L.; Jove, E.; Gonzalez-Cava, J.M.; Méndez Pérez, J.A.; Calvo-Rolle, J.L.; Blanco Alvarez, F. Hybrid model for the ANI index prediction using Remifentanil drug and EMG signal. *Neural Comput. Appl.* **2020**, *32*, 1249–1258. [[CrossRef](#)]
32. Casteleiro-Roca, J.L.; Barragán, A.J.; Segura, F.; Calvo-Rolle, J.L.; Andújar, J.M. Fuel Cell Output Current Prediction with a Hybrid Intelligent System. *Complexity* **2019**, *2019*, 1–10. [[CrossRef](#)]
33. Alaiz-Moretón, H.; Jove, E.; Casteleiro-Roca, J.L.; Quintián, H.; López García, H.; Benítez-Andrades, J.A.; Novais, P.; Calvo-Rolle, J.L. Bioinspired hybrid model to predict the hydrogen inlet fuel cell flow change of an energy storage system. *Processes* **2019**, *7*, 825. [[CrossRef](#)]
34. Casteleiro-Roca, J.L.; Javier Barragan, A.; Segura, F.; Luis Calvo-Rolle, J.; Manuel Andujar, J. Intelligent hybrid system for the prediction of the voltage-current characteristic curve of a hydrogen-based fuel cell. *Rev. Iberoam. Autom. Informática Ind.* **2019**, *16*, 492–501. [[CrossRef](#)]
35. Jove, E.; Casteleiro-Roca, J.; Quintián, H.; Méndez-Pérez, J.; Calvo-Rolle, J. Anomaly detection based on intelligent techniques over a bicomponent production plant used on wind generator blades manufacturing. *Rev. Iberoam. Autom. Informática Ind.* **2020**, *17*, 84–93. [[CrossRef](#)]
36. Vega Vega, R.; Quintián, H.; Calvo-Rolle, J.L.; Herrero, Á.; Corchado, E. Gaining deep knowledge of Android malware families through dimensionality reduction techniques. *Log. J. IGPL* **2019**, *27*, 160–176. [[CrossRef](#)]
37. Jove, E.; Blanco-Rodríguez, P.; Casteleiro-Roca, J.L.; Moreno-Arboleda, J.; López-Vázquez, J.A.; de Cos Juez, F.J.; Calvo-Rolle, J.L. Attempts prediction by missing data imputation in engineering degree. In Proceedings of the International Joint Conference SOCO'17-CISIS'17-ICEUTE'17, León, Spain, 6–8 September 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 167–176.
38. Aláiz-Moretón, H.; Castejón-Limas, M.; Casteleiro-Roca, J.L.; Jove, E.; Fernández Robles, L.; Calvo-Rolle, J.L. A Fault Detection System for a Geothermal Heat Exchanger Sensor Based on Intelligent Techniques. *Sensors* **2019**, *19*, 2740, doi:10.3390/s19122740. [[CrossRef](#)]
39. Jove, E.; Casteleiro-Roca, J.L.; Quintián, H.; Méndez-Pérez, J.A.; Calvo-Rolle, J.L. A new approach for system malfunctioning over an industrial system control loop based on unsupervised techniques. In The 13th International Conference on Soft Computing Models in Industrial and Environmental Applications, San Sebastian, Spain, 6–8 June 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 415–425.
40. Orallo, J.; Quintana, M.; Ramírez, C. *Introducción a la Minería de Datos*; Editorial Alhambra S.A.: Madrid, Spain, 2004.
41. Luis Casteleiro-Roca, J.; Quintián, H.; Luis Calvo-Rolle, J.; Méndez-Pérez, J.A.; Javier Perez-Castelo, F.; Corchado, E. Lithium iron phosphate power cell fault detection system based on hybrid intelligent system. *Log. J. IGPL* **2020**, *28*, 71–82, doi:10.1093/jigpal/jzz072. [[CrossRef](#)]
42. Jove, E.; Gonzalez-Cava, J.M.; Casteleiro-Roca, J.L.; Pérez, J.A.M.; Calvo-Rolle, J.L.; de Cos Juez, F.J. An intelligent model to predict ANI in patients undergoing general anesthesia. In Proceedings of the International Joint Conference SOCO'17-CISIS'17-ICEUTE'17, León, Spain, 6–8 September 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 492–501.
43. Harston, A.M.C.; Pap, R. *Handbook of Neural Computing Applications*; Elsevier Science: Amsterdam, The Netherlands, 2014.



44. Jove, E.; Alaiz-Moretón, H.; García-Rodríguez, I.; Benavides-Cuellar, C.; Casteleiro-Roca, J.L.; Calvo-Rolle, J.L. PID-ITS: an intelligent tutoring system for PID tuning learning process. In Proceedings of the International Joint Conference SOCO'17-CISIS'17-ICEUTE'17, León, Spain, 6–8 September 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 726–735.
45. del Brío, B.; Molina, A. *Redes Neuronales y Sistemas Borrosos*; Ra-Ma: Madrid, Spain, 2006.
46. Wang, L.; Wu, J. Neural network ensemble model using PPR and LS-SVR for stock market forecasting. In Proceedings of the Advanced Intelligent Computing Zhengzhou, China, 11–14 August 2011; pp. 1–8, doi:10.1007/978-3-642-24728-6\\_1. [[CrossRef](#)]
47. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
48. Tomás-Rodríguez, M.; Santos, M. Modelling and control of floating offshore wind turbines. *Rev. Iberoam. Autom. Informática Ind.* **2019**, *16*, 381–390. [[CrossRef](#)]
49. Viñuela, P.; León, I. *Redes de Neuronas Artificiales: Un Enfoque Práctico*; Pearson Educación—Prentice Hall: Madrid, Spain, 2004.
50. Steinwart, I.; Christmann, A. *Support Vector Machines*; Springer Publishing Company, Incorporated: New York, NY, USA, 2008.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).