IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 16 January 2024, accepted 8 February 2024, date of publication 13 February 2024, date of current version 29 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3365812

RESEARCH ARTICLE

Seq2Seq-LSTM With Attention for Electricity Load Forecasting in Brazil

WILLIAM GOUVÊA BURATTO[1], **RAFAEL NINNO MUNIZ**[2], **ADEMIR NIED**[1], **(Member, IEEE), AND GABRIEL VILLARRUBIA GONZÁLEZ**[3]

[1]Electrical Engineering Graduate Program, Department of Electrical Engineering, Santa Catarina State University (UDESC), Joinville 89219-710, Brazil
[2]Production Engineering Graduate Program, Department of Science and Technology, Federal Fluminense University (UFF), Rio das Ostras 28895-532, Brazil
[3]Expert Systems and Applications Laboratory, Faculty of Science, University of Salamanca, 37008 Salamanca, Spain

Corresponding author: William Gouvêa Buratto (william.buratto@edu.udesc.br)

This work was supported in part by Caracterización, Análisis e Intervención en la Prevención de Riesgos Laborales Enentornos de Trabajo Tradicionales Mediante la Aplicación de Tecnologías Disruptivas—Consejeria de Empleo Eindustria under Grant J125; in part by the Coordination for the Improvement of Higher Education Personnel (CAPES), Brazil, under the Doctoral Scholarship under Grant 88887.808258/2023-00; and in part by the Council for Scientific and Technological Development (CNPq) under Grant 310447/2021-6.

ABSTRACT Electricity load forecasting is important to planning the decision-making regarding the use of energy resources, in which the power system must be operated to guarantee the supply of electricity in the future at the lowest possible price. With the rise of the ability of forecasting based on deep learning, these approaches are promising in this context. Considering the attention mechanism promising to capture long-range dependencies, it is highly recommended for sequential data processing, where time series-related tasks stand out. Considering a sequence-to-sequence (Seq2Seq) time series data of the electricity load in Brazil, this paper proposes the use of the long short-term memory (LSTM) with the attention mechanism to perform the time series forecasting. The proposed Seq2Seq-LSTM with attention mechanism outperforms other well-established models. Having a mean absolute error equal to 0.3027 the proposed method is shown to be promising for field applications. The proposed method contributes to time series forecasting by implementing an attention mechanism to Seq2Seq data, therefore, more than one correlated signal can be used to perform the prediction enhancing its capacity when more data is available.

INDEX TERMS Electricity supply industry, energy measurement, forecast uncertainty, time series analysis.

I. INTRODUCTION

In Brazil, the energy dispatch is computed based on a stochastic optimization problem, where the goal is to ensure the availability of energy in the future at a lower cost as possible [1]. Since the power grids are connected in a national interconnected system [2], the generation and transmission are based on the demand of each region, and the management of the use of water (in hydroelectric power plants with dams) is based on this optimization problem [3]. Given this goal, electricity load forecasting is a challenging task.

Load forecasting is an important topic that several researchers have studied since it is relevant for decision-making regarding electrical power management [4], [5], [6], [7]. When renewable energies can not support the electricity demand, classical power sources are used based on fossil energy, which increases the cost of energy and its price [8]. Given that, energy forecasting can be used for energy planning and improved energy security [9].

Improving the characteristics of energy dispatch and costs with more efficiency from a prediction about the electricity costs, attributed together with the community demands, allows evaluating local load and storage showing behavior patterns about system marginal electricity prices [10]. These advantages can be applied to the contribution of planning new generators or alternative sources acquisitions verifying the correlation between variables and their impacts on the prices [11]. These approaches help to select strategies that consequently promote effective optimization of the grid system and reduction of costs of power plants [12].

Modeling and forecasting of electricity are essential tasks that collaborate to present features and advantages in a

The associate editor coordinating the review of this manuscript and approving it for publication was Reinaldo Tonkoski.

competitive market, showing components with potential improvement in the trading and risk management derived from seasonality, rapid spikes, and high volatility [13]. These maximize economic benefits and mitigate power market risks, consequently providing reliable operation and evaluating the profitability of storage in some cases reducing logistic uncertainties and most expensive predictive seasonal costs [14].

The period in the electricity load and prices forecasting are divided into three categories with particular advantages: short, medium, and long-term [15]. The first one is allocated when it desires real-time supply demand to prepare bidding strategies as a demand response program, in the case of the medium-term is collaborative to support import decisions and maintenance schedules, finally, the long-term contributes to power generation and storage planning and controls strategies as low load diesel helping to increase renewable energy insertion mainly in the isolated systems [16].

In Brazil, the operation of the power system depends on the electricity load of the power grid, based on a stochastic optimization problem the national operator system defines the requirements for the generation of each power plant to meet the need over time [17]. In this paper, a time series is considered to evaluate this variation over time, where predictions are made to help manage the power systems' decision-making. This application is especially promising in small grids (isolated systems) where the load variation is higher and classical methods may not perform well.

Decompose the load profile into features and components contributing to reducing the computational time process as the selection of these is a key factor bringing a significant influence in the prediction accuracy to understand the relative importance of each variable in different time horizons from aggregate analysis of opposite environmental conditions that can be used in multiple power levels [18], [19], [20].

The performance of load forecast is evaluated mainly by the accuracy obtained [21], however, other factors must be considered for the prediction model to reach success as the configuration facility and your applicability in the energy sectors, as well as the necessary costs to implement in the government or private power system, finishing the process when reaching the good balance between accuracy, repeatability, applicability, and easiness access [20].

Choosing the appropriate architecture to perform the time series analysis is challenging since deep learning structures may perform better when there are nonlinearities, however, they may need extra computational effort [22]. Combining shallow structures to have an ensemble learning model is also one possible solution [23]. Considering that the models used for time series have a specific nomenclature, in Table 1 the acronyms and symbols used in this paper are presented.

Given these challenges, in this paper, sequence-to-sequence (Seq2Seq) time series data is considered, and the long short-term memory (LSTM) with the attention mechanism is applied to forecast the electricity load in Brazil. The proposed Seq2Seq-LSTM with attention outperforms other well-established models. The contributions of this paper are:

TABLE 1. Abbreviations and symbols used.

Definition	Acronym
Convolutional neural networks	CNNs
Graphics processing unit	GPU
Long short-term memory	LSTM
Mean absolute error	MAE
Mean absolute percentage error	MAPE
Mean squared error	MSE
National Electricity System Operator	ONS
National Interconnected System	SIN
North	N
North East	NE
Random-access memory	RAM
Recurrent neural network	RNN
Root mean squared propagation	RMSprop
Sequence-to-sequence	Seq2Seq
South	S
Southeast/Midwest	SE
Stochastic gradient descent	SGD
Definition	**Symbol**
Candidate cell state	$\tilde{c}_t$
Cell state	c_t
Context vector	Θ
Forget gate	f_t
Hidden state	h_t
Input gate	i_t
Input sequence	x_t
Output gate	o_t
Predicted output	$\hat{y}_t$
Target output sequence	y_t
Time step	t

- A Seq2Seq time series data gives the advantage in the analysis considering that it is possible to evaluate more than one input series to perform the prediction.
- The attention mechanisms capture long-range dependencies which is an advantage for time series analysis since it enables the model to better understand the temporal dynamics within the data.
- A hybrid LSTM network that uses Seq2Seq data, and applies the attention mechanisms, outperforms well-established models.

The remainder of this paper is: In section II related works are discussed regarding the application of time series forecasting. In section III the proposed method is presented. In section IV the results are presented and discussed, and finally, in section VI a conclusion is drawn.

II. STUDY BACKGROUND

Several authors have studied time series forecasting models for diverse applications [24], such as emergency prevention [25], wind speed forecasting [26], fault prediction [27], load forecasting [28], power generation [29], where the use of deep learning structures are becoming popular [30].

According to Sopelsa Neto et al. [31], several models can be applied and it is difficult to define what method is more suitable for a specific application. In this case, the best strategy is to evaluate as many techniques as possible to set the best approach for the application. According to

Souza et al. [32], one major goal of this topic is the guarantee of energy supply.

The reliability of the power grid is important for consumers and it is the objective of the electricity utility suppliers [33]. In addition to load forecasting, the reliability of the electrical power network has been evaluated [34], especially using deep learning methods, such as convolutional neural networks (CNNs) [35], or even hybrid methods [36] that combine CNNs with other models [37]. Focused on time series the use of hybrid deep learning methods has shown to be promising [38].

According to Klaar et al. [39] energy has an important social impact and its prediction may help in the political decision-making scenario. In their research, an ensemble learning method was combined with seasonal decomposition approaches to forecast the energy price in Mexico. By optimizing the models using a hypertuning approach, it is possible to have a mean squared error of 3.37×10^{-9}, which was higher than other state-of-the-art models.

As presented by Seman et al. [40], the ensemble learning methods can outperform even deep learning for some applications. In their evaluation, it was proven that the use of filters to have a hybrid structure can be a promising approach when high frequencies don't need to be considered. Using a hybrid method they had a root mean square error 2.69 times lower than the original ensemble random subspace approach.

Using an aggregating prophet, in [41] the electricity spot prices were predicted considering a study case in Italy. As the authors mentioned the rise in the price caused by the conflict in Ukraine was difficult to predict since high prices like the ones applied in the period were never registered. Using a seasonal trend decomposition method they had an 18% mean absolute percentage error compared to the model with filter, showing to be a great strategy for time series prediction.

Ribeiro et al. [42], evaluated the electricity price of Brazil using a self-adaptive, heterogeneous, decomposed, and ensemble learning approach. They proved that by using a hybrid method it is possible to have a mean absolute percentage error 4.2% lower than other methods. In their method, several models are evaluated and combined to build a structure that outperforms single models.

Considering the use of hybrid methods where denoising filters are applied, in [43] the group method of data handling was combined with the Christiano-Fitzgerald random walk filter for fault prediction. With a root-mean-squared error of 3.44×10^{-12} their method outperformed the standard model (without the filter) and the LSTM.

A combination of different artificial intelligence methods of load electric forecasting can provide the highest accuracy avoiding over-fitting and benefit in the feature extractions resolution [44]. Hybrid models from the application of stochastic optimization as LSTM and wavelet [45] generally have the capability to process sequential data faster than simple algorithms enabling finding a set of suitable weights and recurrent neural networks [46] helping to calibrate the prediction time.

LSTM has advantages for long time series and combined with the attention mechanism can be a powerful time series forecasting [47]. The LSTM, considering a sequence-to-sequence problem, was used for the prediction of water levels in [48]. The authors proved that this method can be even better when combined with the wavelet transform. considering a mean squared error of 0.0020 their method was better than other well-established models.

Considering the load forecasting task, Yamasaki et al. [49] applied a hybrid method that combines prediction models to detrend functions for very short-term load forecasting. Their hybrid approach combines the gradient boosting regressor, extreme gradient boosting, k-nearest neighbors, and support vector regression through automated machine learning. Based on the combination of all models in a hybrid structure they had a root mean squared error of 1,931.8 MW.

Ma et al. [50] presented a complete review of short-term load forecasting presenting the applications focused on energy management systems. As they mentioned, several models can be applied for load forecasting such as autoregessive integrated moving average model, support vector regression, random forest, and gradient boosting regression. In addition to these models, they highlighted the LSTM, which is the one applied here. Other applications are found in the load forecasting review of Zhu et al. [51] and Wang et al. [52].

In [53] the transformer-based is considered for electrical load forecasting. They showed that based on this approach can be used for contextual data. Considering the transformer-based model, they had an accuracy increase of mean absolute percentage error of 2.57% in a 36 h load ahead forecasting.

III. METHODOLOGY

The LSTM is a recurrent neural network (RNN) architecture that is designed to overcome the vanishing gradient problem in traditional RNNs [54]. LSTMs are a popular choice in the field of deep learning for various sequential data processing tasks, such as speech recognition [55], classification [56], natural language processing [57], time series forecasting [58], and more.

LSTMs have proven to be effective in a wide range of applications due to their ability to model sequential data with long-term dependencies. They have been further extended and improved with variations such as attention mechanisms to handle more complex tasks [59]. The LSTM with the attention mechanisms is the method applied in this paper as a baseline for time series prediction.

The LSTM cell with the updates is given by:

$$
\begin{aligned}
f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
\tilde{c}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\
o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}
$$

where f_t is the forget gate, i_t is the input gate, $\tilde{c}_t$ is the candidate cell state, c_t is the cell state, o_t is the output gate, and h_t is the hidden state [60].

The LSTM encoder can be represented by:

$$h_t^{\text{enc}} = \text{LSTM}^{\text{enc}}(x_t, h_{t-1}^{\text{enc}}, c_{t-1}^{\text{enc}})$$
$$c_t^{\text{enc}} = \text{Cell}^{\text{enc}}(x_t, h_{t-1}^{\text{enc}}, c_{t-1}^{\text{enc}})$$

where x_t is the input sequence at time step t. The decoder is given by:

$$h_t^{\text{dec}} = \text{LSTM}^{\text{dec}}(y_t, h_{t-1}^{\text{dec}}, c_{t-1}^{\text{dec}})$$
$$c_t^{\text{dec}} = \text{Cell}^{\text{dec}}(y_t, h_{t-1}^{\text{dec}}, c_{t-1}^{\text{dec}})$$

where y_t is the target output sequence. When the attention mechanism is applied, it can be described as:

$$e_t^i = \text{Score}(h_t^{\text{dec}}, h_i^{\text{enc}})$$
$$\alpha_t^i = \text{Softmax}(e_t)$$
$$\Theta_t = \sum_{i=1}^{T} \alpha_t^i h_i^{\text{enc}}$$

where i is the time step, e_t^i and α_t^i are used in the attention mechanism to compute context vector Θ. Figure 1 presents the attention mechanism a($\Theta\mathbf{x}_i$, $\Theta\mathbf{x}_j$) employed in the model proposed by Veličković et al. [61].

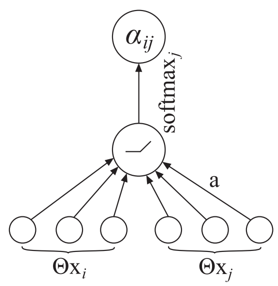


FIGURE 1. Attention mechanism.

Using the attention mechanism it is possible to dispense the recursion and convolutions entirely, resulting in models that require less time to train and are more parallelizable. As mentioned by Vaswani et al. [62], the self-attention can yield more interpretable models. Interpretable models are promising as their results make it clearer what is being accomplished by the model [63].

Extending the mechanism for employing multi-headed attention is beneficial as it helps to stabilize the self-attention learning process [61]. An example of multi-head attention, considering 3 heads, is present in Figure 2, in which different colors denote independent attention computations.

Finally, the output layer of the LSTM is:

$$o_t = \text{LSTM2Seq}(c_t, h_t^{\text{dec}}, y_t)$$
$$\hat{y}_t = \text{Softmax}(o_t)$$

where $\hat{y}_t$ is the predicted output probability distribution over the vocabulary for the time step [64]. The structure of this model is presented in Figure 3.

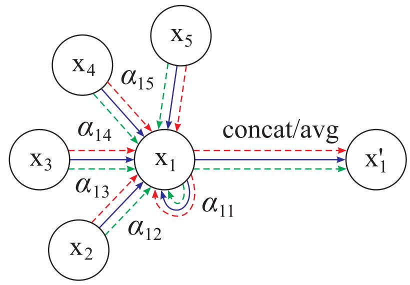


FIGURE 2. Multi-headed attention.

In the structure of the proposed method initially, the time series is loaded to an encoder where the data is normalized. In this step, the number of hidden neurons is the main parameter to specify the size of the network, this parameter is evaluated in the results section presented in this paper. Given this input, the model is trained considering an optimization method, that is also evaluated in this paper, until a maximum of epochs or the early stop criterion is satisfied.

In this paper, the use of 100 epochs is considered since the early stop works before the model reaches this number of epochs, therefore there is no overfitting and the model is properly trained. After training, the predictions are made and the predicted signal is compared to the observed signal. In this structure a sequence provides features to help the training of other sequences, therefore when there are correlated time series, the predictions of the seq2seq are enhanced by the proposed approach.

A. DATASET

In Brazil, there is a National Interconnected System (in Portuguese *Sistema Interligado Nacional* - SIN) that connects all the regions of the country. Still, there are isolated systems, however, they don't represent the majority of the systems, being used in remote places, mainly in the Amazon region where renewable energies are used [65]. The regions of Brazil are North (N), North East (NE), Southeast/Midwest (SE), and South (S). Loads of each region are presented in Figure 4.

The dataset is based on information received by the Supervision and Control System from the National Electricity System Operator (in Portuguese *Operador Nacional do Sistema Elétrico* - ONS), Brazil. Load data is by subsystem on a daily basis, measured in average megawatt (MWmed).[1]

Until February 2021, the dataset represents the load supplied by power plants dispatched and/or scheduled by the ONS; between March 2021 and April 2023, it is considered additionally the generation forecast for plants not dispatched by the ONS. From April 29, 2023, besides the data previously considered, the estimated value of micro and mini-distributed generation was incorporated, based on forecast meteorological information.

In Brazil, there are major differences in rainfall during the rainy season, and this variation directly impacts the plan of

[1]https://dados.ons.org.br/dataset/carga-energia

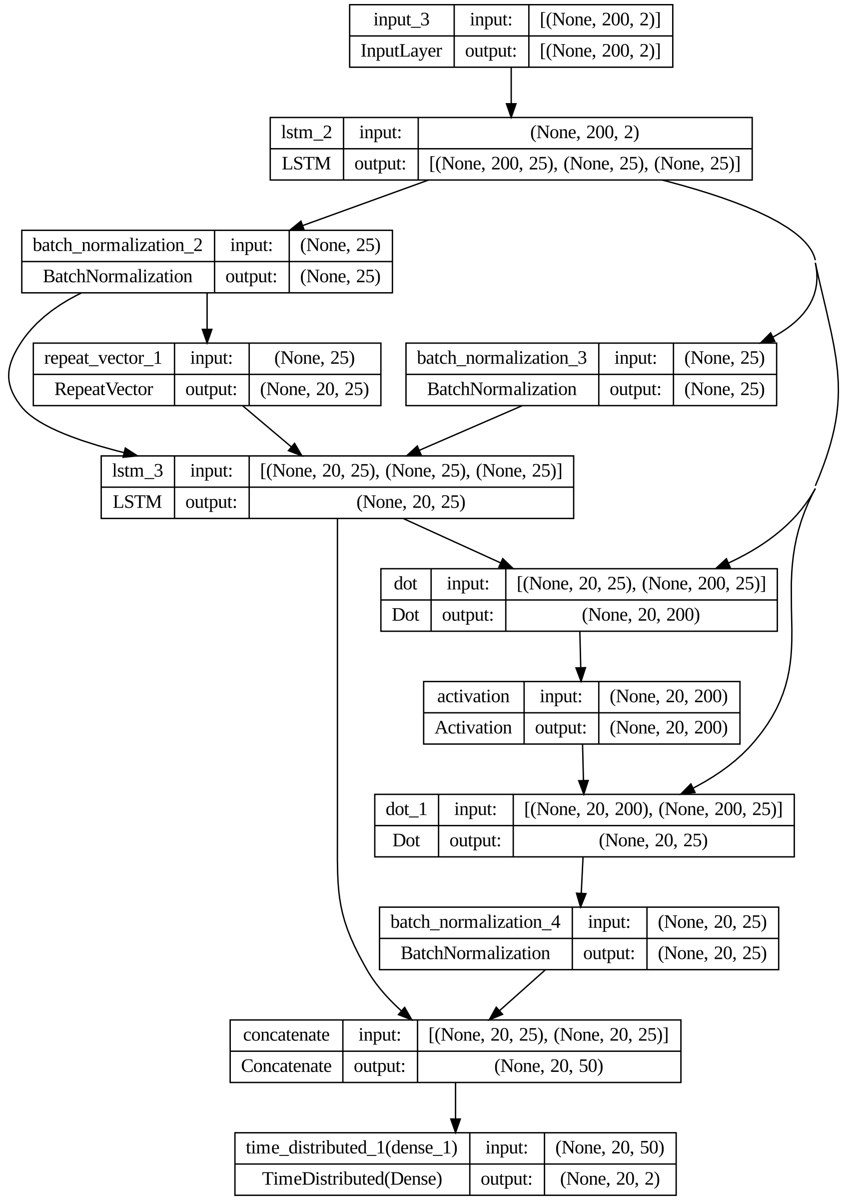


FIGURE 3. Structure of the proposed Seq2Seq-LSTM with attention.

the power dispatch over time [66]. Considering the variation of the rainy season (which happens once a year), in this paper a period of one year is evaluated to have the complete picture of the time series. The forecast is daily, as the electrical power system dispatch follows a daily plan.

The use of co-features can enhance the prediction model by having more information to train the model to deal with the variation of the system [67]. In this paper, the co-features of predictions are based on seq2seq data, where a sequence provides information on the prediction of other time series. The season variation is considered since the data evaluates a yearly period that has all variations of seasons. Weather conditions and social variables were not used since the dataset doesn't have this information and the focus is based on the evaluation of the trend of the time series.

The region SE was the major load of the SIN, which makes this region have a major impact on it. The result of summing all the loads of each region is the load of the country, this result is presented in Figure 5. This time series is considered for the study case presented in this paper.

B. EXPERIMENT SETUP

The algorithm proposed in this paper was written in Python, and the computations were performed in the Google Colab

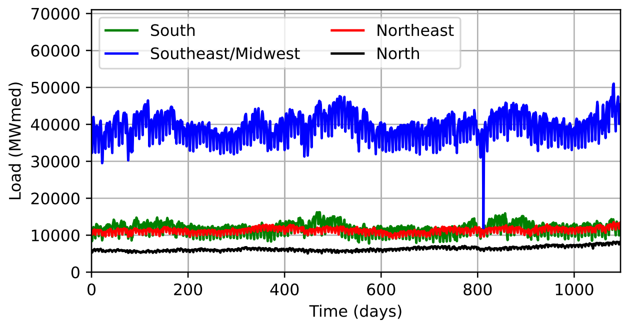


FIGURE 4. Load by region from October 2020 to October 2023, ONS, Brazil.

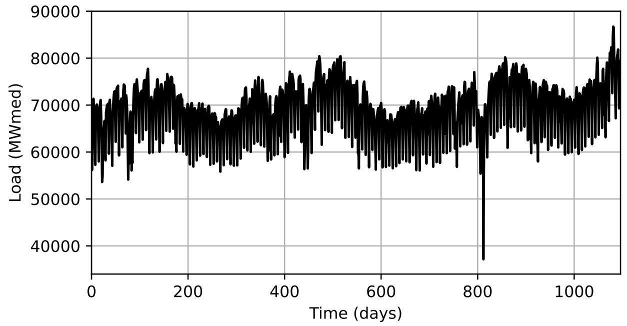


FIGURE 5. Electricity load of Brazil from October 2020 to October 2023, ONS, Brazil.

(back-end Google Compute Engine). The experiments were using a graphics processing unit (GPU) NVIDIA Tesla T4 (16 GB) and 12.7 GB of random-access memory (RAM).

For model assessment, the mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are considered, which are computed by:

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

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|, \tag{2}$$

$$\mathrm{MAPE}_{\%} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100, \tag{3}$$

where n are the number of samples, y is the observed value, and $\hat{y}$ is predicted value [68].

IV. RESULTS AND DISCUSSION

In this section, the results will be presented and discussions about them will be done. The initial evaluation is going to be using the standard LSTM considering a Seq2Seq time series data. After the first evaluations, the attention mechanism will be considered and a complete assessment will be performed.

A. PREPROCESSING

Given that the optimization problem evaluated by the ONS is based on a stochastic evaluation, two scenarios will be considered, one with more generation requests (x_2) and another with less demand (x_1). These scenarios compared to the original signal are presented in Figure 6.

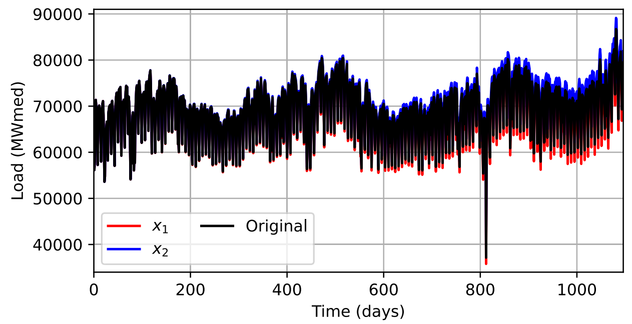


FIGURE 6. Original and possible two scenarios for the electricity load.

Given that the possible scenarios have a maximum and minimum trend, this variation is considered for the system to take account of the worst possible case. The maximum and minimum trend variations are presented in Figure 7. These trend variations are considering two possible scenarios given a stochastic problem. Training and test splitting are applied after normalizing the input data.

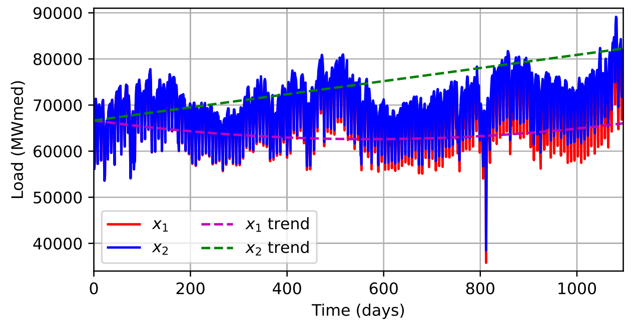


FIGURE 7. Maximum and minimum considered trend variations.

When the load follows the wort scenarios, with major variations, the results become even further than previously considered, these time series are presented in Figure 8.

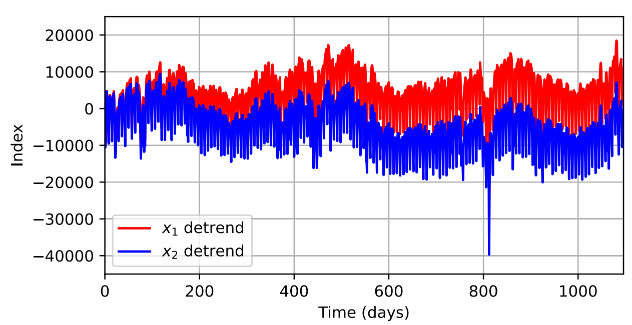


FIGURE 8. When load follows the trends scenarios.

Finally, normalizing the two-time series and splitting the data into train and test, we have the data that is used to evaluate the method presented in this paper. These time series are shown in Figure 9.

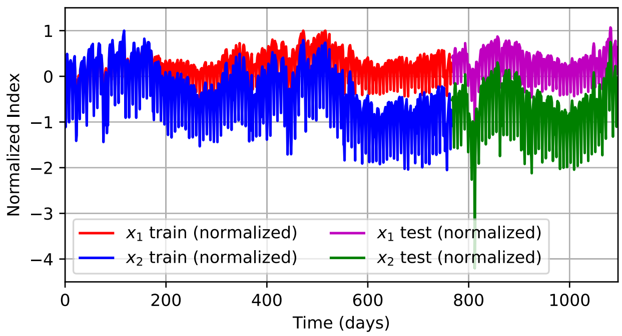


FIGURE 9. Normalized time series split in train and test datasets.

B. SEQ2SEQ-LSTM

To evaluate the Seq2Seq-LSTM model initially, 25 hidden neurons are considered and the Adam optimizer is applied. In the initial experiment, the Seq2Seq-LSTM model starts to have overfitting after 80 epochs as can be seen in Figure 10. To be sure that the model was properly trained a maximum of 100 epochs is defined and the early stop is applied (the training is stopped when there are no more improvements).

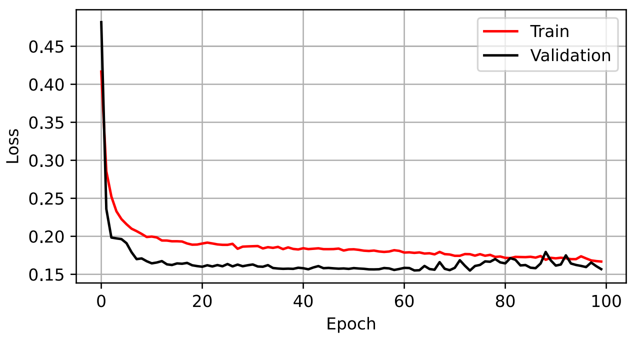


FIGURE 10. Train and validation loss over the epochs.

To have a complete evaluation of the Seq2Seq-LSTM model the number of neurons and the optimizer are evaluated. The Adam, root mean squared propagation (RMSprop), and stochastic gradient descent (SGD) optimizer are considered. The results are presented in Table 2.

Since the models have their best results considering different measures, the MAE is used for a global evaluation. Using the Adam optimizer, the best MAE was based on 5 neurons, in this configuration, the MSE was 0.1638, the MAE was 0.3129, and the MAPE was 116.01.

Based on the RMSprop most MAE results were a bit higher, considering the SGD better results were found. The best MAE was 0.3074 using 5 neurons with the SGD optimizer. Most of the best MSEs were found using the Adam, however, as the MAE is used for the main evaluation the SGD is the optimizer that returns the lower error given this metric.

TABLE 2. Results of Seq2Seq-LSTM model.

Optimizer	Neurons	MSE	MAE	MAPE	Time (s)
Adam	5	0.1638	0.3129	116.01	404.58
	10	0.1605	0.3257	113.54	449.98
	25	0.1568	0.3165	116.97	457.86
	50	0.1534	0.3158	130.97	450.97
	100	0.1559	0.3238	116.04	323.04
RMSprop	5	0.1630	0.3321	115.39	449.16
	10	0.1663	0.3401	112.61	405.80
	25	0.1584	0.3270	119.43	390.40
	50	0.1685	0.3413	120.77	450.30
	100	0.1687	0.3162	129.75	450.24
SGD	5	0.1769	0.3074	121.45	451.39
	10	0.1733	0.3151	108.43	450.93
	25	0.1634	0.3096	139.97	399.36
	50	0.1623	0.3214	102.98	451.66
	100	0.1648	0.3245	101.67	407.39

The distribution of the MAE results over the time series is presented in Figure 11. As can be observed the shutdown that happens in the power system at the beginning of the time series evaluated, which results in low values in the generation, didn't affect the prediction results. This indicates that the proposed model could be applied to identify possible shutdowns been an extra advantage of this approach.

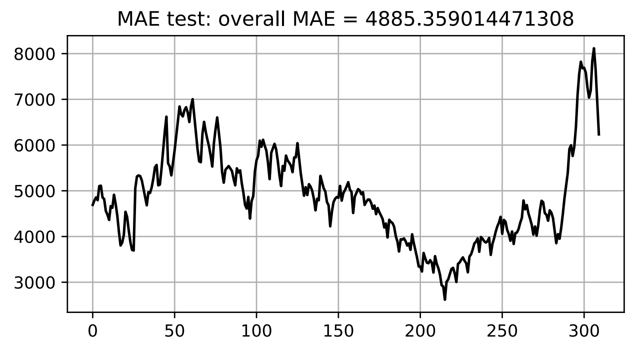


FIGURE 11. MAE results for the test dataset.

C. SEQ2SEQ-LSTM WITH ATTENTION

The results of prediction errors using the attention mechanism are presented in Table 3, these setups are based on the equivalent variations of the hyperparameters that were evaluated considering the standard Seq2Seq-LSTM.

The Seq2Seq-LSTM with Attention had the best MAE using the SGD with 25 neurons, by using this configuration an MSE of 0.1722, MAE of 0.3027, and MAPE of 140.72 was obtained. Interesting results also emerged from this optimizer using 10 neurons, in which the MAPE was 93.97, the lowest result so far, however, when five neurons were evaluated, this optimizer resulted in the worst MAPE, showing that this measure is not the best on to define the setup configuration of these models.

Using the Adam optimizer the second-best MAE was found, this is also valid to state that the Seq2Seq-LSTM using the attention mechanism was better than the original model considering this metric. If the MSE was considered the

TABLE 3. Results of Seq2Seq-LSTM with attention model.

Optimizer	Neurons	MSE	MAE	MAPE	Time (s)
Adam	5	0.2276	0.3732	177.75	452.84
	10	0.1604	0.3038	142.69	451.82
	25	0.1888	0.3096	161.00	392.09
	50	0.1978	0.3196	139.64	298.72
	100	0.2500	0.3430	145.31	330.28
RMSprop	5	0.1549	0.3086	138.07	425.82
	10	0.2047	0.3126	138.99	381.41
	25	0.2106	0.3192	156.34	377.49
	50	0.1947	0.3103	131.89	449.55
	100	0.1660	0.3056	116.88	390.68
SGD	5	0.2081	0.3356	227.07	450.15
	10	0.1697	0.3279	93.97	425.67
	25	0.1722	0.3027	140.72	453.42
	50	0.2047	0.3108	174.84	450.10
	100	0.2000	0.3089	150.83	412.48

attention mechanism may not be interesting since it creates more complexity without major improvements in this case.

The higher architecture complexity doesn't reflect in a longer time to compute the model using the attention mechanism, this result can be related to the early stop, where a model with this function may converge faster even using extra parameters.

The results of the predictions considering the training and testing time series data are presented in Figure 12. As observed the predictions have fewer nonlinearities than the original signal, this result is fine since the main goal of the evaluation is to have a trend of variation.

The best MSE result using the proposed method (with attention mechanism) was equal to 0.1549 using the RMSprop optimizer and 5 neurons, in comparison to the original LSTM had an MSE of 0.1630. Based on this result with an MSE 5.23% lower, the Seq2Seq-LSTM with Attention outperforms the original LSTM.

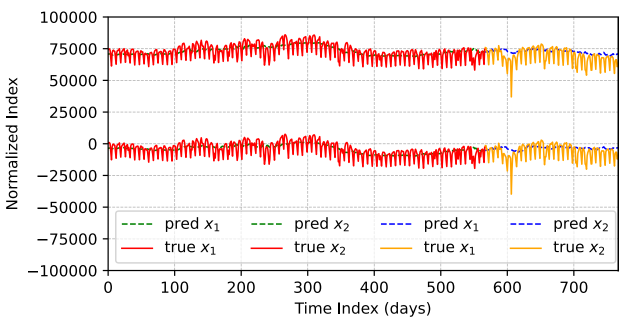


FIGURE 12. Prediction considering the training and testing.

D. EVALUATION BY REGION

Using the Seq2Seq-LSTM with Attention model an extra evaluation is presented in Table 4. Brazilian regions have very different structures of power sources and load features, this characteristic makes the power grid unique, which may be reflected in the evaluation of time series models.

Comparing the results of different regions it is possible to observe that in the North region, the error of the prediction was lower, this happens because in this region there was less

TABLE 4. Evaluation performance of the model considering different regions of Brazil.

Region	MSE	MAE	MAPE	Time (s)
North (N)	0.2544	0.3655	43.10	450.39
North East (NE)	0.4481	0.5059	2,943.79	451.10
Southeast/Midwest (SE)	0.1443	0.2741	335.93	420.51
South (S)	0.1667	0.2921	94.11	450.34

variation in the load given the features of the SIN, observing that there is more wind energy generation in this region and less industrialization.

V. FINAL REMARKS AND CONCLUSION

The use of Seq2Seq-LSTM with attention mechanism to electricity load forecasting in Brazil has proven to be a significant advancement. This approach combines the strengths of the sequence-to-sequence approach, LSTM networks, and attention mechanisms to tackle the complex and dynamic nature of electricity load data.

The proposed Seq2Seq-LSTM with attention model excels at capturing the temporal dependencies and patterns in electricity load data. By considering historical loads and their relationships over time, the model provides more accurate and context-aware predictions.

Compared to classic forecasting methods, the Seq2Seq-LSTM with attention has demonstrated superior results with an MAE equal to 0.3027, this model outperforms other approaches. Additionally, the proposed method can handle non-linear load movements, sharp fluctuations, and abrupt changes, making it a valuable tool for energy evaluation.

Future work can be performed considering hypertuning the proposed model, thus it would not be necessary to evaluate each parameter individually to have the best structure. The optimization of hyperparameters tunes the structure to have the best performance considering the used data, and it is a promising approach to have the best possible model.

REFERENCES

[1] B. Colonetti, E. Finardi, P. Larroyd, and F. Beltrán, "A novel cooperative multi-search benders decomposition for solving the hydrothermal unit-commitment problem," *Int. J. Electr. Power Energy Syst.*, vol. 134, Jan. 2022, Art. no. 107390.

[2] K. G. Di Santo, E. Kanashiro, S. G. Di Santo, and M. A. Saidel, "A review on smart grids and experiences in Brazil," *Renew. Sustain. Energy Rev.*, vol. 52, pp. 1072–1082, Dec. 2015.

[3] H. Gomes e Souza, E. C. Finardi, B. H. Brito, and F. Y. K. Takigawa, "Partitioning approach based on convex hull and multiple choice for solving hydro unit-commitment problems," *Electric Power Syst. Res.*, vol. 211, Oct. 2022, Art. no. 108285.

[4] L. B. S. Morais, G. Aquila, V. A. D. de Faria, L. M. M. Lima, J. W. M. Lima, and A. R. de Queiroz, "Short-term load forecasting using neural networks and global climate models: An application to a large-scale electrical power system," *Appl. Energy*, vol. 348, Oct. 2023, Art. no. 121439.

[5] G. Tri. Ribeiro, J. G. Sauer, N. Fraccanabbia, V. C. Mariani, and L. D. S. Coelho, "Bayesian optimized echo state network applied to short-term load forecasting," *Energies*, vol. 13, no. 9, p. 2390, May 2020.

[6] F. L. C. da Silva, K. da Costa, P. C. Rodrigues, R. Salas, and J. L. López-Gonzales, "Statistical and artificial neural networks models for electricity consumption forecasting in the Brazilian industrial sector," *Energies*, vol. 15, no. 2, p. 588, Jan. 2022.

[7] A. Ozcan, C. Catal, and A. Kasif, "Energy load forecasting using a dual-stage attention-based recurrent neural network," *Sensors*, vol. 21, no. 21, p. 7115, Oct. 2021.

[8] A. Meng, P. Wang, G. Zhai, C. Zeng, S. Chen, X. Yang, and H. Yin, "Electricity price forecasting with high penetration of renewable energy using attention-based LSTM network trained by crisscross optimization," *Energy*, vol. 254, Sep. 2022, Art. no. 124212.
[9] W. Sulandari, M. H. Lee, and P. C. Rodrigues, "Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks," *Energy*, vol. 190, Jan. 2020, Art. no. 116408.
[10] L. Tschora, E. Pierre, M. Plantevit, and C. Robardet, "Electricity price forecasting on the day-ahead market using machine learning," *Appl. Energy*, vol. 313, May 2022, Art. no. 118752.
[11] B. Oh, D.-H. Lee, and D. Lee, "Oil-price based long-term hourly system marginal electricity price scenario generation," *IEEE Access*, vol. 10, pp. 25051–25061, 2022.
[12] W. Yang, S. Sun, Y. Hao, and S. Wang, "A novel machine learning-based electricity price forecasting model based on optimal model selection strategy," *Energy*, vol. 238, Jan. 2022, Art. no. 121989.
[13] N. Bibi, I. Shah, A. Alsubie, S. Ali, and S. A. Lone, "Electricity spot prices forecasting based on ensemble learning," *IEEE Access*, vol. 9, pp. 150984–150992, 2021.
[14] A. N. Alkawaz, A. Abdellatif, J. Kanesan, A. S. M. Khairuddin, and H. M. Gheni, "Day-ahead electricity price forecasting based on hybrid regression model," *IEEE Access*, vol. 10, pp. 108021–108033, 2022.
[15] E. Semshchikov, M. Negnevitsky, J. Hamilton, and X. Wang, "Cost-efficient strategy for high renewable energy penetration in isolated power systems," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3719–3728, Sep. 2020.
[16] M. F. Azam and M. S. Younis, "Multi-horizon electricity load and price forecasting using an interpretable multi-head self-attention and EEMD-based framework," *IEEE Access*, vol. 9, pp. 85918–85932, 2021.
[17] K. V. D. Santos, B. Colonetti, E. C. Finardi, and V. M. Zavala, "Accelerated dual dynamic integer programming applied to short-term power generation scheduling," *Int. J. Electr. Power Energy Syst.*, vol. 145, Feb. 2023, Art. no. 108689.
[18] A. Yang, W. Li, and X. Yang, "Short-term electricity load forecasting based on feature selection and least squares support vector machines," *Knowl.-Based Syst.*, vol. 163, pp. 159–173, Jan. 2019.
[19] P. C. Albuquerque, D. O. Cajueiro, and M. D. C. Rossi, "Machine learning models for forecasting power electricity consumption using a high dimensional dataset," *Exp. Syst. Appl.*, vol. 187, Jan. 2022, Art. no. 115917.
[20] M. G. Pinheiro, S. C. Madeira, and A. P. Francisco, "Short-term electricity load forecasting—A systematic approach from system level to secondary substations," *Appl. Energy*, vol. 332, Feb. 2023, Art. no. 120493.
[21] S. Li, X. Kong, L. Yue, C. Liu, M. A. Khan, Z. Yang, and H. Zhang, "Short-term electrical load forecasting using hybrid model of Manta ray foraging optimization and support vector regression," *J. Cleaner Prod.*, vol. 388, Feb. 2023, Art. no. 135856.
[22] A. C. R. Klaar, S. F. Stefenon, L. O. Seman, V. C. Mariani, and L. D. S. Coelho, "Optimized EWT-Seq2Seq-LSTM with attention mechanism to insulators fault prediction," *Sensors*, vol. 23, no. 6, p. 3202, Mar. 2023.
[23] S. F. Stefenon, M. H. D. M. Ribeiro, A. Nied, V. C. Mariani, L. D. S. Coelho, V. R. Q. Leithardt, L. A. Silva, and L. O. Seman, "Hybrid wavelet stacking ensemble model for insulators contamination forecasting," *IEEE Access*, vol. 9, pp. 66387–66397, 2021.
[24] A. Medeiros, A. Sartori, S. F. Stefenon, L. H. Meyer, and A. Nied, "Comparison of artificial intelligence techniques to failure prediction in contaminated insulators based on leakage current," *J. Intell. Fuzzy Syst.*, vol. 42, no. 4, pp. 3285–3298, Mar. 2022.
[25] S. F. Stefenon, M. H. D. M. Ribeiro, A. Nied, K.-C. Yow, V. C. Mariani, L. D. S. Coelho, and L. O. Seman, "Time series forecasting using ensemble learning methods for emergency prevention in hydroelectric power plants with dam," *Electric Power Syst. Res.*, vol. 202, Jan. 2022, Art. no. 107584.
[26] R. G. da Silva, S. R. Moreno, M. H. D. M. Ribeiro, J. H. K. Larcher, V. C. Mariani, and L. D. S. Coelho, "Multi-step short-term wind speed forecasting based on multi-stage decomposition coupled with stacking-ensemble learning approach," *Int. J. Electr. Power Energy Syst.*, vol. 143, Dec. 2022, Art. no. 108504.
[27] S. F. Stefenon, R. Bruns, A. Sartori, L. H. Meyer, R. G. Ovejero, and V. R. Q. Leithardt, "Analysis of the ultrasonic signal in polymeric contaminated insulators through ensemble learning methods," *IEEE Access*, vol. 10, pp. 33980–33991, 2022.
[28] M. H. D. M. Ribeiro, R. G. da Silva, G. T. Ribeiro, V. C. Mariani, and L. D. S. Coelho, "Cooperative ensemble learning model improves electric short-term load forecasting," *Chaos, Solitons Fractals*, vol. 166, Jan. 2023, Art. no. 112982.
[29] S. F. Stefenon, C. Kasburg, R. Z. Freire, F. C. S. Ferreira, D. W. Bertol, and A. Nied, "Photovoltaic power forecasting using wavelet neuro-fuzzy for active solar trackers," *J. Intell. Fuzzy Syst.*, vol. 40, no. 1, pp. 1083–1096, Jan. 2021.
[30] N. W. Branco, M. S. M. Cavalca, S. F. Stefenon, and V. R. Q. Leithardt, "Wavelet LSTM for fault forecasting in electrical power grids," *Sensors*, vol. 22, no. 21, p. 8323, Oct. 2022.
[31] N. F. S. Neto, S. F. Stefenon, L. H. Meyer, R. G. Ovejero, and V. R. Q. Leithardt, "Fault prediction based on leakage current in contaminated insulators using enhanced time series forecasting models," *Sensors*, vol. 22, no. 16, p. 6121, Aug. 2022.
[32] B. J. Souza, S. F. Stefenon, G. Singh, and R. Z. Freire, "Hybrid-YOLO for classification of insulators defects in transmission lines based on UAV," *Int. J. Electr. Power Energy Syst.*, vol. 148, Jun. 2023, Art. no. 108982.
[33] M. P. Corso, F. L. Perez, S. F. Stefenon, K.-C. Yow, R. G. Ovejero, and V. R. Q. Leithardt, "Classification of contaminated insulators using K-nearest neighbors based on computer vision," *Computers*, vol. 10, no. 9, p. 112, Sep. 2021.
[34] S. F. Stefenon, L. O. Seman, B. A. Pavan, R. G. Ovejero, and V. R. Q. Leithardt, "Optimal design of electrical power distribution grid spacers using finite element method," *IET Gener., Transmiss. Distrib.*, vol. 16, no. 9, pp. 1865–1876, May 2022.
[35] M. P. Corso, S. F. Stefenon, G. Singh, M. V. Matsuo, F. L. Perez, and V. R. Q. Leithardt, "Evaluation of visible contamination on power grid insulators using convolutional neural networks," *Electr. Eng.*, vol. 105, no. 6, pp. 3881–3894, Dec. 2023.
[36] S. F. Stefenon, G. Singh, K.-C. Yow, and A. Cimatti, "Semi-ProtoPNet deep neural network for the classification of defective power grid distribution structures," *Sensors*, vol. 22, no. 13, p. 4859, Jun. 2022.
[37] S. F. Stefenon, G. Singh, B. J. Souza, R. Z. Freire, and K. Yow, "Optimized hybrid YOLOu-Quasi-ProtoPNet for insulators classification," *IET Gener., Transmiss. Distrib.*, vol. 17, no. 15, pp. 3501–3511, Aug. 2023.
[38] C. Kasburg and S. Frizzo Stefenon, "Deep learning for photovoltaic generation forecast in active solar trackers," *IEEE Latin Amer. Trans.*, vol. 17, no. 12, pp. 2013–2019, Dec. 2019.
[39] A. C. R. Klaar, S. F. Stefenon, L. O. Seman, V. C. Mariani, and L. D. S. Coelho, "Structure optimization of ensemble learning methods and seasonal decomposition approaches to energy price forecasting in Latin America: A case study about Mexico," *Energies*, vol. 16, no. 7, p. 3184, Mar. 2023.
[40] L. O. Seman, S. F. Stefenon, V. C. Mariani, and L. D. S. Coelho, "Ensemble learning methods using the Hodrick–Prescott filter for fault forecasting in insulators of the electrical power grids," *Int. J. Electr. Power Energy Syst.*, vol. 152, Oct. 2023, Art. no. 109269.
[41] S. F. Stefenon, L. O. Seman, V. C. Mariani, and L. D. S. Coelho, "Aggregating prophet and seasonal trend decomposition for time series forecasting of Italian electricity spot prices," *Energies*, vol. 16, no. 3, p. 1371, Jan. 2023.
[42] M. Ribeiro, S. Stefenon, J. de Lima, A. Nied, V. Mariani, and L. Coelho, "Electricity price forecasting based on self-adaptive decomposition and heterogeneous ensemble learning," *Energies*, vol. 13, no. 19, p. 5190, Oct. 2020.
[43] S. F. Stefenon, L. O. Seman, N. F. S. Neto, L. H. Meyer, V. C. Mariani, and L. D. S. Coelho, "Group method of data handling using Christiano–Fitzgerald random walk filter for insulator fault prediction," *Sensors*, vol. 23, no. 13, p. 6118, Jul. 2023.
[44] H. J. Sadaei, P. C. D. L. E. Silva, F. G. Guimar aes, and M. H. Lee, "Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series," *Energy*, vol. 175, pp. 365–377, May 2019.
[45] Z. Chang, Y. Zhang, and W. Chen, "Electricity price prediction based on hybrid model of Adam optimized LSTM neural network and wavelet transform," *Energy*, vol. 187, Nov. 2019, Art. no. 115804.
[46] J. Kim, J. Moon, E. Hwang, and P. Kang, "Recurrent inception convolution neural network for multi short-term load forecasting," *Energy Buildings*, vol. 194, pp. 328–341, Jul. 2019.
[47] X. Ran, Z. Shan, Y. Fang, and C. Lin, "An LSTM-based method with attention mechanism for travel time prediction," *Sensors*, vol. 19, no. 4, p. 861, Feb. 2019.

[48] S. F. Stefenon, L. O. Seman, L. S. Aquino, and L. D. S. Coelho, ''Wavelet-Seq2Seq-LSTM with attention for time series forecasting of level of dams in hydroelectric power plants,'' *Energy*, vol. 274, Jul. 2023, Art. no. 127350.

[49] M. Yamasaki, R. Z. Freire, L. O. Seman, S. F. Stefenon, V. C. Mariani, and L. D. S. Coelho, ''Optimized hybrid ensemble learning approaches applied to very short-term load forecasting,'' *Int. J. Electr. Power Energy Syst.*, vol. 155, Jan. 2024, Art. no. 109579.

[50] P. Ma, S. Cui, M. Chen, S. Zhou, and K. Wang, ''Review of family-level short-term load forecasting and its application in household energy management system,'' *Energies*, vol. 16, no. 15, p. 5809, Aug. 2023.

[51] J. Zhu, H. Dong, W. Zheng, S. Li, Y. Huang, and L. Xi, ''Review and prospect of data-driven techniques for load forecasting in integrated energy systems,'' *Appl. Energy*, vol. 321, Sep. 2022, Art. no. 119269.

[52] H. Wang, K. A. Alattas, A. Mohammadzadeh, M. H. Sabzalian, A. A. Aly, and A. Mosavi, ''Comprehensive review of load forecasting with emphasis on intelligent computing approaches,'' *Energy Rep.*, vol. 8, pp. 13189–13198, Nov. 2022.

[53] A. L'Heureux, K. Grolinger, and M. A. M. Capretz, ''Transformer-based model for electrical load forecasting,'' *Energies*, vol. 15, no. 14, p. 4993, Jul. 2022.

[54] K. Zhou, W. Wang, T. Hu, and K. Deng, ''Time series forecasting and classification models based on recurrent with attention mechanism and generative adversarial networks,'' *Sensors*, vol. 20, no. 24, p. 7211, Dec. 2020.

[55] R. Shashidhar, S. Patilkulkarni, and S. B. Puneeth, ''Combining audio and visual speech recognition using LSTM and deep convolutional neural network,'' *Int. J. Inf. Technol.*, vol. 14, no. 7, pp. 3425–3436, Dec. 2022.

[56] S. Belagoune, N. Bali, A. Bakdi, B. Baadji, and K. Atif, ''Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems,'' *Measurement*, vol. 177, Jun. 2021, Art. no. 109330.

[57] M. K. Nammous and K. Saeed, ''Natural language processing: Speaker, language, and gender identification with LSTM,'' in *Advanced Computing and Systems for Security*, vol. 8. Singapore: Springer, 2019, pp. 143–156.

[58] S. F. Stefenon, C. Kasburg, A. Nied, A. C. R. Klaar, F. C. S. Ferreira, and N. W. Branco, ''Hybrid deep learning for power generation forecasting in active solar trackers,'' *IET Gener., Transmiss. Distrib.*, vol. 14, no. 23, pp. 5667–5674, Dec. 2020.

[59] A. Pranolo, Y. Mao, A. P. Wibawa, A. B. P. Utama, and F. A. Dwiyanto, ''Robust LSTM with tuned-PSO and bifold-attention mechanism for analyzing multivariate time-series,'' *IEEE Access*, vol. 10, pp. 78423–78434, 2022.

[60] S. F. Stefenon, R. Z. Freire, L. H. Meyer, M. P. Corso, A. Sartori, A. Nied, A. C. Rodrigues Klaar, and K.-C. Yow, ''Fault detection in insulators based on ultrasonic signal processing using a hybrid deep learning technique,'' *IET Sci., Meas. Technol.*, vol. 14, no. 10, pp. 953–961, Dec. 2020.

[61] P. Velivković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, ''Graph attention networks,'' in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, 2018, pp. 1–12.

[62] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, ''Attention is all you need,'' in *Proc. Int. Conf. Neural Inf. Process. Syst.*, vol. 31, 2017, pp. 6000–6010.

[63] G. Singh, S. F. Stefenon, and K.-C. Yow, ''Interpretable visual transmission lines inspections using pseudo-prototypical part network,'' *Mach. Vis. Appl.*, vol. 34, no. 3, p. 41, May 2023.

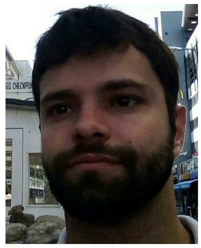
[64] F. Fernandes, S. F. Stefenon, L. O. Seman, A. Nied, F. C. S. Ferreira, M. C. M. Subtil, A. C. R. Klaar, and V. R. Q. Leithardt, ''Long short-term memory stacking model to predict the number of cases and deaths caused by COVID-19,'' *J. Intell. Fuzzy Syst.*, vol. 42, no. 6, pp. 6221–6234, Apr. 2022.

[65] R. N. Muniz, C. T. D. C. Júnior, W. G. Buratto, A. Nied, and G. V. González, ''The sustainability concept: A review focusing on energy,'' *Sustainability*, vol. 15, no. 19, p. 14049, Sep. 2023.

[66] D. L. D. S. Abreu and E. C. Finardi, ''Continuous piecewise linear approximation of plant-based hydro production function for generation scheduling problems,'' *Energies*, vol. 15, no. 5, p. 1699, Feb. 2022.

[67] A. Borré, L. O. Seman, E. Camponogara, S. F. Stefenon, V. C. Mariani, and L. D. S. Coelho, ''Machine fault detection using a hybrid CNN-LSTM attention-based model,'' *Sensors*, vol. 23, no. 9, p. 4512, May 2023.

[68] S. F. Stefenon, R. Z. Freire, L. D. S. Coelho, L. H. Meyer, R. B. Grebogi, W. G. Buratto, and A. Nied, ''Electrical insulator fault forecasting based on a wavelet neuro-fuzzy system,'' *Energies*, vol. 13, no. 2, p. 484, Jan. 2020.



WILLIAM GOUVÊA BURATTO received the B.E. degree in environmental engineering from the State University of Santa Catarina, in 2017, the B.E. degree in electrical engineering from the University of Planalto Catarinense, in 2018, and the master's degree in mechatronics engineering from the Federal Institute of Santa Catarina, Florianópolis, Brazil, in 2020.

He has experience in the areas of thermal treatment of solid waste (pyrolysis and gasification), building electrical projects, energy efficiency in public lighting, work safety in industries, and home automation.

RAFAEL NINNO MUNIZ received the master's degree in energy planning with renewable energy sources and the Ph.D. degree in energy sustainability and intelligent systems. He is currently an Electrical Engineer with an emphasis on energy systems with the Graduate Program in Electrical Engineering, Federal University of Pará (PPGEE/UFPA). He develops research and innovation projects in partnership with Fluminense Federal University (UFF) in the field of sewage sludge treatment through pyrolysis. He is also a Researcher specializing in technologies for waste treatment with energy use. His research interests include sustainability, energy, waste, pyrolysis, gasification, and biodigestion.

ADEMIR NIED (Member, IEEE) received the B.E. degree in electrical engineering from the Federal University of Santa Maria, Santa Maria, Brazil, in 1987, the M.S. degree in industrial informatics from the Federal Technological University of Parana, Curitiba, Brazil, in 1995, and the Ph.D. degree in electrical engineering from the Federal University of Minas Gerais, Belo Horizonte, Brazil, in 2007.

Since 1996, he has been a Faculty Member of Santa Catarina State University, Joinville, Brazil, where he is currently an Associate Professor with the Department of Electrical Engineering. From 2015 to 2016, he was a Visiting Professor with the Wisconsin Electric Machines and Power Electronics Consortium, University of Wisconsin-Madison, Madison, WI, USA. His teaching and research interests include electrical machines, control of electrical drives, neural networks, and renewable energy.

GABRIEL VILLARRUBIA GONZÁLEZ received the master's degree in intelligent systems from the University of Salamanca, in 2012, the master's degree in internet security, in 2014, the master's degree in information systems management, in 2015, and the Ph.D. degree from the Department of Computer Science and Automation, University of Salamanca, Spain.

He is currently an Associate Professor with the University of Salamanca and a Researcher with the Expert Systems and Applications Laboratory (ESALab). Throughout his training, he has followed a well-defined line of research focused on applying multi-agent systems to ambient intelligence environments, particularly concerning the definition of intelligent architectures and information fusion.