ISSN 2007-9737

IoT-Driven Energy Consumption Prediction in a Mexican Residence: A Case Study Utilizing Deep Learning with Attention Mechanism

Baldemar Aguirre-Fraire, Valeria Soto-Mendoza, Jessica Beltrán*

Centro de Investigación en Matemáticas Aplicadas,
Universidad Autónoma de Coahuila,
Mexico

{baldemaraguirre, vsoto, jessicabeltran}@uadec.edu.mx

Abstract. The energy consumption patterns exhibited by individuals carry significant implications for both the environment and energy production. Leveraging historical data, inferential models can foresee future consumption trends, assisting energy providers in planning adequate energy generation while encouraging shifts towards more environmentally sustainable practices. This, in turn, fosters a transition towards more eco-friendly behaviors. This paper outlines the development of a comprehensive system based on IoT that gathers real-world energy consumption data from a household in northeast Mexico, employs deep learning models for prediction, and incorporates a visualization tool to present energy demand. For the prediction, it was conducted a comparative analysis on three advanced deep learning models tailored for sequential data: LSTM, GRU, and Seq2Seq. Additionally, we explored the impact of enhancing each model with an Attention mechanism. Our findings consistently demonstrate that the incorporation of an Attention layer improves model performance, leading to a reduction in error metrics across the board. Specifically, we achieved an average Mean Absolute Percentage Error of 8.83% for daily predictions and 30.44% for hourly forecasts. These results underscore the efficacy of our selected models in accurately predicting energy consumption patterns, marking a notable stride towards informed and sustainable energy management.

Keywords. Energy consumption prediction, deep learning models, recurrent neural networks, attention mechanism.

1 Introduction

The International Energy Agency (IEA) reports that global growth in electricity demand was 2.2% in 2023 and it is projected to accelerate to an average of 3.4% from 2024 through 2026[1]. Households globally consumed 88 million terajoules (TJ) of energy for residential purposes in 2019, constituting nearly a quarter of the total final energy use worldwide[2]. This proportion demonstrated a consistent range of 19-21% across member countries of the Organisation for Economic Co-operation and Development (OECD) and has maintained relative stability over time. Regarding environmental impact, household energy consumption in 2019 accounted for 11% of global CO2 emissions and 14% of emissions from OECD countries. The notable contribution of households to total energy use, in contrast to total CO2 emissions, is influenced by the larger share of electricity in the residential sector's energy mix compared to other sectors [29].

In Mexico, an OECD country, electrical energy consumption has changed over time due to the evolution of energy policies and economic conditions[3]. Until the early 1980s, the majority of electricity in Mexico came from hydroelectric and thermal power plants [8]. In 2013, the

[1] https://www.iea.org/energy-system/electricity
[2] https://www.iea.org/data-and-statistics/data-product/world-energy-balances
[3] https://www.gob.mx/sener

Electricity Industry Law was enacted, allowing for a greater role of the private sector in the generation and sale of electrical energy[3]. This law also established goals to increase the participation of renewable energies in the national electrical system[4]. However, fossil sources, mainly natural gas (54.59%) and oil (13.53%), continued to dominate the national electrical system [33].

Since the 1980s, electrical energy consumption in Mexico has continued to grow [33]. This growth is driven by population increase, urbanization, and economic expansion[5]. Currently, the country faces significant challenges such as the need to increase renewable energy generation and improve the energy efficiency of the electrical system. Additionally, Mexico exhibits a high dependence on energy generation from fossil fuels, posing environmental risks such as greenhouse gas emissions that need to be mitigated [33]. In 2022, around 90% of electrical energy generation in Mexico relies on fossil fuels, contributing to global warming with negative effects on both the environment and public health. Although renewable energies are gaining popularity, their transition involves high initial costs and medium to long-term benefits [43].

Electricity, essential in almost all human activities, faces the challenge of storage, currently not feasible due to high costs and physical limitations. An alternative is generating energy based on real demand, requiring accurate predictive models to maintain a balance between supply and demand. Domestic electrical energy consumption represented 27.44% of the total consumption in Mexico during 2020[3].

Behavioral changes and active citizen participation can constitute a cost-effective strategy to reduce energy demand in the residential sector [21, 30, 39]. This participation can be achieved by providing useful information that aware users about their own behavioral patterns in energy consumption. This requirement implies having access to real-time data that reflects the actual and future behavior of electricity consumption within residences, facilitated by intelligent and automated monitoring systems, such as those implemented in smart homes using the IoT paradigm. Machine learning approaches can be incorporated into these smart systems to identify energy-saving opportunities, energy consumption prediction and awareness.

This work presents a holistic technical solution aimed to encourage citizen participation through the energy consumption behaviour awareness. This approach includes an acquisition system for collecting real-time energy consumption data from a household using IoT devices, storing historical data, predicting energy consumption with deep learning models and visualizing user's behaviours through a dashboard for a case study in a northeastern Mexican residence.

2 Related Work

Early attempts at time series prediction relied heavily on classical statistical methods like AutoRegressive Integrated Moving Average (ARIMA) and exponential smoothing methods. These traditional techniques were widely used due to their simplicity, interpretability, and well-established theoretical foundations [9, 20]. However, in recent years, there has been a notable shift towards utilizing machine learning models, particularly those based on deep learning, due to their ability to learn intricate patterns from historical data, resulting in improved predictive performance [24, 15, 37].

As shown in Table 1, there is no consensus regarding the variables to achieve predictions. Typically, the data analyzed includes: active power (KWh), electric current (A), and electric tension (V). In addition, there are works that also include climatic variables such as temperature (Tmp), humidity (Hum), and wind speed (WS) improving energy demand predictions [25]. In Table 1, we also summarize the sample rate, that refers to the frequency at which data samples are gathered and it vary among different studies. They include frequencies of data collected by seconds (Sec), by minutes (Min), by hours (Hr), and by days (Day).

There is also no consensus for the frequencies of data collection employed for making predictions. Moreover, depending on the frequency of data collection the prediction time interval in the future

[4] https://dof.gob.mx/
[5] https://www.inegi.org.mx/

ISSN 2007-9737

is established. These intervals also vary from minutes, hours, days and weeks (Week) and all these time periods are intervals in the short term.

We can observe also from Table 1 that many of the related work is based on the analysis of a single facility, although some works include data coming from more facilities. Besides, we summarized the metrics used to evaluate the prediction models. Due to the nature of the problem, which is a regression problem, the most popular metrics are the Mean Squared Error (MSE), the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE) and the R-squared (R2).

In Table 2 we display a comparison among various studies focused on the prediction of electrical consumption that use different classical and deep learning methods. In addition, we summarize the interval of time in which data was collected for each previous work, the number of observations in the dataset and the sample rate. Approaches with large amount of data [36, 23, 22] allow for a detailed analysis of electrical consumption fluctuations. However, this large amount of data can lead to long processing times and complicate data handling and storage, therefore prediction intervals can be affected and limited.

There is considerable diversity in the variables used, sample rate, short-term prediction interval, and performance metrics among different studies. When evaluating differences in the variables used, it is evident that each study selected variables based on their perceived relevance to the prediction model. Some studies focus more on electrical factors, such as current and voltage, while others incorporate environmental factors, such as temperature and humidity. In contrast, our approach incorporates all mentioned variables collected by minute for making short-term predictions evaluated with four of the five metrics found in the literature, as shown in Table 1. Furthermore, recent studies have witnessed a growing focus on crafting visualization tools for effectively communicating predictions and energy consumption behaviors to users. This trend is evident in works such as [34, 6, 17, 3].

It's important to mention that despite advances in machine learning methods for residential electricity consumption prediction, significant challenges exist [10, 4]. These challenges include the need for high-quality data, variability in consumer behavior and seasonal influences, and the need for models that can adapt to changes in consumption patterns over time [40, 27]. Therefore, machine learning and other data modeling techniques are promising tools to address these challenges, and their role in electricity consumption prediction is expected to continue growing in the future. In our approach, we analyze a dataset considering 468 days of data collected every minute, giving a total of 673,078 observations. The primary objective of our analysis is to predict household energy consumption, leveraging three distinct machine learning methods: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Sequence-to-Sequence (Seq2Seq), both in their standalone configurations and enhanced with an attention mechanism, as shown in Table 2.

3 Theoretical Background

3.1 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) represent a type of artificial neural networks tailored for sequential data processing [35, 42]. Different from the conventional feedforward neural networks, RNNs are characterized by their cyclic connections, which enable them to effectively handle sequential data [13]. This attribute makes them particularly adept at tasks like time series prediction, where the temporal order of data points holds crucial predictive information.

Figure 1 presents a schematic of a RNN architecture. This illustration demonstrates how each input element, denoted as x, is processed by an RNN cell to produce an output, $\hat{y}$. Furthermore, the cell generates an activation that is forwarded to the subsequent cell, enabling the RNN to incorporate information from past inputs. The cumulative loss L is calculated by summing the errors across all outputs. The lower part of the figure details the internal mechanics of each cell, which involve the weighting of input and previous

ISSN 2007-9737

Table 1. Comparison of literature review. Source: Own elaboration

Work	Number of facilities	Data type collected						Sample rate				Prediction interval				Performance metrics				
		KWh	A	V	Tmp	Hum	WS	Sec	Min	Hr	Day	Min	Hr	Day	Week	MSE	RMSE	MAE	MAPE	R2
[12]	1	✓			✓				✓					✓			✓	✓	✓	
[5]	1	✓			✓						✓			✓			✓	✓	✓	
[22]	1	✓	✓	✓					✓			✓	✓	✓	✓	✓	✓	✓	✓	
[36]	5	✓		✓				✓				✓					✓	✓	✓	
[23]	1	✓	✓	✓					✓			✓		✓		✓	✓	✓		
[28]	1	✓			✓	✓	✓			✓			✓				✓		✓	
[19]	1	✓							✓			✓					✓			✓
Our approach	**1**	**✓**	**✓**	**✓**	**✓**	**✓**	**✓**		**✓**				**✓**	**✓**		**✓**	**✓**	**✓**	**✓**	

Table 2. Comparison of dataset size and methods. Source: Own elaboration

Work	Duration of data collection in			Number of observations	Sample rate	Methods
	Days	Months	Years			
[12]	30	1	0	43,200	Minute	SARIMA-PSO-LSSVR SARIMA-MetaFA-LSSVR
[5]	990	33	2.8	990	Day	ANN's ensemble framework
[31]	730	24	2	17,520	Hour	deep RNN
[22]	1410	47	3.9	2,030,400	Minute	CNN-LSTM
[25]	720	24	2	17,280	Hour	LSTM-GA
[36]	786	26.2	2.2	4,034,880	Second	SWT-Transformer
[23]	1410	47	3.9	2,030,400	Minute	XGBoost, LSTM, CNN
[32]	365	12	1	8,760	Hour	RNN, LSTM GRU, TST
Our approach	**468**	**15.6**	**1.3**	**673,078**	**Minute**	**LSTM-Attention, GRU-Attention Seq2Seq-Attention**

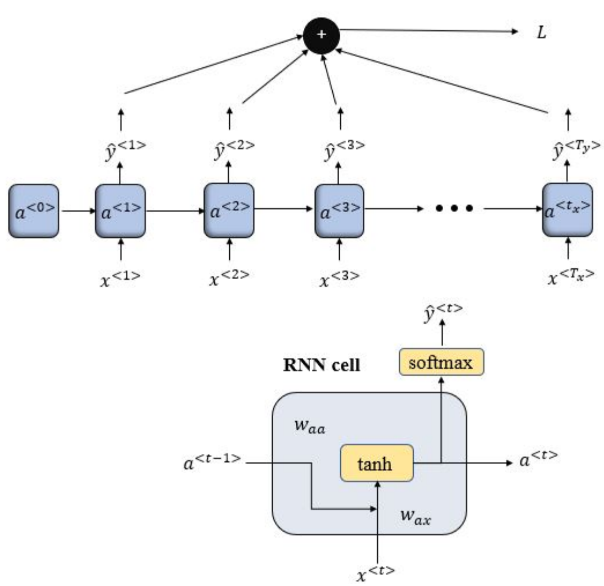

Fig. 1. Recurrent Neural Network representation. Source: Own elaboration, based on Andrew Ng notation from deeplearning.ai

data, along with the application of activation functions such as $tanh$ and $softmax$.

A prevalent issue in RNN architectures is the vanishing gradient problem. This occurs when dealing with long sequences, leading to the diminishing effect of multiplication operations, which in turn makes it challenging to sustain the influence of initial inputs on later outputs

3.2 Long Short-Term Memory

Long Short-Term Memory (LSTM) networks are an advanced form of RNN designed to capture long-term dependencies in sequential data, addressing the vanishing gradient problem that plagues early RNNs [18]. This problem, where gradients shrink as they backpropagate over time, is mitigated in LSTMs through the use of memory cells and three types of gates: input, output, and forget [14].

A memory cell stores information for long durations, while the gates control the flow of information, allowing the network to selectively remember or forget details. This structure enables LSTMs to effectively learn and maintain critical information over long sequences, making them ideal for analyzing complex temporal patterns and dependencies [16].

3.3 Gated Recurrent Unit

Gated Recurrent Unit (GRU) networks are another form of RNNs that also attend the vanishing gradient problem which allows them to handle long-term sequential data [11]. It can be seen as a simplified version of the LSTMs where the forget and input gates are combined into the update gate and the memory cell is combined with the output gate into the hidden state.

Likewise to LSTMs, these gate and state control the flow of information and how much should be retained and passed or forgotten. However, in comparison with LSTMs, the GRUs have fewer parameters and it can be faster to train.

3.4 Attention Mechanism

The attention mechanism has been a groundbreaking development in neural network research, significantly enhancing performance across various domains involving sequential data [7, 26, 41]. Its core principle is to selectively concentrate on pertinent segments of the input sequence based on their relative importance.

Incorporating the attention mechanism into architectures like LSTMs and GRUs can improve their efficacy in processing sequential data. This integration involves adding an attention layer to the existing framework, tasked with evaluating the significance of each input at specific time steps [7, 26].

By adopting this strategy, LSTM and GRU models gain the capability to model complex sequences with greater precision, thereby improving their overall modeling effectiveness [41].

3.5 Sequence to Sequence

The Sequence to Sequence (Seq2Seq) model is a neural network architecture specifically designed to handle tasks where both input and output are sequences of potentially varying lengths [38]. In the context of energy consumption prediction, this model can be adeptly used to process input sequences spanning multiple days to forecast energy usage for a single subsequent day or extend predictions across several days.

The Seq2Seq model comprises two primary components: the encoder and the decoder. The encoder processes the input sequence to produce a context vector, a condensed representation capturing the essence of the input [38]. This transformation is facilitated by a Recurrent Neural Network (RNN), which may employ Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) mechanisms for enhanced memory and processing capability.

On the flip side, the decoder takes this context vector and iteratively generates the output sequence. Through various layers, it deciphers the context to produce the final sequence prediction.

Moreover, the architecture can be augmented with an attention mechanism, enabling it to selectively concentrate on pertinent segments of the input sequence during different phases of the output generation [26]. This feature is especially beneficial in refining the model's focus and improving the accuracy and relevance of its predictions in complex scenarios, such as fluctuating energy demands.

3.6 Hyperparameters of the Models

Given that the models discussed in this section are based on neural network architectures, the tuning of the following hyperparameters is crucial:

— Number of Units: This refers to the quantity of neurons within each layer. While a higher number of neurons enables the network to discern more complex patterns in the data, it is essential to balance this capability to avoid overfitting the model.

— Batch Size: The batch size denotes the number of training examples utilized in one iteration of model training. Its size directly influences the speed of convergence of the model, impacting both training efficiency and stability.

— Number of Epochs: An epoch represents a full cycle of passing the entire training dataset through the learning algorithm. Increasing the number of epochs allows for more comprehensive adjustment and learning by the model but requires careful monitoring to prevent overtraining.

— Learning Rate: This scalar value dictates the step size at which the model's weights are updated during the optimization process. A well-chosen learning rate ensures efficient convergence to minimize the loss function.

— Dropout Rate: Defined as the probability of randomly excluding neurons during training, dropout acts as a regularization technique to mitigate overfitting by preventing dependency on any individual neuron.

The proper adjustment of these hyperparameters might improve the model's learning efficiency. A method for fine-tuning these hyperparameters is known as Random Search. This approach entails defining a range of potential values for each hyperparameter and subsequently testing random combinations of these values. After assessing a sufficient number of configurations, the combination that yields the most favorable outcomes is selected.

3.7 Error Metrics

To evaluate the performance of models used to predict energy consumption, a variety of metrics can be employed. In this paper, we employ four different metrics described next. In all cases, n is the number of observations, y_i is the actual value, and $\hat{y}$ is the predicted value.

Mean Squared Error (MSE). It measures the average squared difference between the predictions and actual values:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2. \tag{1}$$

Mean Absolute Error (MAE). It obtains the average absolute difference between the predicted and actual values:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \tag{2}$$

Root Mean Squared Error (RMSE). It is the square root of the Mean Squared Error, which gives a measure at the same magnitude of data of the average of the errors:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{3}$$

Mean Absolute Percentage Error (MAPE). It is the average percentage difference between predicted and actual values:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100. \tag{4}$$

4 Data Acquisition

4.1 Capture System

Efficient and continuous data collection is crucial in the digital age. In the context of domestic energy consumption and weather data, real-time data collection of this information can assist homeowners and utility companies in monitoring energy usage and making informed decisions.

Through a comprehensive literature review, it was identified that most previous studies rely on historical databases to develop predictive models of domestic electrical consumption.

However, these databases, while publicly available, are not recent and do not focus on case studies within Mexico.

ISSN 2007-9737

In response to this limitation, this work integrates various technologies for data collection. We used an energy smart meter[6], a Raspberry Pi 4[7], the Python programming language[8], a cloud-based MongoDB Atlas cluster[9] for database management, and the OpenWeather API[10] for real-time weather data acquisition are employed. Figure 2 depicts a diagram with the main components of the system and below we describe in detail each system component.

1. Smart Meter Configuration: the AT-Q-SY1 smart meter is an electrical measuring device that records energy consumption at regular time intervals (in this case, every minute) and transmits it digitally. These devices must be configured to transmit energy consumption data to another device.

2. Raspberry Pi 4: a Raspberry Pi 4 model B was used to receive readings from the smart meter, which was configured to host Docker containers. The container executes Python scripts to collect energy consumption and weather data each 60 seconds. The Python script was used to communicate with the smart meter quantifying home energy consumption. Additionally, a connection was made to the OpenWeather API to obtain weather variables for the house's geographical coordinates simultaneously. All collected data is structured in JSON format and sent to the cloud database.

3. Data Storage in MongoDB Atlas: MongoDB is a NoSQL database that can efficiently handle structured data in JSON format. Data collected from the Raspberry Pi 4 model B is sent to the MongoDB cluster in the cloud for storage. MongoDB also allows querying and analysis of stored data.

4. Dashboard for Data Visualization: We used the framework Dash[11] from the library Plotly[12] in Python to implement the dashboard. The data visualization tool was hosted in cloud-based Heroku platform[13].

The total cost for implementing the whole system is around $200 USD without considering hosting costs. Depending on the user needs, the dashboard can run locally to avoid cost derived from the visualization tool.

4.2 Dataset Description

The dataset used in this work was published in [2] with a detailed description in [1] and extended up to February 15, 2024. It comprises data from approximately one year and two months. The data source, variable, description, data type, and unit of the variables employed are outlined in Table 3. Variables selected for predicting household electricity consumption were chosen based on analysis of the correlation matrix, retaining those with a correlation coefficient exceeding 0.3 and omitting those with null or very weak correlation with the response variable (active power). The selected variables are in bold in Table 3.

5 Data Preprocessing

The data preprocessing phase involves several steps aimed at cleaning, transforming, and preparing the data for subsequent use in model training. Below is a detailed description of each preprocessing step:

5.1 Data Imputation

— Numerical Data: We used linear interpolation to fill in missing values in numerical columns.

— Categorical Data: We used forward filling to propagate the last known value in categorical columns.

[6] `https://at-ele.com/`
[7] `https://www.raspberrypi.com/products/raspberry-pi-4-model-b/`
[8] `https://www.python.org/`
[9] `https://www.mongodb.com/atlas/database`
[10] `https://openweathermap.org/`
[11] https://dash.plotly.com/
[12] https://plotly.com/
[13] https://www.heroku.com/

doi: 10.13053/CyS-29-4-4986

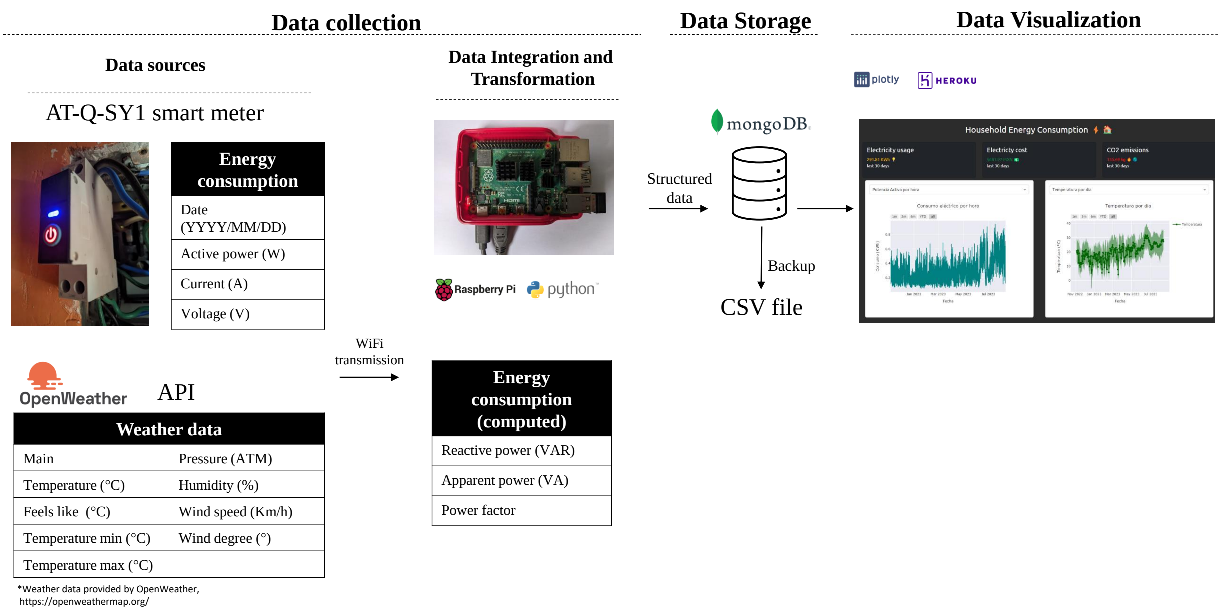

Fig. 2. Overall system diagram. Source: Own elaboration

Table 3. Detailed description of collected data. Source: Own elaboration

Source	Variable	Description	Data Type	Unit
Smart Meter	***date***	Timestamp of the sample	Date Time	YYYY-mm-dd HH:mm
	active_power	Active power (P)	Float	Watt
	current	Electric current (I)	Float	Ampere
	voltage	Electric tension (V)	Float	Volt
	reactive_power	Computed reactive power (Q)	Float	Volt-ampere reactive
	apparent_power	Computed reactive power (S)	Float	Volt-ampere
	power_factor	Computed power factor (PF)	Float	-
OpenWeather	***main***	General weather conditions	Categorical	-
	temp	Temperature	Float	℃
	feels_like	Thermal sensation	Float	℃
	temp_min	Minimum temperature	Float	℃
	temp_max	Maximum temperature	Float	℃
	pressure	Atmospheric pressure	Integer	ATM
	humidity	Humidity percentage	Integer	%
	speed	Wind speed	Float	Km/h
	deg	Wind direction	Integer	°

5.2 Data Aggregation

— By Hour: Original data is aggregated and summed hourly, resulting in a new time series with one-hour intervals.

— By Day: Original data is aggregated and summed daily, resulting in a new time series with one-day intervals.

5.3 One-Hot Encoding

One-hot encoding is applied to the “main” categorical column from the weather data. The values for this variable are: clear, clouds, drizzle, dust, fog, haze, mist, rain and thunderstorm.

This technique converts categorical variables into binary vectors that represent the presence or absence of each category and allow machine learning algorithms to deal with categorical data in a numerical fashion.

5.4 Data Splitting

The dataset is divided into training (80%), validation (10%), and test (10%) sets.

As shown in Figure 3, this split is performed while preserving the characteristics of time series, meaning that the temporal sequences remain intact in each set.

5.5 Series to Supervised Dataset

Time series are transformed into supervised dataset, consisting of an input and and target for each sample. This is done by defining an input and output length and creating sliding windows along the original time series.

In Figure 4 is depicted an example of using an input length of 5 and an output length of 1.

5.6 Data Scaling

To ensure uniformity among features, a crucial requirement in deep learning models for both convergence and optimal performance, we employed the Min-Max scaling method to rescale the data within the range of 0 to 1. This method, is depicted by the equation:

$$X_{\text{scaled}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}. \tag{5}$$

With this method, the maximum and minimum values for each feature are identified. The process involves subtracting the minimum value from each data point and dividing it by the range between the maximum and minimum values. Consequently, this transformation ensures that each feature spans from a minimum value of 0 to a maximum value of 1.

The determination of maximum and minimum values is exclusively performed using the training dataset to prevent any form of data leakage, maintaining the integrity of the scaling process by avoiding the influence of validation or test data on the computation of scaling parameters

6 Experimentation and Results

We experimented with data aggregated in the two manners explained in the preprocessing section: by hour and by day. For each type of time interval, we compared the models LSTM, GRU and Seq2Seq, all with and without the Attention mechanism. We adjusted the models with the training data and used the validation dataset to obtain the best hyperparameters.

For each model, we defined a search space that includes the number of units in the LSTM or GRU layers, the learning rate, batch size, and dropout rate. We used a combination of integers and discrete values to define these search spaces, thus ensuring a thorough yet efficient exploration. The defined values are listed in Table 4. In addition, we ensured to evaluate the default hyperparameter values. We used Random search and run multiple iterations of training and validation (40 and 20 for the day and hour interval respectively), we selected the set of hyperparameters that yielded the best

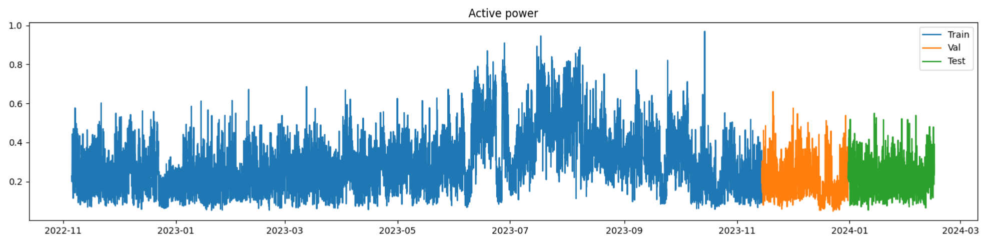

Fig. 3. Dataset split by hour into train, validation and test. Source: Own elaboration

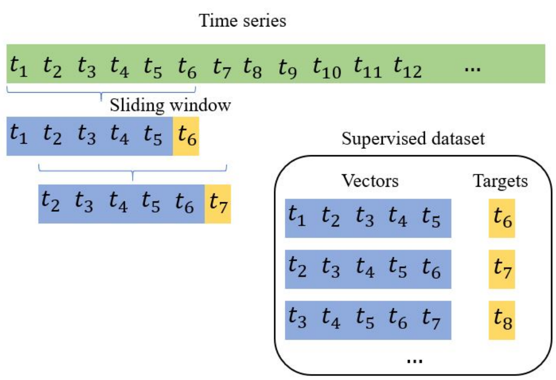

Fig. 4. Illustration of converting a time series to a supervised dataset. Source: Own elaboration

Table 4. Hyperparameter search space. Source: Own Elaboration

Hyperparameter	Values
n_units	128 - 513
learning_rate	0.001, 0.0005, 0.0001, 0.00001
dropout_rate	0.1, 0.2, 0.3, 0.4, 0.5
batch_size	1 - 168
epochs	50 - 201

performance in terms of a custom error metric, in this case, the root mean squared error (RMSE). This metric is relevant in our context as we are working with multivariate time series data and aim to minimize the discrepancy between the model predictions and the actual values.

Once the best hyperparameters for each model were identified, we proceeded to train the selected models using these hyperparameters on the test dataset. We performed 20 iterations of training and evaluation for each model and calculated the average error and standar deviation among the metrics, including MSE, RMSE, MAE, and MAPE. These metrics provide us with a holistic view of each model's performance on the test data that allow us to compare their relative effectiveness. Furthermore, they provide valuable insights into the models' generalization capability and their ability to make accurate predictions in real-world applications.

6.1 Model's Architectures

6.1.1 LSTM

The layers included in the LSTM are the following:

— A Bidirectional LSTM layer with n_{units} neurons that is defined in the hyperparameter space and the default activation functions (tanh and sigmoid).

— A dropout layer with the dropout rate defined in the hyperparameter space.

— An output TimeDistributed dense layer with one neuron with a linear activation function, aiming to predict the target values.

— Finally, we used the Adam optimizer and the custom loss function RMSE.

6.1.2 GRU

The layers included in the GRU architecture are as follows:

- A Bidirectional GRU layer with n_{units} neurons, which incorporates information from both past and future time steps.
- A dropout layer with the specified dropout rate to prevent overfitting.
- An output TimeDistributed dense layer with linear activation function, aiming to predict the target values.
- Similar to the LSTM architecture, Adam optimizer is utilized with the custom loss function, root mean squared error (RMSE).

6.1.3 Seq2Seq with LSTM

The architecture for the Seq2Seq model with LSTM consists of:

- A Bidirectional LSTM layer in the encoder part with n_{units} neurons, capturing bidirectional context information from the input sequences.
- A RepeatVector layer that repeats the output of the encoder for each time step in the decoder part.
- A unidirectional LSTM layer in the decoder part with n_{units} neurons, generating the output sequence based on the context vector from the encoder.
- A dropout layer to regularize the network and mitigate overfitting.
- An output TimeDistributed dense layer with linear activation function to produce the final output sequence.
- Adam optimizer is employed with the custom loss function RMSE for training.

6.1.4 Versions with Attention Mechanism

To all previous architectures (LSTM, GRU, and Seq2Seq) we created versions with Attention mechanism. In all architectures, we included the following layers:

- A Bidirectional encoder layer with n_{units} neurons, which incorporates information from both past and future time steps.
- An attention layer, where both the query and the value are the output of the recurrent layers (LSTM or GRU) of the model.
- A dropout layer with the specified dropout rate to prevent overfitting.
- The attention output and the output of the additional recurrent layer are concatenated together.
- A unidirectional decoder layer with n_{units} neurons, generating the output sequence based on the context vector from the encoder.
- Finally, a TimeDistributed dense layer with linear activation function is applied to predict the target values.

These versions with attention mechanism enhance the model's ability to focus on relevant parts of the input sequences while making predictions, potentially improving performance in tasks involving sequential data. It is important to note that in the analysis of daily household energy consumption, none of the mentioned architectures used bidirectional neural networks, primarily due to the limited availability of data for that time interval.

6.2 Results

The analysis presented in Table 5 reveals significant insights into the prediction errors across different time frames in the test set. A key observation is the disparity in prediction error when analyzed hourly versus daily. In this context, the MAPE metric emerges as a particularly useful tool for comparison. Its percentage-based nature offers a more intuitive understanding of errors across varying units, avoiding potential confusion

Table 5. Comparison of Average Model Performance by Hour and Day. Source: Own Elaboration

By Hour

Model	MSE	SD	RMSE	SD	MAE	SD	MAPE %	SD
LSTM	0.0066	$\pm 2.0e^{-4}$	0.0814	$\pm 1.2e^{-3}$	0.0622	$\pm 1.4e^{-3}$	33.03	± 1.98
LSTM-Attention	0.0064	$\pm 2.3e^{-4}$	0.0802	$\pm 1.4e^{-3}$	0.0613	$\pm 2.3e^{-3}$	32.46	± 2.89
GRU	0.0065	$\pm 1.2^{-4}$	0.0807	$\pm 7.9^{-4}$	0.0614	$\pm 1.1e^{-3}$	32.51	± 1.75
GRU-Attention	0.0065	$\pm 1.6e^{-4}$	0.0803	$\pm 1.0^{-3}$	0.0606	$\pm 1.5e^{-3}$	31.23	± 2.38
Seq2Seq	**0.0063**	$\pm 1.3e^{-4}$	**0.0792**	$\pm 8.7^{-4}$	**0.0597**	$\pm 6.7e^{-4}$	30.86	± 1.09
Seq2Seq-Attention	**0.0063**	$\pm 1.5e^{-4}$	0.0797	$\pm 9.8^{-4}$	0.0598	$\pm 9.2e^{-4}$	**30.44**	± 1.23

By Day

Model	MSE	SD	RMSE	SD	MAE	SD	MAPE %	SD
LSTM	0.3728	± 0.0160	0.6104	± 0.0130	0.4808	± 0.0110	9.07	± 0.13
LSTM-Attention	0.3750	± 0.0130	0.6123	± 0.0106	0.4769	± 0.0100	8.99	± 0.13
GRU	0.3894	± 0.0270	0.6237	± 0.0212	0.4883	± 0.0192	9.23	± 0.27
GRU-Attention	0.3878	± 0.0185	0.6226	± 0.0145	0.4832	± 0.0154	9.13	± 0.26
Seq2Seq	**0.3621**	± 0.0110	**0.6017**	± 0.0092	0.4732	± 0.0077	8.88	± 0.10
Seq2Seq-Attention	0.3662	± 0.0236	0.6048	± 0.0190	**0.4669**	± 0.0203	**8.83**	± 0.30

that might arise from other metrics. Notably, the MAPE for hourly predictions is significantly higher than for daily predictions, reflecting the greater variability and complexity encountered in the former scenario. However, the other metrics also reflects how models were adjusted and are also considered in the comparison.

An interesting aspect of the analysis is the performance of the Seq2Seq model, which outperforms its counterparts in both daily and hour predictions. In the case of daily predictions, the Attention layer does not significantly benefit the model as in The Seq2Seq model emerges as the top performer, offering the best outcomes and the flexibility to be easily adapted for predicting more than single future points.

Figure 5 presents a comparison between the actual values and those predicted by the Seq2Seq with Attention model on the test data, which achieved the best performance metrics.

Similarly, Figure 6 illustrates a comparison between the actual values and predictions made by the Seq2Seq model, also noted for its superior metrics.

These figures demonstrate that both models effectively capture the underlying trends in the data. Nonetheless, further efforts are necessary to address the challenge posed by atypical data points.

7 Dashboard of Energy Consumption

Another significant contribution of this research is the development of a dashboard, designed for intuitive visualization by end-users. The dashboard's primary objective is to enhance household occupants' awareness of their energy consumption patterns over recent hours or days, thereby promoting more conscious energy use behaviors.

Illustrated in Figure 7, the dashboard features a section for the indicators: energy usage, electricity costs, and CO2 emissions for the past 30 days. These indicators are calculated based on historical data, regional energy pricing, and CO2 emissions. The latter is computed according to the fact that Mexico emits 439gCO2 for each KW hour of electricity uses[14]. The aim of these indicators is to offer a concise overview of the current impact of household behaviors on both financial expenditures and environmental health. By highlighting the potential negative consequences, it seeks to motivate a positive change in consumption habits.

The graphs below within the dashboard serve as an informative role, detailing fluctuations in energy consumption and ambient temperature. Users have the flexibility to adjust the display settings to view data either on a daily or hourly basis. This section is particularly designed to provide insights

[14] https://www.climate-transparency.org/

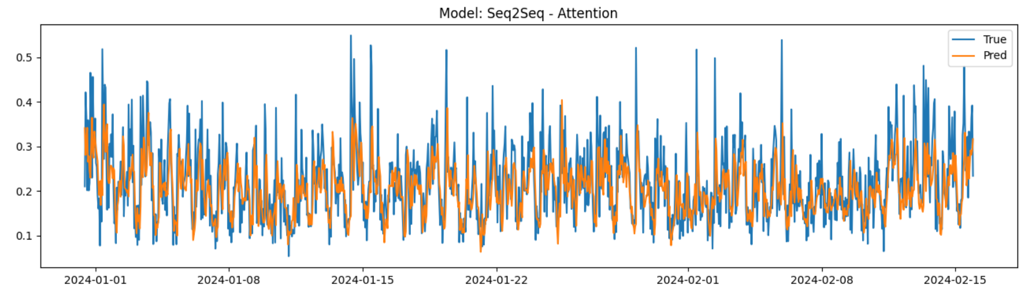

Fig. 5. Hourly predictions made for the test data with Seq2Seq-Attention. Source: Own elaboration

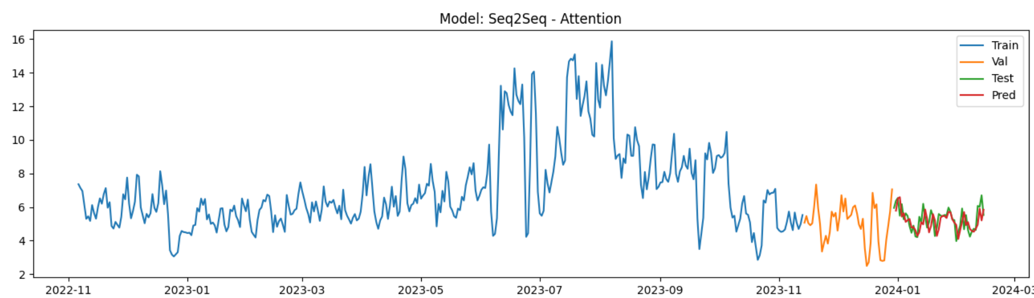

Fig. 6. Daily predictions made for the test data with Seq2Seq-Attention. Source: Own elaboration

into the relationship between energy use and external factors, enabling users to make informed decisions about their consumption patterns.

8 Conclusion and Future Work

In this study, we developed a comprehensive system designed to capture household energy consumption data, analyze time series information, generate predictions, and present key indicators. This system aims to foster positive changes in energy consumption behaviors among informed users. The dataset used in our research originates from a real-world household setting in the northeast region of Mexico, presenting unique challenges due to the variability inherent in human behavior.

Given the diversity in techniques and datasets used in related work, direct comparisons are challenging. However, our literature review revealed that hourly and daily data aggregation is among the most common approaches. Therefore, we adopted these aggregation levels and compared the performance of prevalent machine learning techniques for sequential data, incorporating an attention mechanism to enhance model performance.

We evaluated three deep learning models specialized in handling sequential data: LSTM, GRU, and Seq2Seq, applying metrics such as MSE , RMSE, MAE, and MAPE for performance comparison. Furthermore, we explored the impact of augmenting each model with an Attention layer, observing significant improvements as evidenced by reduced error metrics.

Looking ahead, our research will explore several promising directions. These include extending the prediction horizon to cover longer time intervals, either through increasing the Seq2Seq model's output size or employing a multistep forecasting strategy. Additionally, with access to a more extensive dataset, we plan to experiment with data aggregated on a weekly and monthly basis.

ISSN 2007-9737

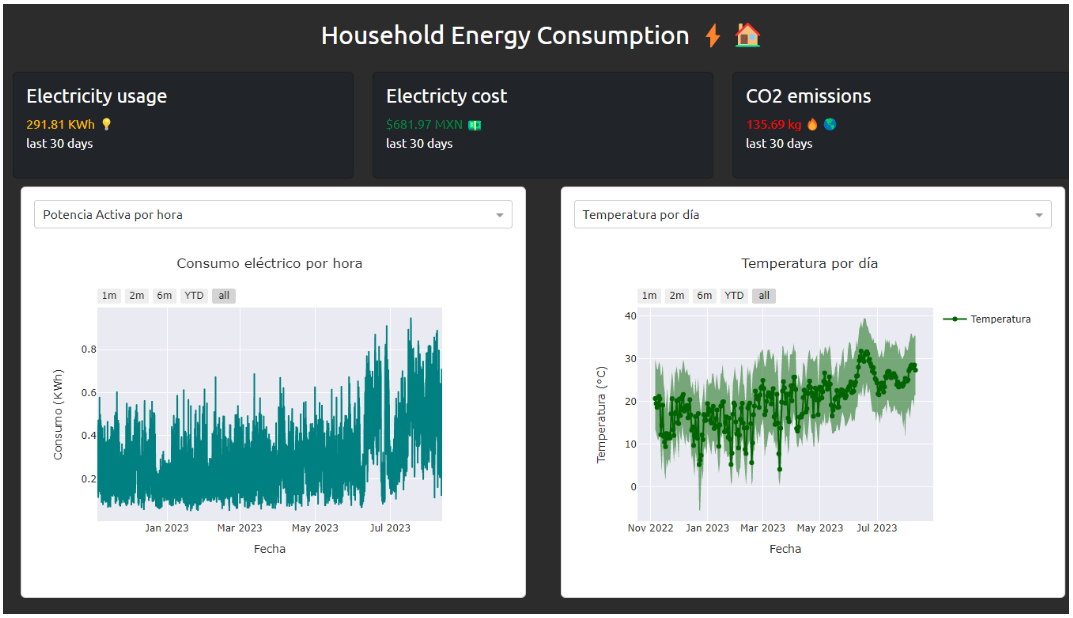

Fig. 7. Proposed dashboard to monitor energy consumption behavior. Source: Own elaboration

Another avenue of exploration is the evaluation of the Transformer model, which has demonstrated superior performance in other studies, although larger datasets are required for optimal performance.

Finally, a critical component of our future work will involve assessing the impact of the dashboard on influencing user behavior towards more sustainable energy consumption patterns.

Ethics Statement

The data acquisition system was installed in one house, informed consent was obtained from the owner and participant data has been fully anonymized. An ethical committee approved the data acquisition protocol number CIMA-CE-2022-P02 accordingly with the notice of privacy of the Universidad Autonoma de Coahuila (UAdeC).

Declaration of competing interest

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

Data Availability

The dataset up to January 5, 2024 is available here[15]. Full data will be made available on request.

Acknowledgments

We acknowledge to the National Council of Humanities, Science and Technology, CONAHCYT, for supporting a scholarship to the first author with the number 1243854.

[15] `https://data.mendeley.com/datasets/tvhygj8rgg/1`

References

1. **Aguirre-Fraire, B., Beltrán, J., Soto-Mendoza, V. (2024).** A comprehensive dataset integrating household energy consumption and weather conditions in a North-Eastern Mexican urban city. Data in Brief, Vol. 54, pp. 110452.
2. **Aguirre-Fraire, B., Beltrán, J., Soto, V. (2024).** Household energy consumption enriched with weather data in Northeast of Mexico.
3. **Al-Kababji, A., Alsalemi, A., Himeur, Y., Fernandez, R., Bensaali, F., Amira, A., Fetais, N. (2022).** Interactive visual study for residential energy consumption data. Journal of Cleaner Production, Vol. 366, pp. 132841.
4. **Almughram, O., Zafar, B., Slama, S. B. (2022).** Home energy management machine learning prediction algorithms: A review. 2nd International Conference on Industry 4.0 and Artificial Intelligence (ICIAI 2021), Atlantis Press, pp. 40–47.
5. **Alobaidi, M. H., Chebana, F., Meguid, M. A. (2018).** Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. Applied Energy, Vol. 212, pp. 997–1012.
6. **Alsalemi, A., Bensaali, F., Amira, A., Fetais, N., Sardianos, C., Varlamis, I. (2020).** Smart energy usage and visualization based on micro-moments. Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 2, Springer, pp. 557–566.
7. **Bahdanau, D., Cho, K., Bengio, Y. (2014).** Neural machine translation by jointly learning to align and translate. CoRR, Vol. abs/1409.0473.
8. **Bazán, G. (1981).** Energía en México: Perspectivas y problemas. Fondo de Cultura Económica.
9. **Box, G. E., Jenkins, G. M., Reinsel, G. C., Ljung, G. M. (2015).** Time series analysis: forecasting and control. Wiley.
10. **Chen, M., Yang, X. (2015).** Situations and challenges of household energy consumption in chinese small towns. Energy and Buildings, Vol. 107, pp. 155–162.
11. **Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y. (2014).** Learning phrase representations using rnn encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734.
12. **Chou, J.-S., Tran, D.-S. (2018).** Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. Energy, Vol. 165, pp. 709–726.
13. **Elman, J. L. (1990).** Finding structure in time. Cognitive science, Vol. 14, No. 2, pp. 179–211.
14. **Gers, F. A., Schmidhuber, J., Cummins, F. (2000).** Learning to forget: Continual prediction with lstm. Neural computation, Vol. 12, No. 10, pp. 2451–2471.
15. **Goodfellow, I., Bengio, Y., Courville, A. (2016).** Deep Learning. MIT press.
16. **Graves, A., Schmidhuber, J. (2005).** Framewise phoneme classification with bidirectional lstm and other neural network architectures. Neural Networks, Vol. 18, No. 5-6, pp. 602–610.
17. **Herrmann, M. R., Brumby, D. P., Cheng, L., Gilbert, X. M., Oreszczyn, T. (2021).** An empirical investigation of domestic energy data visualizations. International Journal of Human-Computer Studies, Vol. 152, pp. 102660.
18. **Hochreiter, S., Schmidhuber, J. (1997).** Long short-term memory. Neural computation, Vol. 9, No. 8, pp. 1735–1780.
19. **Huang, J., Kaewunruen, S. (2023).** Forecasting energy consumption of a public building using transformer and support vector regression. Energies, Vol. 16, No. 2.
20. **Hyndman, R. J., Athanasopoulos, G. (2018).** Forecasting: principles and practice. OTexts.
21. **Iea (2022).** Residential behaviour changes lead to a reduction in heating and cooling energy use by 2030 – analysis.
22. **Kim, T.-Y., Cho, S.-B. (2019).** Predicting residential energy consumption using cnn-lstm neural networks. Energy, Vol. 182, pp. 72–81.
23. **Kumar, N., Mishra, S., Baweja, T., Dubey, A., Dhiman, A. (2022).** Time series forecasting for electricity consumption using ml and dl algorithms. **Suhag, S., Mahanta, C., Mishra, S.**, editors, Control and Measurement Applications for Smart Grid, Springer Nature Singapore, Singapore, pp. 255–269.
24. **LeCun, Y., Bengio, Y., Hinton, G. (2015).** Deep learning. Nature, Vol. 521, No. 7553, pp. 436–444.
25. **Luo, X., Oyedele, L. O. (2021).** Forecasting building energy consumption: Adaptive long-short term memory neural networks driven by genetic algorithm. Advanced Engineering Informatics, Vol. 50, pp. 101357.

ISSN 2007-9737

26. **Luong, T., Pham, H., Manning, C. D. (2015).** Effective approaches to attention-based neural machine translation. **Màrquez, L., Callison-Burch, C., Su, J.**, editors, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Lisbon, Portugal, pp. 1412–1421.
27. **Ma, P., Cui, S., Chen, M., Zhou, S., Wang, K. (2023).** Review of family-level short-term load forecasting and its application in household energy management system. Energies, Vol. 16, No. 15, pp. 5809.
28. **Mu, Y., Wang, M., Zheng, X., Gao, H. (2023).** An improved lstm-seq2seq-based forecasting method for electricity load. Frontiers in Energy Research, Vol. 10.
29. **OECD (2023).** How green is household behaviour?: Sustainable choices in a time of interlocking crises. OECD iLibrary.
30. **Pratiwi, S. N., Sakita, Y., Kubota, T., Tanaka, H., Nishiiri, S., Takaguchi, H., Surahman, U. (2023).** Influences of occupant preferences for living environments on household energy consumption: A case study of indonesia. Energy and Buildings, Vol. 300, pp. 113636.
31. **Rahman, A., Srikumar, V., Smith, A. D. (2018).** Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Applied Energy, Vol. 212, pp. 372–385.
32. **Ramos, P. V. B., Villela, S. M., Silva, W. N., Dias, B. H. (2023).** Residential energy consumption forecasting using deep learning models. Applied Energy, Vol. 350, pp. 121705.
33. **Ritchie, H., Roser, M., Rosado, P. (2022).** Energy. Our World in Data. Https://ourworldindata.org/energy.
34. **Ruiz, L., Pegalajar, M., Molina-Solana, M., Guo, Y.-K. (2020).** A case study on understanding energy consumption through prediction and visualization (vimoen). Journal of Building Engineering, Vol. 30, pp. 101315.
35. **Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986).** Learning representations by back-propagating errors. Nature, Vol. 323, No. 6088, pp. 533–536.
36. **Saad Saoud, L., Al-Marzouqi, H., Hussein, R. (2022).** Household energy consumption prediction using the stationary wavelet transform and transformers. IEEE Access, Vol. 10, pp. 5171–5183.
37. **Shen, M., Lu, Y., Wei, K. H., Cui, Q. (2020).** Prediction of household electricity consumption and effectiveness of concerted intervention strategies based on occupant behaviour and personality traits. Renewable and Sustainable Energy Reviews, Vol. 127, pp. 109839.
38. **Sutskever, I., Vinyals, O., Le, Q. V. (2014).** Sequence to sequence learning with neural networks. Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14, MIT Press, Cambridge, MA, USA, pp. 3104–3112.
39. **Tete, K. H., Soro, Y., Sidibé, S., Jones, R. V. (2023).** Urban domestic electricity consumption in relation to households' lifestyles and energy behaviours in burkina faso: Findings from a large-scale, city-wide household survey. Energy and Buildings, Vol. 285, pp. 112914.
40. **Ushakova, A., Mikhaylov, S. J. (2020).** Big data to the rescue? challenges in analysing granular household electricity consumption in the united kingdom. Energy research & social science, Vol. 64, pp. 101428.
41. **Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017).** Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, pp. 6000–6010.
42. **Williams, R. J., Zipser, D. (1989).** A learning algorithm for continually running fully recurrent neural networks. Neural computation, Vol. 1, No. 2, pp. 270–280.
43. **World Wildlife Fund (2023).** Informe de la energía renovable. `https://www.wwf.org.mx/`. Accedido: 2023-11-26.

Article received on 21/05/2024; accepted on 13/02/2025.
**Corresponding author is Jessica Beltrán.*