Applied LSTM Neural Network Time Series to Forecast Household Energy Consumption

Génesis Segura, est[1], José Guamán, est[1], Mónica Mite-León, MSc.[1], Vicente Macas-Espinosa, MSc.[2] and Julio Barzola-Monteses, MSc.[3,4]

[1]*University of Guayaquil, Facultad de Ciencias Matemáticas y Físicas, Ecuador,* monica.miteleon@ug.edu.ec
[2]*University of Guayaquil, Facultad de Arquitectura y Urbanismo, Ecuador, vicente.macase*@ug.edu.ec
[3]*Artificial Intelligence and Information Technology research group, University of Guayaquil, Facultad de Ciencias Matemáticas y Físicas, Ecuador, julio.barzolam@ug.edu.ec*
[4]*Department of Computer Science and Artificial Intelligence, Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación, University of Granada, Spain, jbarzola@correo.ugr.es*

Abstract. **- In Ecuador, energy consumption is accentuated in the residential sector due to population growth and other parameters, which leads to an increase in energy costs, greenhouse gas emissions and fossil fuel subsidies. Hence, there is a need to optimize and reduce energy consumption in buildings. One approach considered is predictive control systems, for which high accuracy consumption predictions are required. In this work we will apply supervised machine learning techniques using neural networks to forecast the energy consumption behavior of a family house; for this purpose, an experimental design is proposed using a dataset of almost four years of energy measurements, four different Long Short-Term Memory (LSTM) architectures are tested and about 200 models are run by varying hyperparameters. Metrics such as root mean square error (RMSE), mean absolute error (MAE) and mean absolute percent error (MAPE) are considered to compare and select the best LSTM model, being the best simple LSTM structure with vectorial output.**

Keywords**. - buildings, energy efficiency, forecasting, LSTM, time series.**

I. INTRODUCTION

Buildings play a dominant role in the clean energy transition, a report developed by the International Energy Agency shows growing trends in energy consumption and carbon emissions related to energy used in the buildings and housing sector across all types. During 2017, the construction and operation of buildings globally accounted for more than one-third (36%) of energy end-use. Carbon dioxide (greenhouse gas) emissions related to energy used by existing buildings, materials manufacturing and new building construction products accounted for about 40% of the global emissions for the same year [1].

In the Ecuadorian case, as reported in the National Energy Balance 2018 [2], national consumption in residential, commercial, service, and public administration buildings was approximately 21% of total consumption. The main source of energy used by buildings was electricity, which accounted for 57.6%, or 29.7% of the residential group and 27.9% of the commercial, services, and public administration group. Total energy demand in 2018 was 3.8% higher than in 2017, and the same behavior has been observed in the historical years, i.e., energy demand in Ecuador is increasing annually.

To mitigate the growth of consumption in buildings, there are some research approaches. One of them is related to energy efficiency and use of predictive control systems based on artificial neural networks (ANN) to forecast electricity consumption.

Predictive control systems considered in buildings require predicting with high accuracy the demand of energy consumption in the short, medium, and long term [3]–[7]. There are some techniques to forecast the electrical energy demand, such as traditional statistical models, thermal mathematical models involving differential equations, models considering machine learning, or deep learning models [8].

In this paper, we will apply supervised deep learning techniques of neural networks to forecast the energy consumption behavior of a family house, considering four different Long Short-Term Memory (LSTM) architectures.

The rest of the paper is composed of the following chapters. In chapter II, related literature on the energy theme in which ANN is applied are presented. In chapter III, the LSTM models used in this work are detailed. In session IV, the methodology is presented. In session V, the results are shown, and in session VI, they are analyzed. Finally, in session VII, the conclusions of the work are presented.

II. RELATED LITERATURE

Energy forecasting in recent years has been of great interest, so there is a good number of related literatures. A study conducted in [9] presents the prediction of hydroelectric production in Ecuador for 2015 using traditional statistical models such as ARIMA and ARIMAX.

The authors of the work in [10] proposed an energy efficiency strategy for residential users in the city of Quito, through the implementation of deep machine learning techniques ANN, established models of typical daily curves to indicate the most used appliances at certain times and propose an efficient economically and environmentally sustainable strategy.

There are other ANN applications in different types of buildings according to a study in [11] in which a calculation method was developed to simplify the energy consumption of buildings that applies to the Latin American socioeconomic context.

Another study in [8] proposed black-box approaches based on the use of ANN to forecast electricity consumption in an educational building, in which the potential and robustness of using LSTM applied to a limited amount of data was analyzed.

Research in [12] verified the effectiveness of an ANN for electrical distribution systems in terms of optimization for low voltage, based on the results obtained from the technical-economic analysis called benefit-cost ratio, which allowed

Digital Object Identifier (DOI):
http://dx.doi.org/10.18687/LACCEI2021.1.1.564
ISBN: 978-958-52071-8-9 ISSN: 2414-6390

them to verify the feasibility of the method proposed against an existing system or with an optimized one.

The authors in a study shown in [13] proposed the use of ANNs for electric power consumption forecasting. They determined the main influential variables in electric power consumption and analyzed some of the most used electric power consumption forecasting methods, choosing ANNs for being the most accurate. In addition, they verified the validity of the proposed method by comparing the forecasted results, resulting in a difference of 31.77 MWh. With this value, they were able to determine potential saving techniques, as well as the economic benefits.

There is also literature related to the application of ANNs [14] for the seismic pre-design of reinforced concrete buildings. For this, they used two ANN models. The first model corresponds to buildings of 8 to 12 floors, and the second, of 4 to 7 floors located in the soft soil of Mexico City. ANN models were used with "forward" feeding and with a feedback learning algorithm. The building designs obtained through ANN were compared with designs made conventionally, and they obtained maximum differences of the order of 15% at the level of structural elements, so the models were accepted.

Likewise, the authors in [15] considered LSTM models with two structures, one standard and the other, sequence to sequence. The dataset considered was of a residential client, and experiments were performed with one-minute and one-hour resolutions.

On the other hand, a study in [16] proposed to combine LSTM ANNs in hybrid structures considering real household energy consumption by the "UK-DALE" project with different time step sizes such as 5, 10, 20, and 30 minutes. Moreover, three error metrics were used.

III. LSTM Models

According to a study in [17], LSTM networks are used as connections to store feedback of input events in the form of activations which can be, short-term and long-term memory. Fig. 1 shows a simplified schematic of an LSTM memory unit.

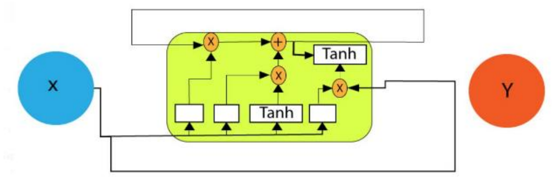

Fig. 1 LSTM schematic of model operation

1) LSTM model with vector output for multi-step forecasting with univariate input data

This model has a basic architecture that helps us to read the input data sequence and forecast with a vector output the energy consumption, this model, as well as others that we will see, have a three-dimensional structure, which waits on the input data that have [samples, time intervals, features].
Fig. 2 shows a schematic of this architecture.

2) LSTM encoder-decoder model for multi-step forecasting with univariate input data

The encoder-decoder model is a recurrent neural network designed to address sequential data difficulties. A schematic of this model is presented in Fig. 3.

Fig. 3 shows the architecture of the encoder-decoder model. This is composed of two parts: the encoder, which will be used to read the input sequence and encode it, and the decoder, which will read the encoded input sequence and make a one-step prediction for each element in the output sequence.

3) LSTM encoder-decoder model for multi-step forecasting with multivariate input data.

The multivariate model will be used with 3 input variables that have been previously selected by applying dimension reduction techniques.

4) CNN-LSTM encoder-decoder model for multi-step forecasting with univariate input data

Fig. 4 shows a scheme of this architecture. It considers a 1D convolutional layer, known as a temporal layer, with a convolutional neural network (CNN) in a decoder encoder structure as the CNN can also be used as an encoder. There are two convolutional layers, the first layer reads the data sequence and gives, as a result, a feature map. The second convolutional layer performs the same process to amplify the salient features. Following this process, the MaxPooling to group the elements is applied. The output data is passed through a *'Flatten'* to flatten the grouping of elements in a vector that later will be used as input data for the decoding process.

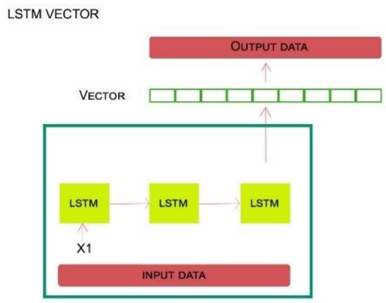

Fig. 2 LSTM model architecture with vector output

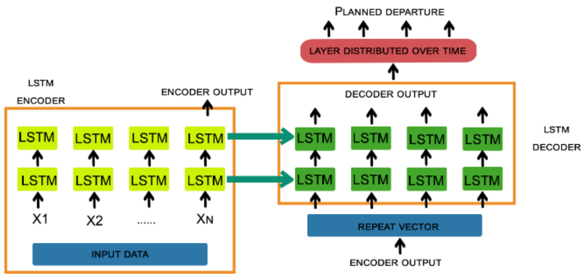

Fig. 3 Encoder-decoder arquitecture

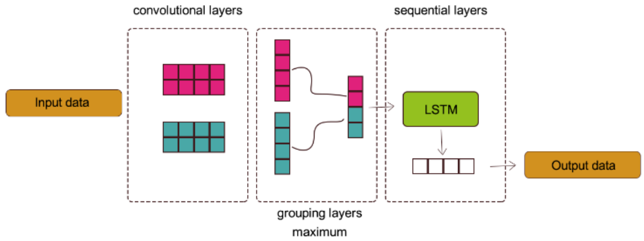

Fig. 4 Architecture of the CNN-LSTM model

IV. METHODOLOGY

A. DataSet Resolution

This paper used a dataset with energy consumption variables of a house located in France. The information is provided by the UCI repository [18]. The dataset is composed of measurements collected by a group of sensors, with a frequency of 1 minute for 47 months, i.e. from December 2006 to November 2010. For the experimentation, the dataset is resampled with a resolution of 15 minutes. The variables included in this dataset are shown in Table I.

The variables in Table I were preprocessed, resulting in the selection of three variables: total active power (TAP), total current intensity (TCI), and remaining energy consumption (G4). In this study, the prediction variable will be TAP (output), while the variables TAP, TCI, and G4 will be regressor variables (input) in each of the models to be analyzed.

TABLE I
DESCRIPTION OF THE DATASET VARIABLES

Index	Variables	Description
1	Date Time	Date and time
2	Global active power (TAP)	Household global active power consumed (KW)
3	Global reactive power (GRP)	Household global reactive power consumed (KW)
4	Voltage (VT)	Average voltage (V)
5	Global intensity (TCI)	Average current intensity (A)
6	Sub_metering_1 (G1)	Active energy for kitchen (Wh)
7	Sub_metering_2 (G2)	Active energy for laundry room (Wh)
8	Sub_metering_3 (G3)	Active energy for air conditioning systems Wh)
9	Sub_metering_4 (G4)	Household remaining energy consumption (Wh)

B. Past value horizon is defined (lookback)

The past values we have selected to define the hyperparameters are 24, 48, 72, and 96. They refer to 6 hours, 12 hours, 18 hours, and one day.

C. Prediction horizon is defined (Delay)

Our prediction horizon for the future data is 24, i.e., it refers to 6 hours in the future since our dataset was given a resolution of 15 minutes.

D. Dataset is defined for training and validation

The dataset contains 136885 records, which were divided into training data (109508) and data for model testing/validation (27377). Fig. 5 shows both datasets.

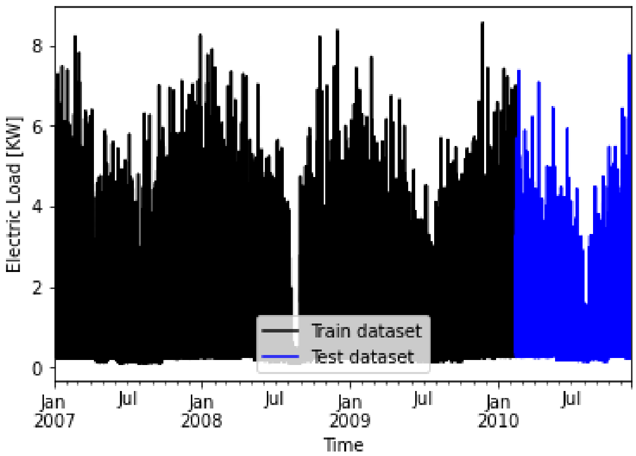

Fig. 5 Gráfico de Train y Test del dataset

E. Validation metrics

The validation metrics will be used to select the best model for each structure. Additionally, due to the random nature of ANNs, the best model is repeated 30 times to establish a statistical distribution of errors, which is represented by box plots.

The metrics used are root mean square error (RMSE), mean absolute error (MAE) and mean absolute percent error (MAPE). Their equations are shown below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \tag{1}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |y_t - \hat{y}_t| \tag{2}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} |\frac{y_t - \hat{y}_t}{y_t}| \tag{3}$$

where n is the number of samples, $\hat{y}_t$ is the estimation of y_t.

On the other hand, each structure will vary hyperparameters of the design which will result in different combinations or models. Table II shows the values of the hyperparameters that will be used during the experimentation for the first three LSTM structures presented in the previous sections.

TABLE II
HYPERPARAMETERS CONSIDERED FOR THE FIRST 3 MODELS

Hyperparameters	Values
n_input	24, 48. 72, 96
n_nodes	24, 48. 72, 96
n_epochs	25, 50
n_batch	32, 64, 128
act_hid	tanh
act_out	Linear

Table III shows the hyperparameter values for the fourth CNN-LSTM hybrid structure, which will produce 128 possible combinations to be tested during experimentation.

TABLE III
HYPERPARAMETERS CONSIDERED FOR THE CNN-LSTM MODEL

Hyperparameters	Value
n_input	24, 48. 72, 96
n_nodes	24, 48. 72, 96
n_epochs	25, 50
n_batch	32, 64
n_filters	32, 64
act_hid	tanh
act_out	Linear

V. RESULTS

A. LSTM con vector output

Table IV shows the ten best combinations based on the RMSE metric. The best fit is achieved with the model that has 24 input data, 96 nodes, 50 epochs, 64 batch, hidden layer 'tanh', and output data with an activation function.

The total execution time was 937.92 minutes, i.e., on average, each combination lasted 9.77 minutes.

B. Univariate LSTM decoder encoder

Table V shows the ten best combinations based on the RMSE metric. The best fit is achieved with the model that has 96 input data, 24 nodes, 25 epochs, 128 batches, and a hidden layer 'tanh'.

The total execution time was 28774.84 minutes, i.e., on average, each combination lasted 312.77 minutes.

TABLE IV
TOP 10 LSTM MODELS WITH VECTOR OUTPUT

Hyperparameters [n_input, n_nodes, n_epochs, n_batch, act_hid, act_out]	RMSE
[24, 96, 50, 64, 'tanh', 'linear']	1.0171
[24, 72, 50, 64, 'tanh', 'linear']	1.0176
[24, 48, 50, 32, 'tanh', 'linear']	1.0188
[24, 72, 50, 32, 'tanh', 'linear']	1.0191
[24, 96, 50, 32, 'tanh', 'linear']	1.0206
72, 96, 50, 64, 'tanh', 'linear']	1.0210
[72, 72, 50, 64, 'tanh', 'linear']	1.0221
[24, 96, 50, 128, 'tanh', 'linear']	1.0223
[24, 48, 50, 64, 'tanh', 'linear']	1.0231
[24, 96, 25, 32, 'tanh', 'linear']	1.0238

TABLE V
TOP 10 UNIVARIATE ENCODER-DECODER LSTM MODELS

Hyperparameters [n_input, n_nodes, n_epochs, n_batch, act_hid, act_out]	RMSE
[96, 24, 25, 128, 'tanh', 'linear']	1.0484
[96, 72, 25, 128, 'tanh', 'linear']	1.0492
[96, 24, 25, 64, 'tanh', 'linear']	1.0520
[72, 24, 25, 128, 'tanh', 'linear']	1.0695
[72, 48, 25, 128, 'tanh', 'linear']	1.0658
[48, 24, 25, 128, 'tanh', 'linear']	1.1162
[48, 48, 25, 128, 'tanh', 'linear']	1.1092
[24, 48, 25, 128, 'tanh', 'linear']	1.1506
[24, 24, 25, 32, 'tanh', 'linear']	1.1452
[24, 48, 50, 128, 'tanh', 'linear']	1.1439

TABLE VI
TOP 10 MULTIVARIABLE LSTM MULTIVARIABLE ENCODER-DECODER MODELS

Hyperparameters [n_input, n_nodes, n_epochs, n_batch, act_hid, act_out]	RMSE
[48, 24, 25, 128, 'tanh', 'linear']	2.3164
[72, 96, 50, 128, 'tanh', 'linear']	2.3235
[72, 24, 50, 128, 'tanh', 'linear']	2.3284
[48, 24, 50, 128, 'tanh', 'linear']	2.3383
[48, 72, 25, 32, 'tanh', 'linear']	2.3479
[96, 24, 50, 128, 'tanh', 'linear']	2.3492
[96, 48, 50, 32, 'tanh', 'linear']	2.3503
[24, 72, 50, 32, 'tanh', 'linear']	2.3939
[24, 72, 50, 64, 'tanh', 'linear']	2.3963
[24, 24, 25, 64, 'tanh', 'linear']	2.3982

TABLE VII
TOP 10 UNIVARIATE CNN-LSTM ENCODER-DECODER MODELS

Hyperparameters [n_input, n_nodes, n_epochs, n_batch, act_hid, act_out, n_filters]	RMSE
[48, 24, 25, 128, 'tanh', 'linear', 64]	1.4092
[48, 24, 25, 128, 'tanh', 'linear', 32]	1.4152
[48, 24, 25, 128, 'tanh', 'linear', 128]	1.4219
[48, 48, 25, 128, 'tanh', 'linear', 32]	1.4239
[24, 72, 25, 128, 'tanh', 'linear', 64]	1.4340
[24, 24, 25, 128, 'tanh', 'linear', 64]	1.4357
[96, 96, 25, 64, 'tanh', 'linear', 64]	1.4469
[96, 24, 25, 64, 'tanh', 'linear', 32]	1.4489
[72, 24, 25, 64, 'tanh', 'linear', 32]	1.4535
[72, 72, 25, 32, 'tanh', 'linear', 32]	1.4575

C. Multivariable LSTM encoder decoder

Table VI shows the ten best combinations based on the RMSE metric. The best fit is achieved with the model that has 48 input data, 24 nodes, 25 epochs, 128 batch, hidden layer 'tanh', and output data with an activation function.

The total execution time was 27631.28 minutes, i.e., on average, each combination lasted 300.34 minutes.

D. Univariable encoder decoder CNN-LSTM

Table VII shows the ten best combinations based on the RMSE metric. The best fit is achieved with the model that has 48 input data, 24 nodes, 25 epochs, 128 batch and filter 64.

The total execution time was 52962.56 minutes, i.e., on average, each combination lasted 413.77 minutes.

VI. DISCUSSION OF RESULTS

Fig. 6 contrasts the average execution time of each ANN structure analyzed with the RMSE error. We note that the vector LSTM has the best performance compared to the other LSTM models. Then, we have the sequential encoder-decoder univariate LSTM structure (S2S-U) which has a similarly low RMSE error to the LSTM-V, but its execution time exceeds 300 minutes, so it would not be as viable for prediction problems and associated computational cost.

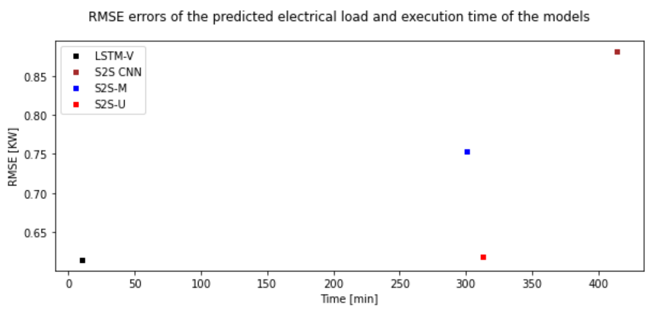

Fig. 6 Comparison of the computational time and RMSE errors of each analyzed structure

A. Comparative RMSE metric

Figure 7 shows the comparison of the RMSE errors of the four analyzed structures. The best model is LSTM-V since the average error is 0.614 KW and, it should be noted that 75% of the values fall below 0.615 KW, while in the S2S-U model, its average error is 0.619 KW, very close to LSTM-V. On the other hand, the S2S-M and S2S-CNN models have too high errors compared to the other models.

B. Comparative MAE metric

Figure 8 shows a comparison of MAE errors of each of the LSTM structures analyzed in this work. For the case of this metric, we note that the structure with the best error is the univariate CNN-LSTM decoder encoder represented in the acronym S2S-LSTM-CNN.

C. Comparative MAPE metric

Finally, for the MAPE error case, Figure 9 shows the comparison between the four structures studied. Where the MAPE error as well as the comparative RMSE, the best error is achieved with the LSTM structure with vector output.

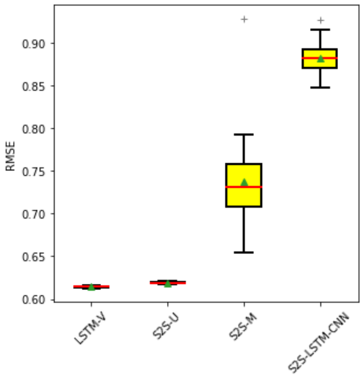

Fig. 7 Comparative boxplot of RMSE errors of the analyzed LSTM structures

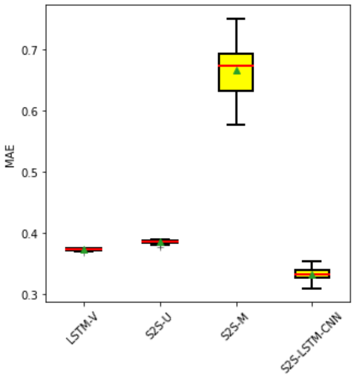

Fig. 8 Comparative boxplot of MAE errors of the analyzed LSTM structure

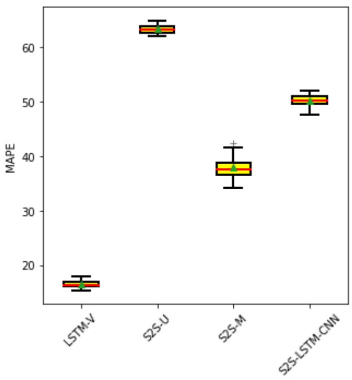

Fig. 9 Comparative box plot of MAPE errors of analyzed LSTM structures

VII. CONCLUSIONS

In the present work, LSTM recurrent neural networks were applied to a dataset with energy consumption data of a family house. This dataset has eight variables. After preprocessing the data, three variables were selected: total active power, as our prediction variable, total current intensity, and energy consumption (G4).

Based on the results obtained from the experimentation described in this manuscript, the best computational time is achieved with the LSTM structure with univariate vector output. This is mainly due to the low level of complexity that its structure has, so it is the fastest in concluding the adjustment and forecast that we wish to acquire.

According to the RMSE metrics, the one with the lowest error concerning the four structures is the LSTM model with univariate vector output, i.e., this model is the one that best forecasts a sequence of values and is ratified with the best processing time.

As for the MAE metrics, the best result was obtained with the univariate CNN-LSTM structure. However, in second place, is the LSTM model with vector output which has a lower execution time compared to this structure which is the highest in execution. Finally, the MAPE metric confirms that the best model is the LSTM with univariate vector output.

The LSTM model with univariate vector output for this particular case of energy consumption dataset of a family dwelling is the best fit to the forecast validation data. This model may be considered as a basis for the analysis of other residential energy studies in which other ANN structures and a larger variation of hyperparameters could be considered.

ACKNOWLEDGMENT

The authors kindly acknowledge the support from University of Guayaquil. This work was supported in part by the grant number FCI-015-2019.

REFERENCES

[1] V. Guillén Mena, F. Quesada Molina, M. López Catalán, D. Orellana Valdés, and A. Serrano, “Energetic efficiency in residential buildings,” *Estoa*, vol. 4, no. 7, pp. 59–67, 2015.

[2] IIGE, “Balance Energético Nacional 2018,” Quito, 2018.

[3] C. Finck, R. Li, and W. Zeiler, “Economic model predictive control for demand flexibility of a residential building,” *Energy*, vol. 176, pp. 365–379, 2019.

[4] H. H. Member, L. Chen, and E. Hu, “Model Predictive Control for Energy-Efficient Buildings : An Airport Terminal Building Study,” pp. 1025–1030, 2014.

[5] A. Afram and F. Janabi-shari, “Theory and applications of HVAC control systems e A review of model predictive control (MPC),” vol. 72, pp. 343–355, 2014.

[6] H. Huang, L. Chen, and E. Hu, “A neural network-based multi-zone modelling approach for predictive control system design in commercial buildings,” *Energy Build.*, vol. 97, pp. 86–97, 2015.

[7] J. Reynolds, Y. Rezgui, A. Kwan, and S. Piriou, “A zone-level , building energy optimisation combining an artificial neural network , a genetic algorithm , and model predictive control,” vol. 151, pp. 729–739, 2018.

[8] J. Barzola-Monteses, M. Espinoza-andaluz, M. Mite-León, and M. Flores-Morán, “Energy Consumption of a Building by using Long Short-Term Memory Network : A Forecasting Study,” in *39th International Conference of the Chilean Computer Science Society, SCCC 2020*, 2020, pp. 1–6.

[9] J. Barzola-Monteses, M. Mite-León, M. Espinoza-Andaluz, J. Gómez-Romero, and W. Fajardo, “Time Series Analysis for Predicting Hydroelectric Power Production: The Ecuador Case,” *Sustainability*, vol. 11, no. 23, p. 6539, Nov. 2019.

[10] D. Bravo Hidalgo and Y. Pérez Guerra, “Eficiencia energética en la climatización de edificaciones,” *Rev. Publicando*, vol. 3, no. 8, pp. 218–238, 2016.

[11] G. Reus Netto, “Metodología de cálculo simplificado para el consumo energético en acondicionamiento de edificios residenciales en clima templado,” p. 200, 2018.

[12] G. X. Carpio Suarez, P. A. Daga López, and P. D. Robles Lovato, “Aplicación de una red neuronal a un sistema eléctrico de distribución mediante el análisis de comportamiento de su carga en bajo voltaje,” p. 102, 2019.

[13] A. Santiesteban Velázquez, J. Osvaldo, N. González, D. R. Peña, and D. J. García, “Pronostico De Consumo De Energía Eléctrica Usando Redes Neuronales Artificiales,” *Tlatemoani*, no. 16, pp. 19–28, 2014.

[14] J. Bojórquez, D. Tolentino, S. E. Ruiz, and E. Bojorquez, “Diseño sísmico preliminar de edificios de concreto reforzado usando redes neuronales artificiales,” vol. 7, 2016.

[15] D. L. Marino, K. Amarasinghe, and M. Manic, “Building energy load forecasting using Deep Neural Networks,” *IECON Proc.*

(Industrial Electron. Conf., pp. 7046–7051, 2016.
[16] K. Yan, W. Li, Z. Ji, M. Qi, and Y. Du, “A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households,” *IEEE Access*, vol. 7, pp. 157633–157642, 2019.
[17] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[18] G. Hebrail and A. Barard, “Individual household electric power consumption Data Set,” *UCI Mach. Learn. Repos. Irvine, CA Univ. California, Sch. Inf. Comput. Sci.*, vol. 1, no. 1, pp. 1–3, 2012.