

Article

# Prediction of Building's Thermal Performance Using LSTM and MLP Neural Networks

Miguel Martínez Comesaña <sup>1,\*</sup> , Lara Febrero-Garrido <sup>2</sup> , Francisco Troncoso-Pastoriza <sup>1</sup>   
and Javier Martínez-Torres <sup>3</sup> 

<sup>1</sup> Department of Mechanical Engineering, Heat Engines and Fluids Mechanics, Industrial Engineering School, University of Vigo, Maxwell s/n, 36310 Vigo, Spain; ftroncoso@uvigo.es

<sup>2</sup> Defense University Center, Spanish Naval Academy, Plaza de España, s/n, 36920 Marín, Spain; lfebrero@tud.uvigo.es

<sup>3</sup> Department of Applied Mathematics I. Telecommunications Engineering School, University of Vigo, 36310 Vigo, Spain; javmartinez@uvigo.es

\* Correspondence: migmartinez@uvigo.es

Received: 22 September 2020; Accepted: 20 October 2020; Published: 23 October 2020



**Abstract:** Accurate prediction of building indoor temperatures and thermal demand is of great help to control and optimize the energy performance of a building. However, building thermal inertia and lag lead to complex nonlinear systems is difficult to model. In this context, the application of artificial neural networks (ANNs) in buildings has grown considerably in recent years. The aim of this work is to study the thermal inertia of a building by developing an innovative methodology using multi-layered perceptron (MLP) and long short-term memory (LSTM) neural networks. This approach was applied to a public library building located in the north of Spain. A comparison between the prediction errors according to the number of time lags introduced in the models has been carried out. Moreover, the accuracy of the models was measured using the CV(RMSE) as advised by AHSRAE. The main novelty of this work lies in the analysis of the building inertia, through machine learning algorithms, observing the information provided by the input of time lags in the models. The results of the study prove that the best models are those that consider the thermal lag. Errors below 15% for thermal demand and below 2% for indoor temperatures were achieved with the proposed methodology.

**Keywords:** neural network; LSTM; MLP; thermal inertia; building performance

## 1. Introduction

The residential sector makes an important contribution to energy consumption worldwide, representing more than the 40% of the total energy use in the European Union (EU) [1]. The EU has established several guidelines and directives to improve energy performance in buildings, such as 2010/31/EU (EPBD) [2] and 2012/27/EU [3], requiring that new buildings comply with nearly zero-energy buildings (NZEB) by 2030 [4] and to reach a decarbonized and highly energy efficient building stock by 2050. Therefore, energy efficiency in buildings is of great importance to the overall sustainability. Knowing exactly and precisely the energy consumption of a building is the first step to be able to optimize its energy performance. However, forecasting the building energy consumption is a difficult issue that many authors have investigated in recent years [5–7].

Different methods have been developed to predict building energy demand. Traditionally, dynamic simulation has been used successfully [8–10] becoming a suitable tool, that enables the assessment of building performance and the calculation of energy use. There are several building energy performance simulation (BEPs) tools available, such as TRNSYS [11], EnergyPlus [12] or DOE-2 [13]. However, the use of dynamic simulation tools requires the knowledge and control of many different parameters of

the building, for example, envelope, materials properties, lighting, equipment, heating ventilation and air-conditioning (HVAC) systems and the behavior of the users, which makes the work of data collection arduous and increases the difficulty of obtaining exact results. The validation of these building energy simulation models is usually carried out with standard criteria such as the coefficient of variation of root mean square error (CV(RMSE)). Currently, these models are considered calibrated if they meet the criteria established by American Society of Heating Refrigerating and Air-Conditioning Engineers (ASHRAE) Guideline 14 [14], which states that a model is calibrated if its CV(RMSE) is below 15% monthly or 30% hourly. On the other hand, new tendencies in building prediction have emerged. Examples of this are the black box models, which stand out for their ability to train based on available data [15]. In this context, several data-driven models for building energy forecasting have been used more and more in recent years due to their robustness, resilience, strong yield and ease of implementation. Amongst them, the most popular data-driven approaches are the ones based on artificial neural networks (ANNs) [16,17].

ANNs are specific mathematical models that attempt to replicate the way a human neural network proceeds. The main and most important feature of these models is the ability to learn; that is, from known data, they are able to extract a pattern and then extrapolate the results to new data. Thus, neural networks have a remarkable ability to model the non-linear relationships between inputs and outputs, and for their massive interconnectivity [10,18,19]. Two of the most used ANNs in the context of building energy prediction are multi-layered perceptron (MLP) [19,20] and longshort-term memory (LSTM) neural networks [5,21]. MLP models are well known for being composed of several layers connecting the inputs to the specific output [10]. They have been applied to numerous scientific fields such as environment [22–24], econometrics [19,25], medical research [26], chemistry [27,28] and even building energy [10,29,30]. On the other hand, LSTMs are more current models, but their application has also expanded to a great number of fields. This type of ANN is characterized by a recurrent neural network (RNN) architecture [21,31]. RNNs differ from traditional feed-forward neural networks in that they have a hidden layer to consider connections with the previous values. This enables this type of neural network to model long-term dependencies; such as thermal inertia in buildings [32–34]. They encompass cyclic connections that make them, in principle, more suitable for modeling time sequence data than forward neural networks [35,36]. Some of the numerous fields where LSTM neural networks were applied are language modeling [36,37], speech recognition [35], tourism flow [38,39], sequential diagnosis [32,40] and energy analysis in buildings [21,41,42].

The aim of this work is to study the thermal inertia of a building by developing a methodology using MLP and LSTM neural networks. This procedure was applied to a public library building located in the north of Spain [43,44]. The available data are hourly observations of thermal demand and indoor temperatures of the building. Moreover, hourly observations of two weather variables (outdoor temperatures and solar radiation) and three time variables (hour of the day, day of the week and hour of the year) are also needed. A comparison between the prediction errors according to the number of time lags (each lag is one hour) introduced into the models, with respect to the studied variables, has been carried out. In addition, the accuracy of the models was measured using the CV(RMSE) as advised by AHSRAE. The novelty of this work lies in the analysis of the building inertia through two different artificial intelligence algorithms observing the information provided by the introduction of more time lags in the models.

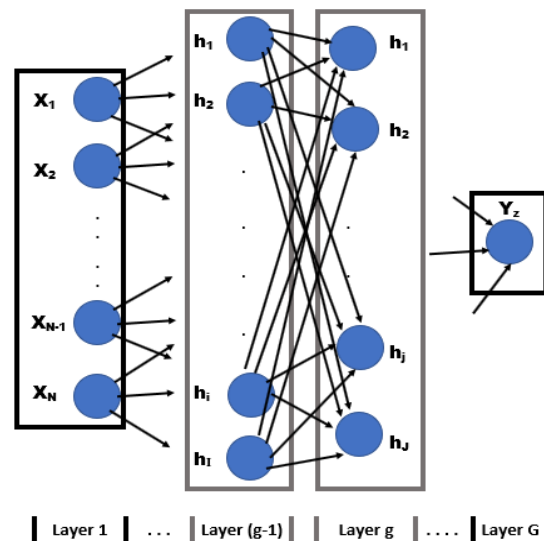
## 2. Materials and Methods

### 2.1. Artificial Neural Networks Developed

The different mathematical models that have been built to carry out the analysis are presented in this section. The models are of two different types of ANNs: An MLP and an LSTM neural network.

### 2.1.1. Multi-Layered Perceptron Neural Networks

In this work the structure of the MLP neural network is variable due to the use of different numbers of time lags. Depending on the structure of these artificial neural networks, different network models can be generated. The most commonly used is the so-called feed-forward model. This model is composed of several layers (see Figure 1). The first layer, or input layer, is where the model inputs are introduced, whereas the last layer, or output layer, is where the results of the trained network are given. In addition, among them, the number of intermediate layers (hidden layers) can be zero, one or more [18,20]. The particular network architectures are differentiated by the number of hidden layers and hidden neurons according to the complexity of the problem [45,46]. The grid of tested values for the hidden layers was between 0 and 4. In contrast, the grid of values for the hidden neurons was composed of specific numbers based on the formulas presented by Shin-ike [47], Doukim et al. [48] or Vujicic et al. [49], which take into account the size of the sample and the number of inputs and outputs of the network. Furthermore, in this case, due to the complexity of the problem, this grid of neurons had to be expanded with the values: 50, 100, 200, 500 and 1000.



**Figure 1.** Multi-layered perceptron (MLP) neural network architecture with  $N$  inputs,  $g$  hidden layers and univariate output.

The MLP neural network training is done with a backward propagation algorithm; errors are propagated through the network and adapted to the hidden layers. Thus, error-correction learning (actual system responses must be known) is used to train the ANN [29,50]. There are several ways to perform the training [19,51] but, the methodology used in this work consists of updating the weights with an average update of the weights (batch learning), which is achieved by incorporating all the patterns in the input file (an epoch) and accumulating all the weight updates. There is also a need for a stop criterion [52,53]. Although there are other options, such as a threshold for the mean square error (MSE) or a limitation on the maximum number of iterations, the most widely used is cross-validation. This method is most effective in stopping training when the best generalization is achieved. It consists of separating a small part of the training data and using it to evaluate the trained network. Thus, training should stop the moment the network performance, quantified by the mean square error (MSE), begins to decrease or stagnates [18–20,31].

Lastly, in all MLP models created in this paper the activation function selected was the reLU function (rectified linear unit:  $\max(0, x)$ ) [54], the kernel initializer was normal and the optimizing algorithm was the adaptive moment estimation (Adam) [55]. On the other hand, the batch size used to train the MLP neural networks was equal to 64 (mini-batch gradient descent [56]) and the patience (limit to stop the training if the performance of the model does not improve) was 100 epochs.

### 2.1.2. Long Short-Term Memory Neural Networks

LSTM is a recurrent neural network architecture specifically for modeling time sequences and their long-range dependencies more accurately than conventional RNNs. LSTM introduces specific units (memory blocks) into the recurrent hidden layer that store the temporal state of the network through self-connections. In addition, the memory block also contains units known as gates to control the flow of information [38,40]. A forget gate was incorporated into the memory blocks to prevent the LSTM model from processing continuous input flows without being segmented in subsequences. Furthermore, more complex LSTM networks also incorporate peephole connections between internal cells and their gates to learn the timing of the outputs [32,35].

The structure of traditional RNNs [57] can be presented by deterministic transitions from previous to current hidden states in form of a function:

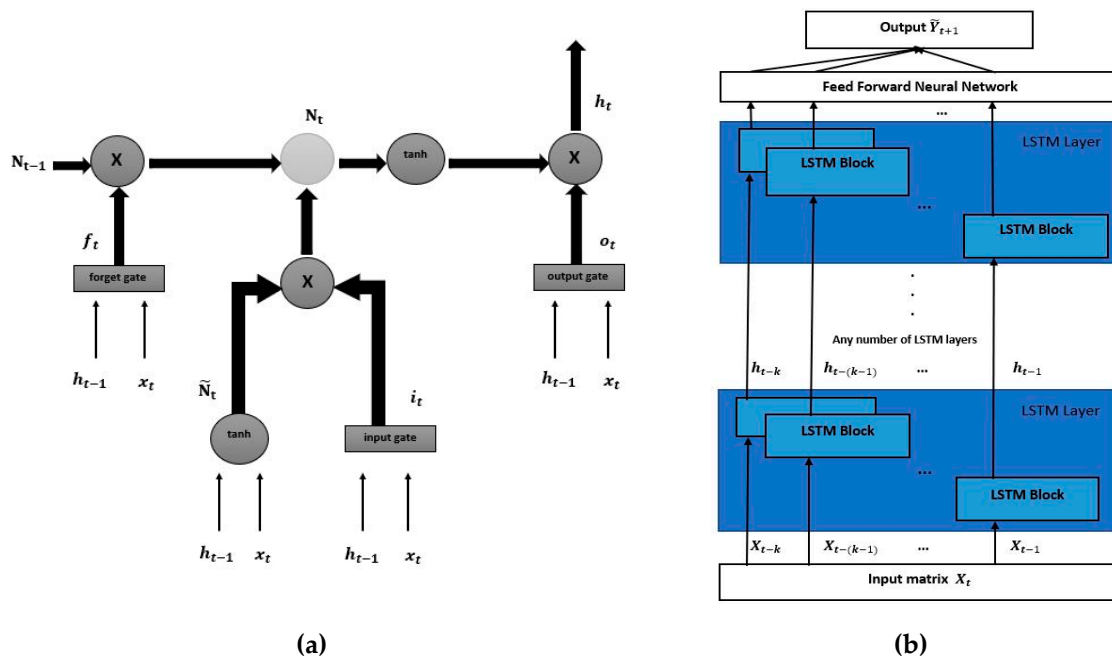
$$h_t^{l-1}, h_{t-1}^l \rightarrow h_t^l$$

where  $h_t^l$  represents a hidden state in layer  $l$  in timestep  $t$ .

However, LSTM counts with a complex dynamic that allows memorizing information for many timesteps. Long-term information is stored in a vector of memory cells  $m_t^l \in \mathbb{R}^n$  [21,37]. The LSTM networks are empowered to decide whether to overwrite, retrieve or maintain this information for the next timestep. Their architecture is the following:

$$h_t^{l-1}, h_{t-1}^l, m_{t-1}^l \rightarrow h_t^l, m_t^l$$

Let  $x_1, x_2, \dots, x_k$  be a classic input sequence (with  $k$  lags), where  $x_t \in \mathbb{R}^D$  denotes the vector of  $D$  real values at the timestep  $t$ . The LSTM architecture defines an input gate  $i_t$ , a forget gate  $f_t$  and an output gate  $o_t$ . At specific time  $t$ , as is observable in Figure 2a where an LSTM cell is shown, the hidden layer output would be  $h_t$ , the cell input state  $\tilde{N}_t$  and the cell output state  $N_t$  [21,38,39,41].



**Figure 2.** (a) Long short-term memory (LSTM) cell structure and (b) forecasting framework of LSTM neural network architecture.

The two values  $N_t$  (Equation (1)) and  $h_t$  (Equation (2)) that are transmitted to the next timestep, following the steps explained in [21,39], can be calculated as follows:

$$N_t = i_t \times \tilde{N}_t + f_t \times N_{t-1}, \tag{1}$$

$$h_t = o_t \times \tanh(N_t) \tag{2}$$

A particular LSTM neural network was built for the prediction of indoor temperatures and thermal demand of the studied building (see Figure 2b). In this case, the input matrix of the network would be the observed data  $X_t$  (see Table 1) and the one-dimensional output the predicted future data  $\tilde{Y}_{t+1}$  (Equation (3)). Once known  $h_t$  (Equation (2)) the network output will be:

$$\tilde{Y}_{t+1} = W_2 \times h_t + b, \tag{3}$$

where  $W_2$  is the weight matrix connecting the last hidden layer to the output layer and  $b$  the bias term in the output layer (see Figure 2b).

**Table 1.** Summary of the data set, with the available variables, used in this study.

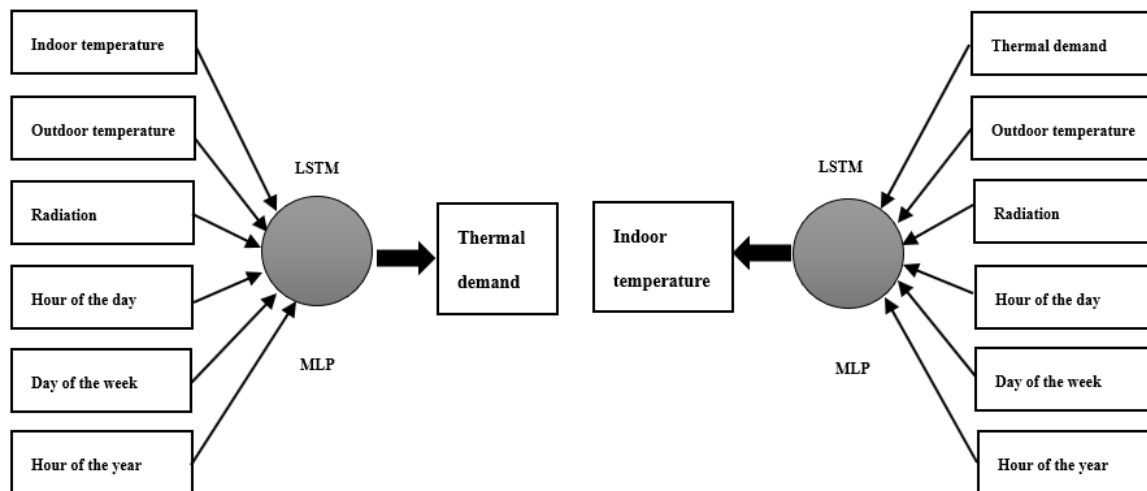
Date	Hour	Weekday	Yearhour	Outdoor Temp. (°C)	Radiation (W/m <sup>2</sup> )	Indoor Temp. (°C)	Thermal Demand (kWh)
05/02/2017 1:00	1	6	865	7.98	0.00	21.02	34.66
05/02/2017 2:00	2	6	866	8.28	0.00	20.99	36.85
05/02/2017 3:00	3	6	867	7.08	0.00	21.00	37.15
05/02/2017 4:00	4	6	868	5.98	0.00	21.04	38.78
05/02/2017 5:00	5	6	869	6.40	0.00	21.05	46.58
05/02/2017 6:00	6	6	870	6.42	0.00	21.03	37.81
05/02/2017 7:00	7	6	871	5.98	0.00	21.05	35.17
05/02/2017 8:00	8	6	872	6.52	1.17	21.06	33.07
05/02/2017 9:00	9	6	873	7.40	31.17	21.04	39.01
05/02/2017 10:00	10	6	874	7.70	180.17	21.04	42.19

Specifically, the LSTM architecture used in this study is also variable due to the different complexity of the models based on the number of time lags introduced. The variable parameters were the number of hidden and LSTM layers and the number of hidden neurons. In this case, the layer structures (LSTM and hidden layers) in the tested grid were: 2-1, 2-2, 3-1 and 3-2. On the other hand, the number of hidden neurons was a grid with eight values between 5 and 500. In addition, as in the MLP models, the activation function of the models was the reLU function [54], the kernel initializer was normal, the optimizing algorithm was Adam [55] and the batch size equal to 64 [56]. In order to avoid overfitting problems, they were trained with the same early stop as MLP neural networks (100 epochs) [52,53].

### 2.1.3. Pre-Processing Data

The data available in this study are hourly observations of two variables related to climate (outdoor temperature and solar radiation) and two related to the energy performance of the building (indoor temperature and thermal demand) between February and March of 2017 (see Table 1).

A continuous sample without large sets of missing values was necessary due to the use of hourly time lags as inputs in the models. This means that to train a model, besides the information of the independent variables at a given moment, the information of past instants of certain variables is also introduced. This study focuses on the prediction of thermal demand and indoor temperature of the studied building. The variables that have been used as explanatory variables of the models are presented in Figure 3. Although only the variables related to thermal conditions have been lagged, to add more information to the models, three time variables have been added (see Table 1). These three variables (hour of day, day of week and hour of year) do not provide any additional information being delayed due to their artificial nature.



**Figure 3.** The variables used as input into the models to predict the thermal demand of the building (left) and the indoor temperatures of the building (right).

Generally, as the number of time lags considered increases, the complexity of the model needs to change (there is more information to be extracted). Therefore, a process based on repetition (cross validation design; data are divided into training, test and validation [58]) was carried out to find the best parameters for each model and for each of the different lagged architectures studied. This process consists of repeating a prediction of the same test sample 10 times with each model (the second week of January 2019), training with the same data, to extract the average performance of each of them and being able to compare them. Thus, the best architecture for each structure of lags and to each of the mathematical models used (LSTM or MLP) is obtained.

After finding the best model for each lag architecture analyzed (see Tables 2 and 3), the thermal inertia of the building has been tested with two new weeks of 2019 (validation data):

- Sample 1: 28/01/2019–03/02/2019
- Sample 2: 11/03/2019–17/03/2019

## 2.2. Data Acquisition of the Experimental Case Study

The process of data acquisition, as well as the characteristics of the energy system of the building studied, is described in this section.

### 2.2.1. Building and HVAC System Description

The case study building where this new methodology has been proved is a public library located in the city of Vigo, in the north west of Spain. It is a three floor building with connected and large open areas that trigger temperature stratification inside the library. The building has a working floor area of 820 m<sup>2</sup> and the average capacity is 220 persons. This total usable floor area includes 412.3 m<sup>2</sup> on the ground floor, 233.5 m<sup>2</sup> on the first floor and 73.0 m<sup>2</sup> on the second floor. Due to the interconnected spaces and the large window area, the building experiences high thermal inertia.

The HVAC system is a ground-source heat pump (GSHP) distributed with a radiant floor (RF). The reversible heat pump is a CIATESA IZA 185 with a nominal heating capacity of 45 kW and nominal cooling capacity of 35.7 kW, with a COP of 3.63 for heating and EER of 3.25 for cooling. It has a glycol-water mixture flow rate of 8450 m<sup>3</sup>/h. The water storage tank contains 200 L. The GSHP is composed of six boreholes 100 m deep. The set point temperature for heating is 20 °C when the library is open and 17 °C when the library is closed. For cooling, the setpoint temperature is 26 °C when the library is open and 29 °C when closed. The tested heating period goes from October to May.

Further information about the building and its HVAC system can be consulted in [43,44,59].

### 2.2.2. Data Acquisition System

The data acquisition (DAQ) system in the building is further described in other articles by Cacabelos et al. [43,59] and Fernández et al. [44]. The rate of this system is one minute.

There are several wall module temperature sensors; six of them placed in the ground floor, four of them in the first floor and one in the second floor, all of them from the manufacturer Honeywell. In addition, the thermal zone temperature is calculated as the average of temperatures of every sensor from each floor. On the other hand, thermal energy consumption is measured with a thermal energy meter Multical 601 manufactured by Kamstrup.

The meteorological data were gathered from a weather station located at 42.17° N latitude and 8.68° E longitude at an altitude of 460 m. This station belongs to the weather station network from the Environment, Land Planning and Infrastructures Department, and it is located only 500 m from the building.

Further information about the DAQ and the weather station can be found in [43,44,59].

### 2.3. Validation and Error Measurement

The error measure considered in this paper to quantify the accuracy of each of the proposed models is the CV(RMSE) (coefficient of variation of the root mean square error):

$$CV(RMSE) = \frac{1}{\bar{Y}} \sqrt{\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N}} \quad (4)$$

This measure was used both to find the best structure depending on the number of time lags introduced in the models and, also, to compare the different models with the test samples (Tables 2 and 3). It has been used in similar studies such as that of Hong et al. [60] and Kuo et al. [61].

## 3. Results and Discussion

The inertia in the thermal conditions of the Science Library of the University of Vigo was analyzed through the variation of the CV(RMSE) (Equation (4)) in the predictions of thermal demand and indoor temperatures of the building. In this analysis the models were trained with two months of hourly observations considering different numbers of time lags (values of certain variables in past hours) with respect to the variables of study to make the predictions. In this way, two specific weeks in 2019 (sample 1 and sample 2) have been taken into account to analyze the thermal inertia of the building by comparing the error obtained by each model according to the number of time lags used.

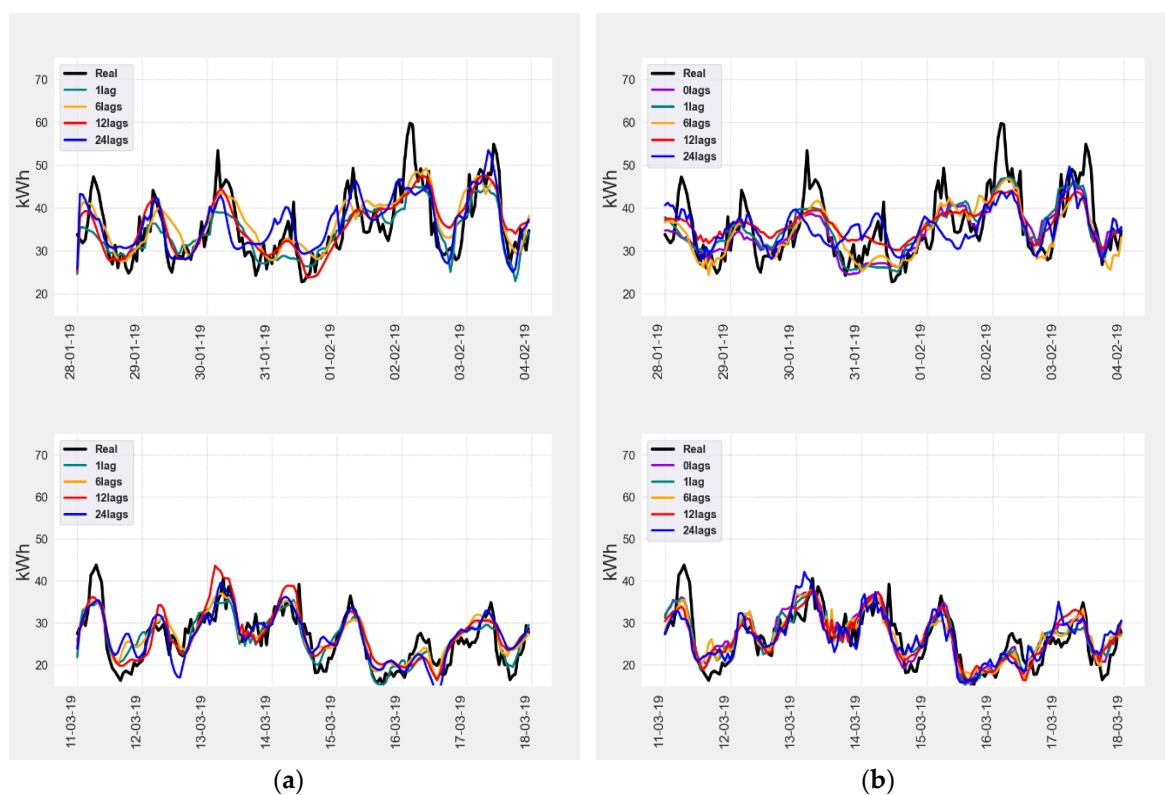
Section 3.1 presents the results to the thermal demand predictions and Section 3.2 the same analysis for indoor temperature predictions. In each section the evolution of the prediction error was analyzed based on the number of hourly time lags introduced in the models. The numerical results are the mean and the standard deviation of the CV(RMSE) obtained in 10 repetitions for each model and sample analyzed. All the specific architectures used in the models are shown in Tables 2 and 3. Lastly, all figures presented in this section were made with the Python programming language [62].

### 3.1. Thermal Demand Analysis

The results of the thermal demand predictions for the two samples studied, through which the thermal inertia of the building was analyzed in this section, are shown in Figure 4. Specifically, the results of the LSTM models are shown in Figure 4a and the results of the MLP models are shown in Figure 4b. Each time-lag structure analyzed is represented in the figures through the prediction with the lowest error among the 10 repetitions of the analysis.

LSTM models were capable of replicating the trend of the thermal demand of the building (except for certain peaks). They also yielded a small variation between the different replications of the analysis. This is shown in Table 2, with average errors of less than 20% and standard deviations

below 0.05, in the two samples considered in this study. In sample 1 (first row of Figure 4a) the lag structure that produced the best results is the one with 6 lags ( $CV(RMSE) = 15.26\%$ ). In addition, it is also the structure with the smallest variation in the 10 experiment repetitions. As shown in Table 2 and Figure 4a, in this sample the average errors are, in general, higher than those of sample 2. Although in Figure 4a all the models are close to the real values, as only the best predictions of each model are shown, in Table 2 the differences between their average performances are presented. Thus, the model that considers 12 lags is the architecture that obtained the highest average error ( $CV(RMSE) = 19.13\%$ ) and presents the largest variability in its results. In sample 2 (second row of Figure 4a), where the models were better adjusted to reality, the architecture that obtained the best average performance, as shown in Table 2, is the 1-lag architecture ( $CV(RMSE) = 14.35\%$ ). In this case, smaller errors are not accompanied by more stable results; the highest stability was achieved with the model with 24 lags. On the contrary, the model that obtained the highest error and the largest variability is the one that considers 12 lags ( $CV(RMSE) = 17.34\%$ ).



**Figure 4.** Results of thermal demand predictions in sample 1 (upper side) and sample 2 (bottom side) considering the different structures of time lags. Each architecture is represented by the prediction curve with the lowest error among the 10 repetitions: (a) LSTM models; (b) MLP models.

MLP models, as shown in Figure 4b, were also capable of predicting the actual thermal demand of the building with errors of less than 18%. As with LSTM models, certain peaks in the validation sample were not perfectly predicted and their variability among analysis repetitions was also small (below 0.02). However, these models yield better results than LSTM models (see Table 2). In sample 1 (first row of Figure 4b) the model with the lowest average error, as with LSTM models, is the model that considered 6 time lags ( $CV(RMSE) = 14.90\%$ ). In addition, this architecture is the one that produced the most stable results (see Table 2). On the contrary, the lag structure that adjusts reality in the worst way and produced the highest average error is the 12-lag structure ( $CV(RMSE) = 19.38\%$ ). It is also the structure that yields the most variable errors. In this case, a higher average error means more variability in predictions. In sample 2 (second row of Figure 4b), as in LSTM models, the predictions, on average, were better than in sample 1 (see Table 2). The architecture of lags that produced the lowest



average error is the architecture of 1 lag (CV(RMSE) = 14.17%); the same structure as with the LSTM models. In this situation it is not the most stable model. On the other hand, the model that yielded the worst performance is the model that considered 24 lags (CV(RMSE) = 17.43%); not being the most variable model. In the case of sample 2, as presented in Table 2, the models with the least errors and the most stable models do not coincide.

**Table 2.** Results of the thermal demand predictions based on the time lags introduced in the neural network model. LSTM errors are presented in the first three columns, whereas MLP errors are presented in the last three. The results shown summarize the 10 repetitions of the experiment, for each architecture and each sample analyzed. The mean of the CV(RMSE) obtained in each of them and the standard deviation (SD) are presented. The different structures of the neural networks that obtain the best accuracy are presented between brackets.

LSTM Model	Sample 1		Sample 2		MLP Model	Sample 1		Sample 2	
	CV RMSE	SD	CV RMSE	SD		CV RMSE	SD	CV RMSE	SD
No lags	-	-	-	-	No lags (50-1)	16.05%	0.007	14.77%	0.005
1 lag (20-20-20-10-10-1)	18.07%	0.047	14.35%	0.035	1 lag (50-1)	16.27%	0.007	14.17%	0.008
6 lags (10-10-5-1)	15.26%	0.006	15.28%	0.012	6 lags (100-1)	14.90%	0.004	15.41%	0.010
12lags (10-10-10-5-5-1)	19.13%	0.066	17.34%	0.038	12lags (8-1)	16.27%	0.006	14.87%	0.004
24 lags (500-500-250-1)	15.90%	0.009	16.15%	0.009	24 lags (200-200-200-200-1)	19.38%	0.017	17.43%	0.007

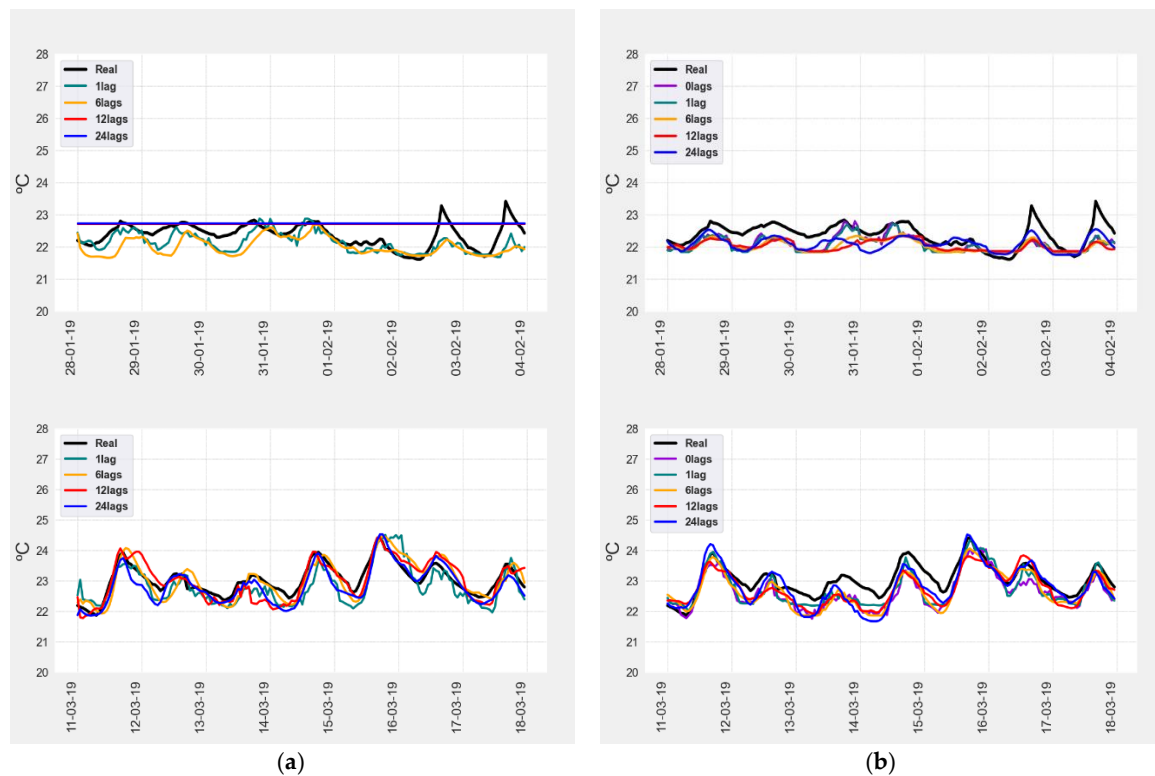
In the case of the thermal demand of the building studied, its thermal inertia is significant but up to a certain limit of time lags. Both LSTM and MLP models show that the best fit to reality is obtained from an architecture with 6 lags in sample 1 and 1 lag in sample 2 (see Table 2). Therefore, the inertia influence is greater with the data of sample 1. Moreover, the LSTM models, for both samples, present the worst performance with 12 lags; but the result, although not surpassing the best one, improved by introducing 24 lags. On the other hand, with MLP models the introduction of more time lags than the optimal ones cause an increase in the average error. MLP models obtained the worst time-lag structure in both cases with the 24-lag structure. Thus, although the best MLP models are more accurate than the best LSTM models, the worst MLP models are less accurate than the worst LSTM models. Regarding the stability of the results obtained, all the models show a small variability and, although there is not much difference between them, MLP models show smaller standard deviations. In addition, the model without time lags in any case obtained the best performance. This means that a thermal inertia exists and that the introduction of time lags in the thermal input variables provides valuable information.

### 3.2. Temperatures Analysis

The analysis of the thermal inertia of the building, for the two samples studied, through the predictions of the indoor temperatures of the building studied, is shown in Figure 5. Specifically, the results of the LSTM models are shown in Figure 5a and the results of the MLP models are shown in Figure 5b. As in the previous section, each time-lag structure studied is represented in the figures through the prediction with the lowest error among the 10 repetitions of the analysis.

LSTM models optimally predict the indoor temperatures of the building analyzed in this study. Table 3 shows that average errors (measured by the CV(RMSE)) are below 3% and their variability is small (below 0.005); both for sample 1 and for sample 2. In sample 1 (first row of Figure 5a), contrary to the previous section, the average errors are lower than those of sample 2. Furthermore, the lowest average error of this sample is obtained with the 1-lag structure (CV(RMSE) = 1.89%). In addition, although the models results are stable, this specific architecture shows the smallest standard deviation among all the structures analyzed (see Table 3). In this case, in Figure 5a it is observed that the two architectures with 12 and 24 lags are straight lines. This fact means that the introduction of more time lags does not provide useful information. Although the average error obtained by these two architectures is small, the models do not efficiently adjust to reality. Thus, the architecture that obtains the highest average error is the architecture with 24 lags (CV(RMSE) = 2.68%). Moreover, it is the least stable model structure in relation to the variability of the errors obtained (see Table 3). There is a correlation between the variability of the errors obtained and the average error; the greater the

variability, the greater the average error. On the other hand, in sample 2 (second row of Figure 5a) the 1-lag structure, as in sample 1, obtained the lowest average error ( $CV(RMSE) = 2.05\%$ ) and the smallest variability (see Table 3). As shown in Figure 5(a), with this sample the different models analyzed show a more similar behavior among themselves than in sample 1. Furthermore, unlike the first sample, the introduction of more time lags in the models eventually reduces the average errors; the best models, are those with 1 and 24 time lags (see Table 3). Therefore, the architecture with the highest average error is the 6-lag architecture ( $CV(RMSE) = 2.23\%$ ). In this case, there is no relation between average error and the variability of the results.



**Figure 5.** Results of the temperatures predictions in sample 1 (upper side) and sample 2 (bottom side) considering the different structures of time lags. Each architecture is represented by the prediction curve with the lowest error among the 10 repetitions: (a) LSTM models; (b) MLP models.

MLP models, presented in Figure 5b, once again provide better results than LSTM models. As shown in Table 3, their average errors are less than 2.5% and their variability is below 0.003. In sample 1 (first row of Figure 5b), the model structure with the lowest average error is the 24-lag structure ( $CV(RMSE) = 1.73\%$ ). On the contrary, the structure that produces the highest average error is the 6-lag structure ( $CV(RMSE) = 2.06\%$ ). In this case, and similarly to LSTM models, there is no direct relation between average error and variability in the results. In addition, as can be seen in the first row of Figure 5b (and the first row of Figure 5a), none of the models are able to replicate the peaks of the indoor temperatures at the weekend. Nevertheless, as shown in Table 3, the models do adjust the data trend efficiently. On the other hand, in sample 2 (second row of Figure 5b), as in LSTM models, the average errors obtained are higher than those of sample 1 (see Table 3). In this case, the architecture that yielded the lowest average error is the 12-lags architecture ( $CV(RMSE) = 1.90\%$ ). Additionally, it is also the architecture that presents the smallest variability among its results. With this sample, the introduction of time lags did provide valuable information because, as shown in Table 3, the best model structures are those with 12 and 24 lags. However, the lag structure that presents the highest average error was 6-lag structure ( $CV(RMSE) = 2.28\%$ ), slightly worse than the average error

of the model without any lag. This means that, although the introduction of time lags provides useful information, the basic model (regardless of thermal inertia) already provides good results.

**Table 3.** Results of the indoor temperatures predictions based on the time lags introduced in the neural network model. LSTM errors are presented in the first three columns, whereas MLP errors are presented in the last three. The results shown summarize the 10 repetitions of the experiment, for each architecture and each sample analyzed. The mean of the CV(RMSE) obtained in each of them and the standard deviation (SD) are presented. The different structures of the neural networks that provide the best accuracy are presented between brackets.

LSTM Model	Sample 1		Sample 2		MLP Model	Sample 1		Sample 2	
	CV(RMSE)	SD	CV(RMSE)	SD		CV(RMSE)	SD	CV(RMSE)	SD
No lags	-	-	-	-	No lags (5-5-5-1)	1.82%	0.001	2.21%	≈ 0
1 lag (10-10-10-5-5-1)	1.89%	0.002	2.05%	0.001	1 lag (5-5-5-5-1)	1.86%	0.001	2.03%	0.003
6 lags (5-5-5-2-2-1)	2.40%	0.003	2.23%	0.004	6 lags (4-4-4-4-1)	2.06%	0.001	2.28%	0.003
12lags (5-5-5-2-1)	2.42%	0.002	2.19%	0.005	12lags (6-6-6-6-1)	2.02%	≈ 0	1.90%	≈ 0
24 lags (5-5-5-2-1)	2.68%	0.004	2.10%	0.004	24 lags (12-12-12-1)	1.73%	≈ 0	2.01%	0.001

Regarding the results of the indoor temperatures of the building, it can be observed that, on the one hand, there are less differences between the performance of the models in the different samples than in the case of the thermal demand. Moreover, in comparison with the thermal demand results, the different errors obtained for each of the architectures analyzed during the repetitions are lower and less variable. This demonstrates that indoor temperature values are more constant throughout a year (always around the set point temperature of the building). On the other hand, there are differences between LSTM and MLP models (see Table 3). With LSTM models the thermal inertia is less significant than in the case of thermal demand. In the two samples analyzed the best model was the one with a single time lag. In contrast, with MLP models the thermal inertia is more significant than in the analysis in the previous section. In both samples, the best model structures were a 24-lag structure for sample 1 and a 12-lag structure for sample 2; they use information from a higher number of past hours. In this case, MLP models show better performance in both the best and the worst case. In addition, even with less variability in general concerning indoor temperature predictions, MLP models again produced the lowest standard deviations. Finally, as in the case of thermal demand predictions, models without time lags do not produce the lowest average error in any of the analyses presented. This indicates that the thermal inertia is also significant in the case of indoor temperatures predictions.

#### 4. Conclusions

A new application of machine and deep learning algorithms and a new methodology to analyze the influence of the thermal inertia of a building are presented in this paper. The study was conducted by analyzing monitored data of thermal demand and indoor temperatures of the Science Library of the University of Vigo. The data were completed with weather variables (outdoor temperatures and solar radiation) and three temporal variables (hour of the day, day of the week and hour of the year). The aim of this analysis is to study the importance of the thermal inertia of the building in predicting thermal demand and indoor temperatures. The methodology used in this study is based, on the one hand, on a classic machine learning model (MLP neural network) contrasting the improvement in the prediction accuracy provided by the thermal inertia of the building. On the other hand, as a comparison, a deep learning model (LSTM neural network) was carried out through an analogous analysis.

The research contribution of this paper is the application of mathematical methods to evaluate the influence of the thermal inertia of the building in the prediction errors of the thermal demand and the indoor temperatures of the building itself. The principal limitation of this research is the analysis of the thermal inertia of the building based on the difference in the errors rates of indoor temperatures depending on the number of time lags. The stable behavior of the indoor temperatures by default, produce that, in general, the prediction error is very low. This makes analysis and detection of the additional information that can provide the introduction of time lags in the models more difficult. On the other hand, the presented methods contribute with advantages over the already existing

research in building simulation and thermal inertia analysis of buildings. With traditional building simulations models, both to make predictions and to analyze the thermal inertia of the building, it is necessary to have a deep knowledge about the specific subject and control many different energy parameters. With the methods presented in this paper, known as black box models, it is possible to do all kinds of analyses without any specific knowledge, and also to do it faster than dynamic simulation methods. These models rely on having a significant amount of data to be able to extract information from it and extrapolate it. Additionally, the variables involved in this study are commonly monitored in numerous buildings. Thus, the methodology presented can be applied to evaluate the energy performance or the influence of thermal inertia of different buildings.

First, the results show that the black box models yield average prediction errors below the proposed values for considering calibrated models. The results also illustrate the greater accuracy of MLP neural networks in predicting both thermal demand and indoor temperatures. LSTM neural networks, created specifically to model time sequences, show a poorer performance than MLP models due to the peculiarity of this analysis. In this study the models were trained with only two months of data to try to predict future time sequences of the variable of interest. LSTM models are focused on analyses where predictions are made for time sequences immediately consecutive to the training data (as a time series). In this specific case, where the data are not continuous in time, LSTM models are not the most efficient. In this paper it was demonstrated that, for analyses similar to the one presented here, MLP models are more appropriate; they are more flexible if the available data, on the one hand, are not continuous in time or, on the other hand, are very few. In terms of thermal demand predictions, for both sample 1 and sample 2, the best MLP models are better than the best LSTM models. While the average errors of the MLP models were 14.90% and 14.17%, respectively, the average errors of the LSTM models were 15.26% and 14.35%. The same applies to indoor temperature predictions. The average errors of the best MLP models were 1.73% in sample 1 and 1.90% in sample 2, and those of the best LSTM models were 1.89% and 2.05%, respectively. Regarding the variability of the predictions, the results of the LSTM and MLP models are similar, but the MLP models, in most cases, have slightly smaller standard deviations.

Furthermore, the results of the thermal inertia analysis of the building are different in relation to the different models used. Although all the analyses show that the introduction of specific time lags reduces the prediction error, the influence of the thermal inertia of the building changes with LSTM and MLP models. In the case of thermal demand analysis, both LSTM and MLP models show the same result. In sample 1 the optimal number of time lags introduced in the model was 6 while in sample 2 the optimal structure was a model with 1 time lag. It was demonstrated that the inertia of the independent variables of these models (indoor temperatures, outdoor temperatures and solar radiation) does not extend too much in time. Differently, the results of LSTM and MLP models do not coincide with the indoor temperatures analysis. LSTM models, both of sample 1 and sample 2, have their optimum in 1 time-lag. In contrast, with MLP models the best structures were 24 time lags in sample 1 and 12 time lags in sample 2. Thus, MLP models are able to better extract the information provided by the introduction of more time lags; they are able to exploit the thermal inertia of the building.

From an energy point of view, the conclusion is that the thermal inertia of the building does provide useful information to the machine learning models by introducing lagged variables. This was demonstrated by the decrease, up to a certain limit, of average errors with the introduction of time lags. Therefore, it was also demonstrated that the dimension of the thermal inertia, measured by the maximum number of time lags that provides valuable information, can be detected by the methods presented in this paper. Moreover, the analysis showed that MLP models are more appropriated than LSTM models if the available data set is not large and continuous over time. For the two variables and the two samples analyzed the average error produced by the optimal MLP models are lower than those provide by the optimal LSTM models. Lastly, it was shown that there is more inertia in thermal demand data than in indoor temperatures data; the introduction of thermal demand lags,

due to the higher variability of their data, provides more useful information to the models than indoor temperature ones.

**Author Contributions:** Conceptualization, M.M.C. and L.F.-G.; methodology, M.M.C. and J.M.-T.; software, M.M.C. and F.T.-P.; validation, L.F.-G., J.M.-T. and F.T.-P.; formal analysis, M.M.C. and L.F.-G.; investigation, M.M.C. and L.F.-G.; resources, J.M.-T. and L.F.-G.; data curation, F.T.-P.; writing—original draft preparation, M.M.C. and L.F.-G.; writing—review and editing, J.M.-T. and F.T.-P.; visualization, M.M.C. and F.T.-P.; supervision, L.F.-G., J.M.-T. and J.T.-P.; project administration, M.M.C. and L.F.-G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Spanish Government (Science, Innovation and Universities Ministry) under the project RTI2018-096296-B-C21.

**Acknowledgments:** This research was supported by the Spanish Government (Science, Innovation and Universities Ministry) under the project RTI2018-096296-B-C21.

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

## References

- Li, D.H.W.; Yang, L.; Lam, J.C. Zero energy buildings and sustainable development implications—A review. *Energy* **2013**, *54*, 1–10. [CrossRef]
- Official Journal of the European Union. *Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the Energy Performance of Buildings*; EU: Brussels, Belgium, 2010.
- European Commission. *Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on Energy Efficiency, Amending DIRECTIVES 2009/125/EC and 2010/30/EU and Repealing Directives 2004/8/EC and 2006/32/EC*; EU: Brussels, Belgium, 2012.
- Nematchoua, M.K.; Marie-Reine Nishimwe, A.; Reiter, S. Towards nearly zero-energy residential neighbourhoods in the European Union: A case study. *Renew. Sustain. Energy Rev.* **2021**, *135*, 110198. [CrossRef]
- Bourdeau, M.; Zhai, X.q.; Nefzaoui, E.; Guo, X.; Chatellier, P. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustain. Cities Soc.* **2019**, *48*, 101533. [CrossRef]
- Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A review on time series forecasting techniques for building energy consumption. *Renew. Sustain. Energy Rev.* **2017**, *74*, 902–924. [CrossRef]
- Fumo, N.; Mago, P.; Luck, R. Methodology to estimate building energy consumption using EnergyPlus Benchmark Models. *Energy Build.* **2010**, *42*, 2331–2337. [CrossRef]
- Harish, V.S.K.V.; Kumar, A. A review on modeling and simulation of building energy systems. *Renew. Sustain. Energy Rev.* **2016**, *56*, 1272–1292. [CrossRef]
- Heiple, S.; Sailor, D.J. Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles. *Energy Build.* **2008**, *40*, 1426–1436. [CrossRef]
- Neto, A.H.; Fiorelli, F.A.S. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy Build.* **2008**, *40*, 2169–2176. [CrossRef]
- University of Wisconsin—Madison. *Solar Energy, L. TRNSYS, a Transient Simulation Program*; The Laboratory: Madison, WI, USA, 1975.
- Crawley, D.B.; Lawrie, L.K.; Winkelmann, F.C.; Buhl, W.F.; Huang, Y.J.; Pedersen, C.O.; Strand, R.K.; Liesen, R.J.; Fisher, D.E.; Witte, M.J.; et al. EnergyPlus: Creating a new-generation building energy simulation program. *Energy Build.* **2001**, *33*, 319–331. [CrossRef]
- James, J. Hirsch & Associates (JH). DOE-2 (version 2.2-047d). Available online: <http://www.doe2.com/> (accessed on 1 September 2020).
- ASHRAE. *Guideline 14-2014—Measurement of Energy, Demand, and Water Savings*; ASHRAE: Atlanta, GA, USA, 2014.
- Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; Pedreschi, D. A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv.* **2018**, *51*, 93. [CrossRef]
- Runge, J.; Zmeureanu, R. Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review. *Energies* **2019**, *12*, 3254. [CrossRef]
- Oh, S. Comparison of a Response Surface Method and Artificial Neural Network in Predicting the Aerodynamic Performance of a Wind Turbine Airfoil and Its Optimization. *Appl. Sci.* **2020**, *10*, 6277. [CrossRef]

18. Hung, N.Q.; Babel, M.S.; Weesakul, S.; Tripathi, N.K. An artificial neural network model for rainfall forecasting in Bangkok, Thailand. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 1413–1425. [[CrossRef](#)]
19. Guresen, E.; Kayakutlu, G.; Daim, T.U. Using artificial neural network models in stock market index prediction. *Expert Syst. Appl.* **2011**, *38*, 10389–10397. [[CrossRef](#)]
20. Fadare, D.A. Modelling of solar energy potential in Nigeria using an artificial neural network model. *Appl. Energy* **2009**, *86*, 1410–1422. [[CrossRef](#)]
21. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Trans. Smart Grid* **2019**, *10*, 841–851. [[CrossRef](#)]
22. Banerjee, P.; Singh, V.S.; Chattopadhyay, K.; Chandra, P.C.; Singh, B. Artificial neural network model as a potential alternative for groundwater salinity forecasting. *J. Hydrol.* **2011**, *398*, 212–220. [[CrossRef](#)]
23. Iglesias, C.; Martínez Torres, J.; García Nieto, P.J.; Alonso Fernández, J.R.; Díaz Muñoz, C.; Piñeiro, J.I.; Taboada, J. Turbidity Prediction in a River Basin by Using Artificial Neural Networks: A Case Study in Northern Spain. *Water Resour. Manag.* **2014**, *28*, 319–331. [[CrossRef](#)]
24. Anjos, O.; Iglesias, C.; Peres, F.; Martínez, J.; García, Á.; Taboada, J. Neural networks applied to discriminate botanical origin of honeys. *Food Chem.* **2015**, *175*, 128–136. [[CrossRef](#)]
25. Gil-Cordero, E.; Cabrera-Sánchez, J.-P. Private Label and Macroeconomic Indexes: An Artificial Neural Networks Application. *Appl. Sci.* **2020**, *10*, 6043. [[CrossRef](#)]
26. Nasser, I.M.; Abu-Naser, S.S. Predicting Tumor Category Using Artificial Neural Networks. *Int. J. Acad. Health Med Res. (IJAHMR)* **2019**, *3*, 1–7.
27. Curteanu, S.; Cartwright, H. Neural networks applied in chemistry. I. Determination of the optimal topology of multilayer perceptron neural networks. *J. Chemom.* **2011**, *25*, 527–549. [[CrossRef](#)]
28. Iglesias, C.; Anjos, O.; Martínez, J.; Pereira, H.; Taboada, J. Prediction of tension properties of cork from its physical properties using neural networks. *Eur. J. Wood Wood Prod.* **2015**, *73*, 347–356. [[CrossRef](#)]
29. Chae, Y.T.; Horesh, R.; Hwang, Y.; Lee, Y.M. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy Build.* **2016**, *111*, 184–194. [[CrossRef](#)]
30. Robinson, C.; Dilkina, B.; Hubbs, J.; Zhang, W.; Guhathakurta, S.; Brown, M.A.; Pendyala, R.M. Machine learning approaches for estimating commercial building energy consumption. *Appl. Energy* **2017**, *208*, 889–904. [[CrossRef](#)]
31. Rahman, A.; Smith, A.D. Predicting heating demand and sizing a stratified thermal storage tank using deep learning algorithms. *Appl. Energy* **2018**, *228*, 108–121. [[CrossRef](#)]
32. Lipton, Z.; Kale, D.; Elkan, C.; Wetzel, R. Learning to Diagnose with LSTM Recurrent Neural Networks. *arXiv* **2015**, arXiv:1511.03677.
33. Mikolov, T.; Zweig, G. Context Dependent Recurrent Neural Network Language Model. In Proceedings of the 2012 IEEE Workshop on Spoken Language Technology, SLT 2012, Miami, FL, USA, 2–5 December 2012. [[CrossRef](#)]
34. Poulouse, A.; Han, D.S. UWB Indoor Localization Using Deep Learning LSTM Networks. *Appl. Sci.* **2020**, *10*, 6290. [[CrossRef](#)]
35. Sak, H.; Senior, A.; Beaufays, F. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. *arXiv* **2014**, arXiv:1402.1128.
36. Sundermeyer, M.; Ney, H.; Schlüter, R. From Feedforward to Recurrent LSTM Neural Networks for Language Modeling. *Audio Speech Lang. Process. IEEE/ACM Trans.* **2015**, *23*, 517–529. [[CrossRef](#)]
37. Sundermeyer, M.; Schlüter, R.; Ney, H. *LSTM Neural Networks for Language Modeling*; Science Department RWTH Aachen University: Aachen, Germany, 2012.
38. Li, Y.; Cao, H. Prediction for Tourism Flow based on LSTM Neural Network. *Procedia Comput. Sci.* **2018**, *129*, 277–283. [[CrossRef](#)]
39. Duan, Y.; Yisheng, L. Travel Time Prediction with LSTM Neural Network. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio De Janeiro, Brazil, 1–4 November 2016; pp. 1053–1058. [[CrossRef](#)]
40. Zhao, H.; Sun, S.; Jin, B. Sequential Fault Diagnosis Based on LSTM Neural Network. *IEEE Access* **2018**, *6*, 12929–12939. [[CrossRef](#)]
41. Kim, T.-Y.; Cho, S.-B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* **2019**, *182*, 72–81. [[CrossRef](#)]

42. Marino, D.L.; Amarasinghe, K.; Manic, M. Building energy load forecasting using Deep Neural Networks. In Proceedings of the IECON 2016—42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 23–26 October 2016; pp. 7046–7051.
43. Cacabelos, A.; Eguía, P.; Míguez, J.L.; Granada, E.; Arce, M.E. Calibrated simulation of a public library HVAC system with a ground-source heat pump and a radiant floor using TRNSYS and GenOpt. *Energy Build.* **2015**, *108*, 114–126. [[CrossRef](#)]
44. Fernández, M.; Eguía, P.; Granada, E.; Febrero, L. Sensitivity analysis of a vertical geothermal heat exchanger dynamic simulation: Calibration and error determination. *Geothermics* **2017**, *70*, 249–259. [[CrossRef](#)]
45. Sheela, K.G.; Deepa, S.N. Review on Methods to Fix Number of Hidden Neurons in Neural Networks. *Math. Probl. Eng.* **2013**, *2013*, 425740. [[CrossRef](#)]
46. Chen, K.; Yang, S.; Batur, C. Effect of multi-hidden-layer structure on performance of BP neural network: Probe. In Proceedings of the 2012 8th International Conference on Natural Computation, Chongqing, China, 29–31 May 2012; pp. 1–5.
47. Shin-ike, K. A two phase method for determining the number of neurons in the hidden layer of a 3-layer neural network. In Proceedings of the SICE Annual Conference 2010, Taipei, Taiwan, 18–21 August 2010; pp. 238–242.
48. Doukim, C.; Dargham, J.; Chekima, A. Finding the Number of hidden Neurons for an MLP Neural Network Using Coarse to Fine Search Technique. In Proceedings of the 10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010), Kuala Lumpur, Malaysia, 10–13 May 2010; pp. 606–609. [[CrossRef](#)]
49. Vujicic, T.; Matijević, T.; Ljucovic, J.; Balota, A.; Sevarac, Z. Comparative Analysis of Methods for Determining Number of Hidden Neurons in Artificial Neural Network. *Artif. Intell. Rev.* **2016**, *48*. [[CrossRef](#)]
50. Pradhan, B.; Lee, S. Landslide risk analysis using artificial neural network model focusing on different training sites. *Int. J. Phys. Sci.* **2009**, *4*, 1–15.
51. Nakama, T. Theoretical analysis of batch and on-line training for gradient descent learning in neural networks. *Neurocomputing* **2009**, *73*, 151–159. [[CrossRef](#)]
52. Li, M.; Soltanolkotabi, M.; Oymak, S. Gradient Descent with Early Stopping is Provably Robust to Label Noise for Overparameterized Neural Networks. In Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, Palermo, Italy, 3–5 June 2020; pp. 4313–4324.
53. Bilbao, I.; Bilbao, J. Overfitting problem and the over-training in the era of data: Particularly for Artificial Neural Networks. In Proceedings of the 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, 5–7 December 2017; pp. 173–177.
54. Eckle, K.; Schmidt-Hieber, J. A comparison of deep networks with ReLU activation function and linear spline-type methods. *Neural Netw.* **2019**, *110*, 232–242. [[CrossRef](#)]
55. Bock, S.; Weiß, M. A Proof of Local Convergence for the Adam Optimizer. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8.
56. Li, M.; Zhang, T.; Chen, Y.; Smola, A. Efficient mini-batch training for stochastic optimization. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014. [[CrossRef](#)]
57. Zaremba, W.; Sutskever, I.; Vinyals, O. Recurrent Neural Network Regularization. *arXiv* **2014**, arXiv:1409.2329.
58. Singh, G.; Panda, R. Daily Sediment Yield Modeling with Artificial Neural Network using 10-fold Cross Validation Method: A small agricultural watershed, Kapgari, India. *Int J. Earth Sci Eng* **2011**, *4*, 443–450.
59. Cacabelos, A.; Eguía, P.; Febrero Garrido, L.; Granada, E. Development of a new multi-stage building energy model calibration methodology and validation in a public library. *Energy Build.* **2017**, *146*. [[CrossRef](#)]
60. Hong, T.; Kim, C.-J.; Jaemin, J.; Kim, J.; Koo, C.; Jeong, K.; Lee, M. Framework for Approaching the Minimum CV(RMSE) using Energy Simulation and Optimization Tool. *Energy Procedia* **2016**, *88*, 265–270. [[CrossRef](#)]
61. Kuo, P.-H.; Huang, C.-J. A High Precision Artificial Neural Networks Model for Short-Term Energy Load Forecasting. *Energies* **2018**, *11*, 213. [[CrossRef](#)]
62. Pilgrim, M.; Willison, S. *Dive into Python 3*; Springer: Berlin, Germany, 2009; Volume 2.



© 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/>).