

BiGRU-CNN neural network applied to short-term electric load forecasting

Lucas Duarte Soares^{a*} , Edgar Manuel Carreño Franco^{**} 

^aUniversidade Estadual do Oeste do Paraná, Foz do Iguaçu, PR, Brasil

*lucas.2012@alunos.utfpr.edu.br, **emfra.unioeste@gmail.com

Abstract

Paper aims: This study analyzed the feasibility of the BiGRU-CNN artificial neural network as a forecasting tool for short-term electric load. This forecasting model can serve as a support tool related to decision-making by companies in the energy sector.

Originality: Despite a large amount of scientific research in this area, the literature still searches for more assertive forecasting models regarding short-term electric load. Thus, the BiGRU-CNN model, based on layers of BiGRU and CNN architecture networks was tested. This model was already proposed and used for other similar tasks, however, it has not been used on load forecasting.

Research method: The code was programmed in *Python* using the *keras* package. The forecasts of all networks were carried out 10 times until an acceptable statistical sample was reached so that future electric load values are as close as possible to reality.

Main findings: The best forecasting model was the proposed BiGRU-CNN network when compared to classical and some hybrid networks.

Implications for theory and practice: This methodology can be applied to short-term electric load forecasting problems. There is evidence that the combination of different layers of neural networks can provide more efficient forecasting results than classical networks with only one architecture.

Keywords

Time series forecasting. Recurrent neural networks. Artificial intelligence. Machine learning.

How to cite this article: Soares, L. D., & Franco, E. M. C. (2022). BiGRU-CNN neural network applied to short-term electric load forecasting. *Production*, 32, e20210087. <https://doi.org/10.1590/0103-6513.20210087>

Received: July 12, 2021; Accepted: Oct. 27, 2021.

1. Introduction

Economic development around the world depends almost exclusively on the availability of electricity in industries, as most of them use it to carry out their vital productive activities (Soliman & Al-Kandari, 2010). Therefore, the electric load forecast is a decision support tool used by companies in the electricity sector to ensure an efficient service (Hahn et al., 2009), as the lack of electricity has a direct impact on the economy and financial health around the world.

There are four types of electric load forecast horizons to achieve different planning objectives and also to assist in the monitoring of critical conditions in the electrical system. According to Setiawan et al. (2009), these forecast horizons can be classified as very short term, short term, medium term, and long term.

Very short-term energy demand forecasts provide future values between one minute and one hour to determine the best strategy for the use of resources during energy generation (Charytoniuk & Chen, 2000). Short-term forecasts are carried out between one hour and one week to assist in operational planning, as electric load is defined one day before its production around the world (Chapagain et al., 2020). Medium-term forecasts are



made between one week and one month aiming to seek higher profits in the electricity market (Pan & Lee, 2012). Finally, long-term forecasts are conducted above one year, being used as a support tool in the dimensioning of new installations of electricity generation, transmission, and distribution companies (Kandil et al., 2002).

According to Ghalehkondabi et al. (2017), knowing the great importance of electric load forecasts for the electricity sector, published articles related to the subject have grown exponentially in recent years, including studies on short-term electric load forecasting. These studies use techniques that, according to Singh & Khatoon (2013), can be divided into three major groups, namely: traditional techniques formed by regression models (Dudek, 2016), multiple regression (Dhaval & Deshpande, 2020; Johannesen et al., 2019; Saber & Alam, 2018), exponential smoothing (Mayrink & Hippert, 2016; Mohammed et al., 2017; Rendon-Sanchez & Menezes, 2019), modified traditional techniques composed by autoregressive integrated moving average models (Alberg & Last, 2018; Amin & Hoque, 2019; Wu et al., 2020a), support vector machine (Chen et al., 2017; Jiang et al., 2018; Li et al., 2018), and computational techniques. According to Carpinteiro & Silva (2000), the traditional and modified traditional techniques present linear forecasting models, thus showing difficulties in forecasting electric load, as their relationship with exogenous variables is complex and non-linear. The computational techniques, also known as artificial intelligence models, has gained notoriety for its satisfactory performance in these scenarios in which linear models present some difficulty (Singh & Khatoon, 2013).

Some models based on artificial intelligence have been used to forecast short-term electric load, namely (Shahidehpour et al., 2002): expert systems (Kandil et al., 2002; Marković & Fraissler, 1993; Rahman & Hazim, 1996), evolutionary computing (Huang & Yang, 1995; Yang et al., 1996), fuzzy systems (Cerne et al., 2018; Coelho et al., 2016; Mukhopadhyay et al., 2018), artificial neural networks (Chandramitasari et al., 2018), and hybrid models (Fallah et al., 2019; Massaoudi et al., 2021; Yan et al., 2019). Regarding these models, artificial neural networks have received higher attention because their models are more accurate and easier to be implemented and have good performance (Shahidehpour et al., 2002). A considerable number of scientific papers using these forecasting neural models to estimate short-term energy requirements are found in the literature because of these specificities.

Recurrent networks are a peculiar type of artificial neural networks that have become the focus of many studies because of their ability to process sequential and temporal information (Medsker & Jain, 2000). Thus, they have been applied to the short-term electricity demand forecast (Bui et al., 2020). According to Hagan et al. (2014), recurrent neural networks are potentially more powerful than feedforward neural networks, but they present difficulties in training due to the vanishing gradient phenomenon, which leads to unsatisfactory forecasting results. Recurrent architectures based on gates have been used to solve this problem, such as the gated recurrent unit (GRU), which can explore essential long-term information in short-term electricity demand forecasting (Dudek, 2020; Gao et al., 2019; Kuan et al., 2017; Niu et al., 2016; Xiuyun et al., 2018).

The combination of a GRU flowing signals in a specific direction with another GRU carrying information in the opposite direction forms the bidirectional gated recurrent unit (BiGRU) (Luo et al., 2018). This process that generated the BiGRU network can provide more efficient short-term electric load forecasts than the original GRU (Lv et al., 2020). On the other hand, the convolutional neural network (CNN) can also be used to generate more efficient results of short-term electric load forecasts (Massaoudi et al., 2020a; Wu et al., 2021). This kind of network has been gaining notoriety for their results in the field related to pattern recognition, ranging from image processing to voice recognition (Albawi et al., 2017).

Different types of artificial neural network architectures can be used to create models with short-term energy demand forecasts closer to reality. Wu et al. (2020b) combined GRU and CNN networks to form the GRU-CNN forecasting model tested in a real-world experiment, in which its MAPE and RMSE were lower than those of the individual GRU and CNN networks, showing that the proposed hybrid model can use the temporal data more completely to obtain a more accurate short-term energy demand forecast. Sajjad et al. (2020) proposed the CNN-GRU model to be an effective alternative to other hybrid short-term energy demand forecasting models in terms of computational complexity and precision of results due to the representative resources of the extraction potential of the CNN network and efficient gate structure of the multilayer GRU network. Since then, other works have used similar approaches with different results (Xuan et al., 2021).

In addition to the hybrid model constituted by the layers of neural networks of GRU and CNN architecture, another hybrid model found in the literature is the one from the LSTM and CNN networks. Boubaker et al. (2021) used the CNN-LSTM and CNN-BiLSTM models to forecast the solar irradiation demand of a photovoltaic system, but their results were worse than the LSTM, BiLSTM, GRU and BiGRU networks when compared by the RMSE and MAPE metrics. On the other hand Massaoudi et al. (2020b) took advantage of the CNN-LSTM

hybrid neural network to perform the short-term electrical energy demand, where its results were superior to the BiGRU and BiLSTM networks regarding the RMSE, MAE and R^2 comparison metrics.

Given the possibility of using future energy demand values as a tool to support decision-making, this study aims to improve the short-term electricity demand forecast of a company in the electricity sector with the following proposed model based on different layers of artificial neural networks named BiGRU-CNN. The precision of short-term energy demand forecasting can affect the costs and revenues for electricity generators and transmission or distribution operators and, therefore, the profitability and sustainability of these organizations (Islam et al., 2019). Thus, the proposed BiGRU-CNN forecasting neural model of distinct layers was compared with the classical artificial neural networks MLP, CNN, RNN, GRU, and LSTM and the hybrid models GRU-CNN e CNN-BiGRU to verify if its results are more accurate. The historical series of the electric load of a company was used as input in the neural models to make short-term energy demand forecasts.

It is important to clarify that the BiGRU-CNN predictive model has already been used in scientific works in other areas of knowledge, such as: Electoral outcomes (Hadi et al., 2019); Chinese question classification (Liu et al., 2019); Aspect Based Opinion Mining (Sindhu et al., 2021a); Sentiment analysis (Sindhu et al., 2021b) and Multilingual named entity recognition (Ayifu et al., 2019). However, to date, there are no works related to short-term electricity demand forecast. Thus, this article seeks to verify whether this model has the capacity to be used in practical applications by companies in the Electric Energy sector in their decision-making.

2. Theoretical framework

2.1. Gated recurrent unit neural networks

A classical recurrent neural network has a memory function suitable for modeling sequential data, but these algorithms cannot deal with long-distance dependency problems due to gradient explosion and gradient disappearance phenomena (Li et al., 2020). The input gate, forget gate, and output gate, which constitute the LSTM neural network, widely used in the field of sequential forecasts, although with a very complex structure, are created to resolve these impasses (Xiuyun et al., 2018). Another disadvantage of this network is its training time, which is much longer than that of other algorithms. Thus, the gated recurrent unit (GRU) network, which is a special LSTM case, takes advantage of this aspect because it has fewer parameters due to the lack of output gate in its structure (Wang et al., 2018).

The GRU neural network is characterized by gate mechanisms that are especially suited to dealing with time-sequential tasks (Deng et al., 2019). These gate mechanisms are simplified in recurrent cells to significantly increase computational efficiencies in an attempt to maintain the same forecasting performance of the LSTM network (Lv et al., 2020). According to Li et al. (2020), the GRU artificial neural network has two control gates called reset gate and update gate, as shown in Figure 1. The first gate (reset gate) determines how much information needs to be forgotten from the hidden state of the previous instant of time. The information from the previous moment is ignored if its value is close to 0. On the other hand, the hidden information from the past time instant is retained in the current memory when the value is close to 1. The update gate, i.e., the second gate, is responsible for the amount of information in the hidden state of the previous time instant that will be brought to the current hidden state. In this case, the information of the hidden state of the previous instant will be ignored if its value is close to 0, but the information is retained in the current hidden state if the value is close to 1.

The GRU artificial neural network structural unit has two inputs at different time instants, being the current input vector x_t and the output vector h_{t-1} of the previous time instant, in which the output of each gate can be obtained through logical operations and non-linear input transformations (Wang et al., 2018). Equations 1, 2, 3 and 4, which control the functioning of the GRU neural network cell in Figure 1 (Li et al., 2020) are shown below, in which z_t is the update gate, r_t is the reset gate, \tilde{h}_t is the candidate hidden state of the current hidden node, h_t is the current hidden state, x_t is the current input of the artificial neural network, and h_{t-1} is the hidden state of the previous time instant. The activation function sigmoid is represented by σ , w represent the weights for each input x_t and u represent the weights for hidden state of the previous time instant h_{t-1} .

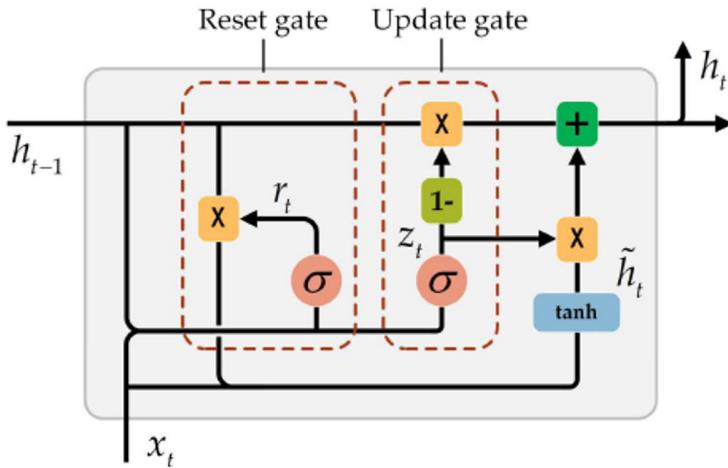


Figure 1. GRU artificial neural network structural unit.

Among these variables, the update gate z_t determines the integration between the information of a new input with the historical information and the reset gate r_t establishes the proportion of the information state in the model (Xiuyun et al., 2018).

$$z_t = \sigma(w_{zx}x_t + u_{zh}h_{t-1}) \quad (1)$$

$$r_t = \sigma(w_{rx}x_t + u_{rh}h_{t-1}) \quad (2)$$

$$\tilde{h}_t = \tanh(w_{hx}x_t + r_t \odot u_{hh}h_{t-1}) \quad (3)$$

$$h_t = (1 - z_t) \odot \tilde{h}_t + z_t \odot h_{t-1} \quad (4)$$

2.2. Bidirectional gated recurrent unit neural networks

The GRU neural network employs the recurrent structure to store and retrieve information for long periods, but its performance in practice may not be as satisfactory as in theory because the network only accesses past information (Deng et al., 2019). The bidirectional GRU (BiGRU) network has a future layer in which the data sequence is in the opposite direction to overcome this problem. Thus, this network uses two hidden layers to extract information from both the past and the future and both are connected in the same output layer (Luo et al., 2018). These characteristics enable the bidirectional structure to assist the recurrent neural networks to extract more information and, consequently, improve the performance of the learning process (Zhang et al., 2018).

Figure 2, taken from Li et al. (2020), shows a BiGRU neural network with two intermediate layers, in which the output layer overlays and normalizes the results of the forward and backward layers at each moment. Its Equations 5, 6, 7, 8 and 9) are shown below, in which \overrightarrow{h}_t^1 and \overrightarrow{h}_t^2 are the output vectors of forward layers of the first and second layers of the BiGRU artificial neural network at time t . On the other hand, the vectors \overleftarrow{h}_t^1 and \overleftarrow{h}_t^2 represent the outputs of the first and second backward layer of the network at the same instant of time t . f is the GRU neural network processing, g is the activation function and w , b are the weight and bias matrices respectively.

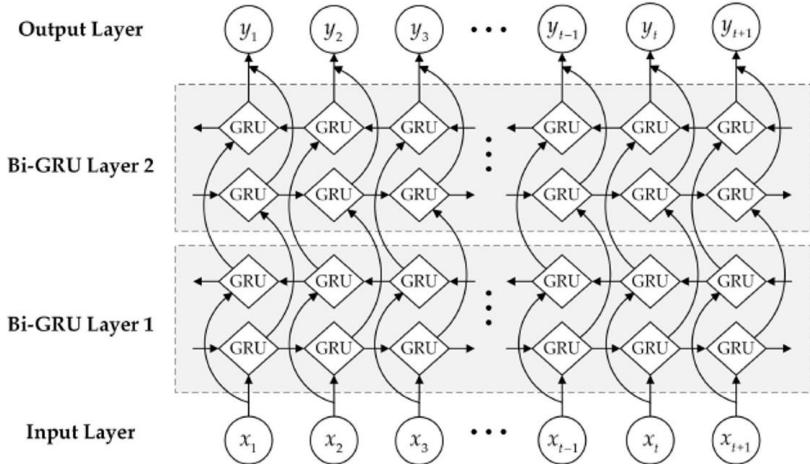


Figure 2. BiGRU neural network structural unit.

.Finally, y_t is the response of the network using past and future information.

$$\overline{h_t^1} = f\left(w_{x\overline{h}^1}x_t + w_{\overline{h}^1\overline{h}^1}\overline{h_{t-1}^1} + b_{\overline{h}^1}\right) \quad (5)$$

$$\overline{\overline{h}_t^1} = f\left(w_{x\overline{h}^1}x_t + w_{\overline{h}^1\overline{h}^1}\overline{h_{t+1}^1} + b_{\overline{h}^1}\right) \quad (6)$$

$$\overline{h_t^2} = f\left(w_{\overline{h}^1\overline{h}^2}\overline{h_t^1} + w_{\overline{h}^2\overline{h}^2}\overline{h_{t-1}^2} + b_{\overline{h}^2}\right) \quad (7)$$

$$\overline{\overline{h}_t^2} = f\left(w_{\overline{h}^1\overline{h}^2}\overline{h_t^1} + w_{\overline{h}^2\overline{h}^2}\overline{h_{t+1}^2} + b_{\overline{h}^2}\right) \quad (8)$$

$$y_t = g\left(w_{\overline{h}^2 y}\overline{h_t^2} + w_{\overline{h}^2 y}\overline{h_{t+1}^2} + b_y\right) \quad (9)$$

2.3. Convolutional neural networks

Convolutional neural network (CNN) is a type of deep artificial neural network often applied to deal with tasks in which data has high local correlations, such as visual images, video prediction, and text categorization, as this specific network can capture the same pattern located in different regions (Tian et al., 2018). Although the CNN network was specially designed to solve image classification problems, in which the network is fed by two-dimensional data, this algorithm is also applied in the field of time series analysis, in which one-dimensional data is used, as the concept of weight sharing is used to increase performance in solving non-linear problems, as seen in electric load forecasts (Sajjad et al., 2020). Basically, weight sharing applies invariance translations

in the neural network model to assist in filtering the learning resource regardless of the spatial properties (Albawi et al., 2017).

The CNN neural network is a feed-forward network, allowing encoding important information contained in the input data with far fewer parameters than in other deep learning models (Zhang et al., 2018). Its standard structure is formed by convolution layers, pooling layers, and, finally, fully connected layers (Tudose et al., 2020). Figure 3, taken from Wu et al. (2021), shows these layers organized generically to compose the CNN network, in which the convolution layer has the function of extracting effective resources from the input data through its multiple internal convolutional kernels, and the pooling layer, added after the convolution layer, can keep strong features and discard weak features to reduce complexity and avoid overfitting. The fully connected layer integrates all local resources to form a global resource used in the calculation of the final result (Tian et al., 2018).

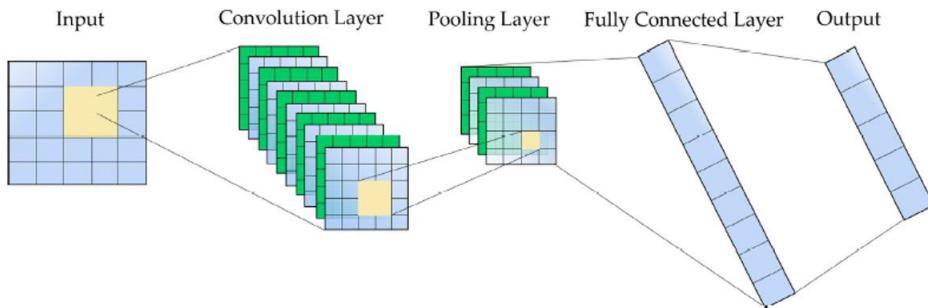


Figure 3. Convolutional neural network structure.

Figure 4 shows a convolutional process from a one-dimensional input (Tudose et al., 2020). This process is controlled by Equation 10, in which $*$ represents the convolutional operation, I corresponds to the one-dimensional input of the current layer, K denotes the one-dimensional Kernel, and S is the convolution output, also known as the feature map.

$$S(i) = (I * K)(i) = \sum_m I(i+m)K(m) \quad (10)$$

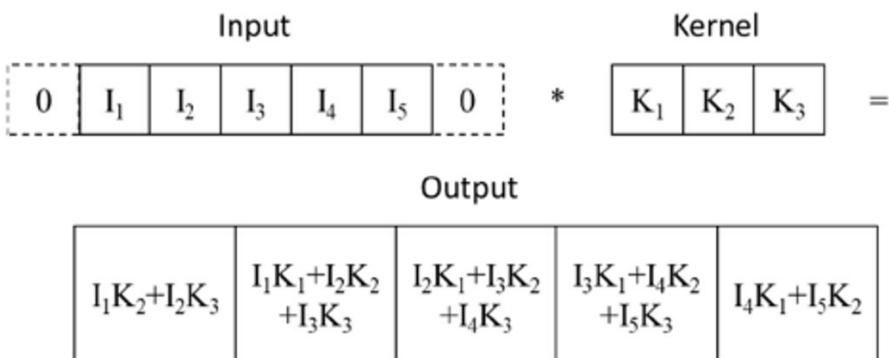


Figure 4. Convolutional process from a 1D input.

3. Proposed method

The BiGRU-CNN model proposed in this study to perform short-term electric load forecasts was built by a BiGRU layer followed by a CNN layer. The input data set referring to the energy demand historical series was manipulated to then feed the BiGRU layer, which performed its processing to extract long-term temporal dependencies. These time-dependent features, which are represented by two hidden state vectors with past and future information, were introduced into the CNN layer so that significant local relationships are captured through the convolution and pooling layers. After this procedure, the data set was structured in several dimensions that had to go through the flatten layer to become one-dimensional again and, therefore, be introduced in the fully connected layer to perform short-term electric load forecasts. The structure of the proposed model mentioned above is shown in Figure 5.

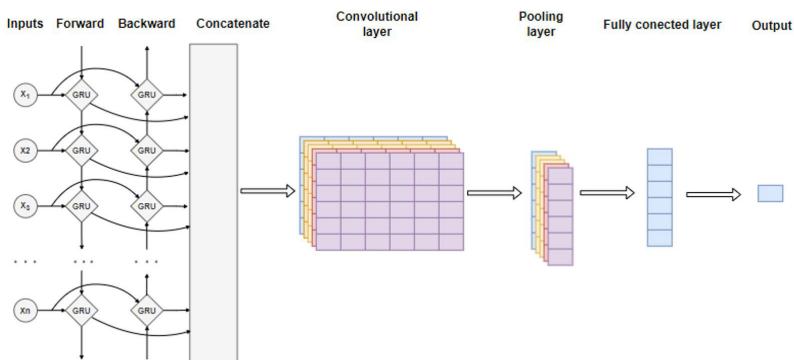


Figure 5. Flowchart of the proposed BiGRU-CNN neural network.

3.1. Time series

A time series is any set of data observed in an orderly manner over time (Morettin & Toloi, 2006). The electric load time series used in this study belongs to the Duke Energy company and is composed of 44,553 hourly observations performed from 1:00 pm on October 1, 2012, to 1:00 am on October 11, 2017. Figure 6 shows the energy demand data over time, in which the ordinate axis is given in megawatts (MW) and the abscissa axis is each time observation given in hours. The file in csv (comma-separated values) format that contained the data from the time series was obtained from <https://www.kaggle.com/robikscube/hourly-energy-consumption>.

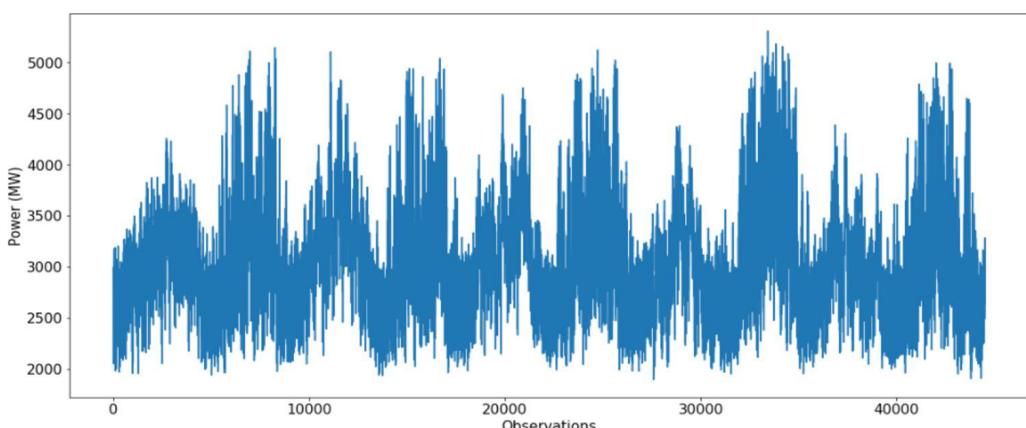


Figure 6. Time series of energy demand.

3.2. Short-term electric load forecast

The proposed BiGRU-CNN model and the well-known classical networks regarding short-term electric load forecast MLP, RNN, GRU, and LSTM were used to forecast in a 24-hour forecasting horizon. Each forecast was performed at an interval of one hour. Thus, 24 forecasts were needed to obtain the forecasting horizon. The future values of short-term electric load requirements were obtained as follows: only one simulation was carried out to train and validate each of the different forecasting models and these models performed multi-step forecasts recursively after being trained and validated. In this type of forecast, future values are fed back into artificial neural networks as if they were observations of the training or validation sample, thus avoiding that the model is trained for each new forecast. In other words, the recursive forecasting model is trained and validated only once to perform 24 effective forecasts, which generates considerable time savings due to the time required to train the networks.

The electric load data were divided into three distinct data set after the time series was transformed into a supervised machine learning problem. The first data set consisted of the training data set, formed by the first 36,199 data; the second data set was that of validation, consisting of 8,330 data after the training data set; and, finally, the test data set was built only by the 24 remaining data.

The training data set aimed to learn the artificial neural networks of the patterns of electric load time series. The internal parameters of the neural networks found during training were tested in the validation data set to verify the forecasting capacity of the network in the data not seen in the previous step. The test data set would be used to evaluate the effective electric load forecasts if the network performance were similar in the training and validation data sets. Effective forecasts are those used for practical purposes, while forecasts made in the training and validation data sets serve to ascertain whether the parameters of the neural networks found during training are capable of generalizing results. The accuracy measures MAPE (mean absolute percentage error), MAE (mean absolute error), and RMSE (root mean square error) presented by Equations 11, 12 and 13, respectively, were used to evaluate the forecasts provided by all artificial neural network models in training, validation, and testing data sets, where x is the desired value, y is the forecasted value, and n is the number of elements in the sample.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{x_i - y_i}{y_i} \right| \quad (11)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (13)$$

3.3. Network code

The algorithm was implemented on Python inside the *Google Colab* environment. The *MinMaxScaler* function of the *sklearn.preprocessing* package was used to normalize the data before being introduced into the neural networks after the temporal data were obtained in the program. Normalization was necessary because energy demand data have a high variation, which could affect the algorithm performance during training, thus providing results not consistent with reality. According to Upadhyaya et al. (2019), preliminary data processing can generate better forecasting results related to short-term energy demand. Equation 14 shows how data normalization was performed by the *MinMaxScaler* function.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (14)$$

The time series was transformed into a supervised machine learning problem after pre-processing the historical data, that is, a sequence of input and output pairs was created so that a decision could be made and then compared with the desired output. The internal parameters of artificial neural networks are modified during training by the *Adam* (adaptive moment estimation) algorithm to allow the difference between the network response and the desired output for a given set of inputs to be minimal. The *Adam* training algorithm was chosen considering the study by Kingma & Ba (2015) where it was considered superior to other algorithms. The ten neurons in each of the two middle layers of all forecasting models have the rectified linear unit (ReLU) as their activation function (Equation 15). The reason for this choice is related to the ability to improve the forecasting performance of recurrent neural networks, according to Talathi & Vartak (2015). Regarding the CNN architecture layers of the neural networks, the kernel_size quantity was set at 6 and the number of filters was set at 8. The training of all artificial neural networks was performed in 150 epochs with a batch size of 32.

$$f(x) = \max(0, x) \quad (15)$$

4. Experiments

Analyzing the results in Table 1 from neural networks of classical MLP, CNN, RNN, GRU and LSTM architectures with regard to short-term electricity demand forecasts, it is evident that the CNN model showed the worst forecasting performance over a 24-hour forecast horizon, this was expected because the CNN network is not able to obtain relevant information from the extraction of nonlinear relations between adjacent samples in local regions.

Table 1. MAPE, MAE and RMSE error for all data.

Model	MAPE	MAE	RMSE
MLP	4.52%	134.36	171.08
CNN	5.58%	165.72	228.16
RNN	3.57%	106.05	138.51
GRU	4.10%	121.50	153.36
LSTM	4.10%	122.04	158.05
GRU-CNN	14.51%	428.46	625.36
CNN-BiGRU	5.10%	151.64	196.09
BiGRU-CNN	3.42%	100.75	122.22

The MLP model showed the second worst forecasting performance over a 24-hour forecast horizon, considering the mean of the lowest MAPE, MAE, and RMSE. It can be explained by the fact that this network is not able to extract the temporal dynamic behavior of the energy demand data because its structure does not have information feedback devices. Following are the LSTM and GRU models, which had very similar performances, mainly in the mean of MAPE and MAE, because their structures have analogous gate mechanisms that work as long-term memory to store essential features responsible for the generating process of the time series. The ability of these two networks to learn long-term temporal patterns was not enough to provide the best short-term energy demand forecast results, which leads to the hypothesis that essential information for this purpose is contained in more recent time features. The formulated hypothesis gains more strength when the performance of the RNN network was superior to the LSTM and GRU networks, as the former does not use the mechanisms that constitute long-term memory and, therefore, very distant temporal information is forgotten during training due to the vanishing gradient phenomenon. Therefore, forecasts from the RNN network are based on recent temporal information and its mean was responsible for placing the model in first place.

In an attempt to increase the efficiency, a GRU architecture network layer followed by a CNN architecture network layer was used to compose the GRU-CNN model, however this model provided predictive results worse than those of the individual networks.

Adding a bidirectional GRU layer at the output of a CNN to form the CNN-BiGRU hybrid model was enough to improve the efficiency of the predictive results when compared with the individual CNN model. However, the short-term energy demand forecasts of the CNN-BiGRU model were not superior to those of classical MLP, RNN, GRU, LSTM architecture networks when compared by MAPE, MAE and RMSE errors. The proposed BiGRU-CNN model had the best forecasting performance, considering the simple mean of the three accuracy metrics when compared to the others.

Therefore feeding a CNN layer to extract local trends and then introducing them into a BiGRU layer so that past and future long-term temporal correlations can be obtained was not as satisfactory as the BiGRU-CNN model. The priority of choosing which layer will be the first architecture of a hybrid neural network model is fundamental for the impact on predictive results.

The analysis of the means of the MAPE, MAE, and RMSE shows that the forecasting performance of the RNN and BiGRU-CNN networks was close, especially in the MAPE error. Thus, more computer simulations were performed to verify whether this behavior is perpetuated and if the BiGRU-CNN network is superior to other networks in its capacity to forecast future values of short-term energy demand. The new simulations are divided into two scenarios using the same conditions and hyperparameters in the networks, differing only regarding the time series size. The first and second scenarios have 77.5 and 66.3% of the original energy demand historical series, respectively (Tables 2 and 3).

Table 2. MAPE, MAE and RMSE error for 77.5% data.

Model	MAPE	MAE	RMSE
MLP	4.69%	140.98	171.36
CNN	14.19%	418.35	533.40
RNN	7.16%	214.51	244.32
GRU	8.22%	246.16	276.39
LSTM	7.02%	209.32	249.83
GRU-CNN	14.63%	430.55	592.70
CNN-BiGRU	9.99%	294.19	377.78
BiGRU-CNN	3.97%	118.95	149.25

Table 3. MAPE, MAE and RMSE error for 66.3% data.

Model	MAPE	MAE	RMSE
MLP	9.65%	284.70	342.53
CNN	6.23%	188.05	234.30
RNN	11.24%	333.21	397.03
GRU	7.73%	228.78	283.18
LSTM	7.75%	231.02	287.56
GRU-CNN	14.18%	420.82	562.35
CNN-BiGRU	8.31%	249.59	304.16
BiGRU-CNN	5.83%	173.23	234.04

The analysis of the tables with simulations involving 77.5 and 66.3% of the original historical series shows that the same relationship between the RNN and BiGRU-CNN networks is not maintained, because in the first one the results of MLP were closer to the BiGRU-CNN network and in the second one the performance of the CNN network was the least different. These results show that the most relevant information to generate more efficient short-term electricity demand forecasts are contained in the short-term temporality when 100% of the data is used.

Therefore, the proposed BiGRU-CNN model showed superiority in these three scenarios relative to the MLP, RNN, GRU, LSTM, GRU-CNN and CNN-BiGRU networks. It is important to highlight that the introduction of the bidirectional mechanism in the GRU-CNN network raised its position from the worst model to the best in the 3 available scenarios. Also, most of the errors increased with a reduction in the number of observations related to the electric load feed to the networks. This event is not unusual, as the lesser the data, the lesser the temporal information to be modeled by the networks during training. Thus, the network parameters found in the training are not able to reliably represent the generating process of the electric load time series and, consequently, their forecasts will not be consistent with reality. The aforementioned processes are better understood with the visualization of Figure 7.

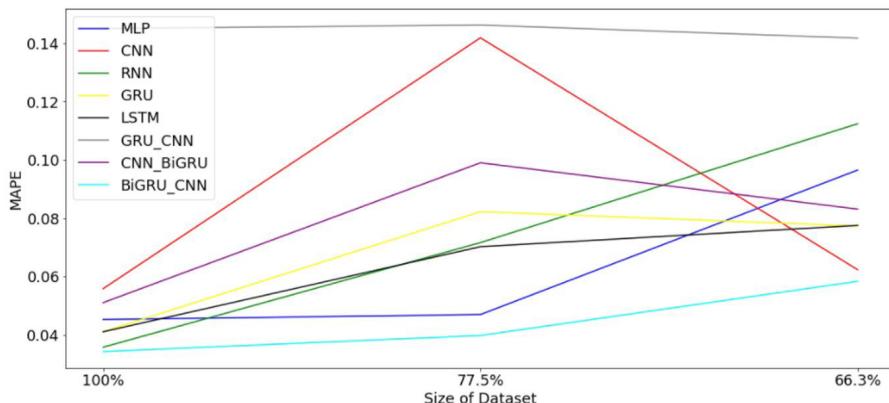


Figure 7. MAPE evolution for all models.

5. Conclusion

This study proposed the BiGRU-CNN model to for short-term electric load forecasts to assist companies in the energy sector in their decision-making. The experimental results showed that feeding a BiGRU layer with the time series to extract its long-term temporal correlations and then introducing these time features into a CNN layer so that local trends can be captured proved to be efficient when compared by the MAPE, MAE, and RMSE errors with the MLP, RNN, GRU, LSTM, GRU-CNN and CNN-BiGRU networks. Further studies need to be conducted by changing the hyperparameters of neural networks and the time series that feed them to ensure that the proposed BiGRU-CNN model can be superior to other forecasting models in terms of short-term electricity demand forecasting.

References

- Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, Aug. 21-23). Understanding of a convolutional neural network. In *International Conference on Engineering and Technology - ICET* (pp. 1-6). Antalya, Turkey: IEEE.
- Alberg, D., & Last, M. (2018). Short-term load forecasting in smart meters with sliding window-based ARIMA algorithms. *Vietnam Journal of Computer Science*, 5(3-4), 241-249. <http://dx.doi.org/10.1007/s40595-018-0119-7>.
- Amin, M. A. A., & Hoque, M. A. (2019, March 13-15). Comparison of ARIMA and SVM for short-term load forecasting. In S. Chakrabarti, & A. Mukherjee (Eds.), *9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference - IEMECON* (pp. 205-210). Jaipur, India: IEEE.
- Ayifu, M., Wushouer, S., & Palidan, M. (2019). Multilingual named entity recognition based on the BiGRU-CNN-CRF hybrid model. *International Journal of Information and Communication Technology*, 15(3), 223-242. <http://dx.doi.org/10.1504/IJICT.2019.102996>.
- Boubaker, S., Benghanem, M., Mellit, A., Lefza, A., Kahouli, O., & Kolsi, L. (2021). Deep neural networks for predicting solar radiation at Hail Region, Saudi Arabia. *IEEE Access: Practical Innovations, Open Solutions*, 9, 36719-36729. <http://dx.doi.org/10.1109/ACCESS.2021.3062205>.
- Bui, V., Nguyen, V. H., Pham, T. L., Kim, J., & Jang, Y. M. (2020, Feb. 19-21). RNN-based deep learning for one-hour ahead load forecasting. In *International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2020* (pp. 587-589). Fukuoka, Japan: IEEE <http://dx.doi.org/10.1109/ICAIIC48513.2020.9065071>.
- Carpinteiro, O. A. S., & Silva, A. P. A. (2000, Nov. 25). A hierarchical neural model in short-term load forecasting. In C. H. C. Ribeiro, & F. M. G. França (Eds.), *Proceedings of Sixth Brazilian Symposium on Neural Networks: Vol. 1* (pp. 120-124). Rio de Janeiro, Brazil: IEEE.

- Cerne, G., Dovzan, D., & Skrijanc, I. (2018). Short-term load forecasting by separating daily profiles and using a single fuzzy model across the entire domain. *IEEE Transactions on Industrial Electronics*, 65(9), 7406-7415. <http://dx.doi.org/10.1109/TIE.2018.2795555>.
- Chandramitasari, W., Kurniawan, B., & Fujimura, S. (2018, Aug. 29-30). Building deep neural network model for short term electricity consumption forecasting. In A. Pranolo, A. Prahara, A. Azhari, & A. Aktawan (Eds.), *International Symposium on Advanced Intelligent Informatics: Revolutionize Intelligent Informatics Spectrum for Humanity - SAIN* (pp. 43-48). Yogyakarta, Indonesia: IEEE.
- Chapagain, K., Kittipiyakul, S., & Kulthanavit, P. (2020). Short-term electricity demand forecasting: impact analysis of temperature for Thailand. *Energies*, 13(10), 1-29. <http://dx.doi.org/10.3390/en13102498>.
- Charytoniuk, W., & Chen, M. S. (2000). Very short-term load forecasting using artificial. *IEEE Transactions on Power Systems*, 15(1), 263-268. <http://dx.doi.org/10.1109/59.852131>.
- Chen, Y., Xu, P., Chu, Y., Li, W., Wu, Y., Ni, L., Bao, Y., & Wang, K. (2017). Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. *Applied Energy*, 195, 659-670. <http://dx.doi.org/10.1016/j.apenergy.2017.03.034>.
- Coelho, V. N., Coelho, I. M., Coelho, B. N., Reis, A. J. R., Enayatifar, R., Souza, M. J. F., & Guimarães, F. G. (2016). A self-adaptive evolutionary fuzzy model for load forecasting problems on smart grid environment. *Applied Energy*, 169, 567-584. <http://dx.doi.org/10.1016/j.apenergy.2016.02.045>.
- Deng, Y., Jia, H., Li, P., Tong, X., Qiu, X., & Li, F. (2019, June 19-21). A deep learning methodology based on bidirectional gated recurrent unit for wind power prediction. In *Proceedings of the 14th IEEE Conference on Industrial Electronics and Applications - ICIEA 2019* (pp. 591-595). Xi'an, China: IEEE. <http://dx.doi.org/10.1109/ICIEA.2019.8834205>
- Dhaval, B., & Deshpande, A. (2020). Short-term load forecasting with using multiple linear regression. *Iranian Journal of Electrical and Computer Engineering*, 10(4), 3911-3917. <http://dx.doi.org/10.11591/ijece.v10i4.pp3911-3917>.
- Dudek, G. (2016). Pattern-based local linear regression models for short-term load forecasting. *Electric Power Systems Research*, 130, 139-147. <http://dx.doi.org/10.1016/j.epsr.2015.09.001>.
- Dudek, G. (2020). Multilayer perceptron for short-term load forecasting: from global to local approach. *Neural Computing & Applications*, 32(8), 3695-3707. <http://dx.doi.org/10.1007/s00521-019-04130-y>.
- Fallah, S. N., Ganjkhani, M., Shamshirband, S., & Chau, K. (2019). Computational intelligence on short-term load forecasting: a methodological overview. *Energies*, 12(3), 393. <http://dx.doi.org/10.3390/en12030393>.
- Gao, X., Li, X., Zhao, B., Ji, W., Jing, X., & He, Y. (2019). Short-term electricity load forecasting model based on EMD-GRU with feature selection. *Energies*, 12(6), 1-18. <http://dx.doi.org/10.3390/en12061140>.
- Ghalehkondabi, I., Ardjmand, E., Weckman, G. R., & Young, W. A. (2017). An overview of energy demand forecasting methods published in 2005-2015. *Energy Systems*, 8, 411-447. <http://dx.doi.org/10.1007/s12667-016-0203-y>.
- Hadi, K. A., Lasri, R., & Abderrahmani, A. E. (2019). Social data analytics for forecasting electoral outcomes. *International Journal of Innovative Technology and Exploring Engineering*, 8(8), 2468-2471.
- Hagan, M. T., Demuth, H. B., & Beale, M. H. (2014). *Neural network design* (2nd ed.). Oklahoma: OSU.
- Hahn, H., Meyer-Nieberg, S., & Pickl, S. (2009). Electric load forecasting methods: tools for decision making. *European Journal of Operational Research*, 199(3), 902-907. <http://dx.doi.org/10.1016/j.ejor.2009.01.062>.
- Huang, C., & Yang, H. (1995, Nov. 21-23). A time series approach to short term load forecasting through evolutionary programming structures. In *Proceedings of the International Conference on Energy Management and Power Delivery - EMPD* (Vol. 2, pp. 583-588). Singapore: IEEE.
- Islam, M. A., Che, H. S., Hasanuzzaman, M., & Rahim, N. A. (2019). Energy demand forecasting. In M. Hasanuzzaman & N. A. Rahim (Eds.), *Energy for sustainable development: demand, supply, conversion and management*. London: Academic Press/Elsevier.
- Jiang, H., Zhang, Y., Muljadi, E., Zhang, J. J., & Gao, D. W. (2018). A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization. *IEEE Transactions on Smart Grid*, 9(4), 3331-3350. <http://dx.doi.org/10.1109/TSG.2016.2628061>.
- Johannesen, N. J., Kolhe, M., & Goodwin, M. (2019). Relative evaluation of regression tools for urban area electrical energy demand forecasting. *Journal of Cleaner Production*, 218, 555-564. <http://dx.doi.org/10.1016/j.jclepro.2019.01.108>.
- Kandil, M. S., El-Debeiky, S. M., & Hasanien, N. E. (2002). Long-term load forecasting for fast developing utility using a knowledge-based expert system. *IEEE Transactions on Power Systems*, 17(2), 491-496. <http://dx.doi.org/10.1109/TPWRS.2002.1007923>.
- Kingma, D. P., & Ba, J. L. (2015, May. 7-9). Adam: a method for stochastic optimization. In Y. Bengio & Y. LeCun (Eds.), *3rd International Conference on Learning Representations - ICLR 2015 - Conference Track Proceedings* (pp. 1-15). San Diego: OpenReview.net.
- Kuan, L., Yan, Z., Xin, W., Yan, C., Xiangkun, P., Wenxue, S., Zhe, J., Yong, Z., Nan, X., & Xin, Z. (2017, Nov. 26-28). Short-term electricity load forecasting method based on multilayered self-normalizing GRU network. In F. Gao (Ed.), *IEEE Conference on Energy Internet and Energy System Integration - EI2* (pp. 1-5). Beijing, China: IEEE. <http://dx.doi.org/10.1109/EI2.2017.8245330>
- Li, P., Luo, A., Liu, J., Wang, Y., Zhu, J., Deng, Y., & Zhang, J. (2020). Bidirectional gated recurrent unit neural network for Chinese address element segmentation. *ISPRS International Journal of Geo-Information*, 9(11), 635. <http://dx.doi.org/10.3390/ijgi9110635>.
- Li, Y., Che, J., & Yang, Y. (2018). Subsampled support vector regression ensemble for short term electric load forecasting. *Energy*, 164, 160-170. <http://dx.doi.org/10.1016/j.energy.2018.08.169>.
- Liu, J., Yang, Y., Lv, S., Wang, J., & Chen, H. (2019). Attention-based BiGRU-CNN for Chinese question classification. *Journal of Ambient Intelligence and Humanized Computing*, 10(13), 1-12. <https://doi.org/10.1007/s12652-019-01344-9>
- Luo, X., Zhou, W., Wang, W., Zhu, Y., & Deng, J. (2018). Attention-based relation extraction with bidirectional gated recurrent unit and highway network in the analysis of geological data. *IEEE Access: Practical Innovations, Open Solutions*, 6, 5705-5715. <http://dx.doi.org/10.1109/ACCESS.2017.2785229>.
- Lv, P., Liu, S., Yu, W., Zheng, S., & Lv, J. (2020). EGA-STLF: a hybrid short-term load forecasting model. *IEEE Access: Practical Innovations, Open Solutions*, 8, 31742-31752. <http://dx.doi.org/10.1109/ACCESS.2020.2973350>.
- Marković, M. L., & Fraissler, W. F. (1993). Short-term load forecast by plausibility checking of announced demand: An expert-system approach. *European Transactions on Electrical Power*, 3(5), 353-358. <http://dx.doi.org/10.1002/etep.4450030506>.

- Massaoudi, M., Refaat, S. S., Abu-Rub, H., Chihi, I., & Oueslati, F. S. (2020a). PLS-CNN-BiLSTM: an end-to-end algorithm-based savitzky-golay smoothing and evolution strategy for load forecasting. *Energies*, 13(20), 1-29. <http://dx.doi.org/10.3390/en13205464>.
- Massaoudi, M., Refaat, S. S., Chihi, I., Trabelsi, M., Abu-Rub, H., & Oueslati, F. S. (2020b). Short-term electric load forecasting based on data-driven deep learning techniques. In *IECON - The 46th Annual Conference of the IEEE Industrial Electronics Society* (pp. 2565-2570). Singapore: IEEE.
- Massaoudi, M., Refaat, S. S., Chihi, I., Trabelsi, M., Oueslati, F. S., & Abu-Rub, H. (2021). A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for short-term load forecasting. *Energy*, 214, 118874. <http://dx.doi.org/10.1016/j.energy.2020.118874>.
- Mayrink, V., & Hippert, H. S. (2016). A hybrid method using exponential smoothing and gradient boosting for electrical short-term load forecasting. In C. Rodríguez, & J. B. Gómez (Eds.), *IEEE Latin American Conference on Computational Intelligence - LA-CCI*. Cartagena, Colombia: IEEE.
- Medsker, L. R., & Jain, L. C. (2000). *Recurrent neural networks: design and applications*. Boca Raton: CRC Press.
- Mohammed, J., Bahadoorsingh, S., Ramsamooj, N., & Sharma, C. (2017, June 18-22). Performance of exponential smoothing, a neural network and a hybrid algorithm to the short term load forecasting of batch and continuous loads. In *IEEE Manchester PowerTech*. Manchester, UK: IEEE.
- Morettin, P. A., & Toloi, C. M. C (2006). Análise de séries temporais (2. ed.). São Paulo, Brazil Blucher.
- Mukhopadhyay, P., Mitra, G., Banerjee, S., & Mukherjee, G. (2018, Dec. 21-23). Electricity load forecasting using fuzzy logic: Short term load forecasting factoring weather parameter. In *7th International Conference on Power Systems - ICPS* (pp. 812-819). Pune, India: IEEE.
- Niu, M., Sun, S., Wu, J., Yu, L., & Wang, J. (2016). An innovative integrated model using the singular spectrum analysis and nonlinear multi-layer perceptron network optimized by hybrid intelligent algorithm for short-term load forecasting. *Applied Mathematical Modelling*, 40(5-6), 4079-4093. <http://dx.doi.org/10.1016/j.apm.2015.11.030>.
- Pan, X., & Lee, B. (2012). A comparison of support vector machines and artificial neural networks for mid-term load forecasting. In *IEEE International Conference on Industrial Technology, ICIT* (pp. 95-101). Athens, Greece: IEEE.
- Rahman, S., & Hazim, O. (1996). Load forecasting for multiple sites: development of an expert system-based technique. *Electric Power Systems Research*, 39(3), 161-169. [http://dx.doi.org/10.1016/S0378-7796\(96\)01114-5](http://dx.doi.org/10.1016/S0378-7796(96)01114-5).
- Rendon-Sanchez, J. F., & Menezes, L. M. (2019). Structural combination of seasonal exponential smoothing forecasts applied to load forecasting. *European Journal of Operational Research*, 275(3), 916-924. <http://dx.doi.org/10.1016/j.ejor.2018.12.013>.
- Saber, A. Y., & Alam, A. K. M. R. (2018). Short term load forecasting using multiple linear regression for big data. In *IEEE Symposium Series on Computational Intelligence - SSCl* (pp. 1-6). Honolulu, HI, USA: IEEE.
- Sajjad, M., Khan, Z. A., Ullah, A., Hussain, T., Ullah, W., Lee, M. Y., & Baik, S. W. (2020). A novel CNN-GRU-based hybrid approach for short-term residential load forecasting. *IEEE Access : Practical Innovations, Open Solutions*, 8, 143759-143768. <http://dx.doi.org/10.1109/ACCESS.2020.3009537>.
- Setiawan, A., Koprinska, I., & Agelidis, V. G. (2009). Very short-term electricity load demand forecasting using support vector regression. In *International Joint Conference on Neural Networks* (pp. 2888-2894). Atlanta, GA, USA: IEEE. <http://dx.doi.org/10.1109/IJCNN.2009.5179063>
- Shahidehpour, M., Yamin, H., & Li, Z. (2002). *Market operations in electric power systems: forecasting, scheduling, and risk management* (1st ed.). Hoboken: Wiley. <http://dx.doi.org/10.1002/047122412X>.
- Sindhu, C., Som, B., & Singh, S. P. (2021a). Aspect based opinion mining leveraging weighted bigru and CNN module in parallel. In *International Conference on Intelligent Technologies - CONIT* (pp. 1-7). Hubli, India: IEEE. <http://dx.doi.org/10.1109/CONIT51480.2021.9498441>.
- Sindhu, C., Som, B., & Singh, S. P. (2021b). Aspect-oriented sentiment classification using BiGRU-CNN model. In *5th International Conference on Computing Methodologies and Communication - ICCMC* (pp. 984-989). Erode, India: IEEE..
- Singh, A. K., & Khatoon, S. (2013). An overview of electricity demand forecasting techniques. *National Conference on Emerging Trends in Electrical, Instrumentation & Communications Engineer*, 3(3), 38-48.
- Soliman, S. A., & Al-Kandari, A. M. (2010). *Electrical load forecasting: modeling and model construction* (1st ed.). Oxford: Butterworth-Heinemann.
- Talathi, S. S., & Vartak, A. (2015). Improving performance of recurrent neural network with relu nonlinearity. *Neural and Evolutionary Computing*, 1, ArXiv:1511.03771. Retrieved in 2021 November 04, from <http://arxiv.org/abs/1511.03771>
- Tian, C., Ma, J., Zhang, C., & Zhan, P. (2018). A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network. *Energies*, 11(12), 3493. <http://dx.doi.org/10.3390/en11123493>.
- Tudose, A. M., Sidea, D. O., Picioroaga, I. I., Boicea, V. A., & Bulac, C. (2020). A CNN based model for short-term load forecasting: a real case study on the Romanian power system. In *55th International Universities Power Engineering Conference - UPEC*. Turin, Italy: IEEE.
- Upadhyaya, D., Thakur, R., & Singh, N. K. (2019). A systematic review on the methods of short term load forecasting. In *2nd International Conference on Power Energy Environment and Intelligent Control - PEEIC* (pp. 6-11). Greater Noida, India: IEEE.
- Wang, Y., Liao, W., & Chang, Y. (2018). Gated recurrent unit network-based short-term photovoltaic forecasting. *Energies*, 11(8), 2163. <https://doi.org/10.3390/en11082163>.
- Wu, F., Cattani, C., Song, W., & Zio, E. (2020a). Fractional ARIMA with an improved cuckoo search optimization for the efficient short-term power load forecasting. *Alexandria Engineering Journal*, 59(5), 3111-3118. <http://dx.doi.org/10.1016/j.aej.2020.06.049>.
- Wu, K., Wu, J., Feng, L., Yang, B., Liang, R., Yang, S., & Zhao, R. (2021). An attention-based CNN-LSTM-BiLSTM model for short-term electric load forecasting in integrated energy system. *International Transactions on Electrical Energy Systems*, 31(1), 1-15. <http://dx.doi.org/10.1002/2050-7038.12637>.
- Wu, L., Kong, C., Hao, X., & Chen, W. (2020b). A short-term load forecasting method based on GRU-CNN hybrid neural network model. *Mathematical Problems in Engineering*, 2020, 1-10. <http://dx.doi.org/10.1155/2020/1428104>.

- Xiuyun, G., Ying, W., Yang, G., Chengzhi, S., Wen, X., & Yimiao, Y. (2018). Short-term load forecasting model of gru network based on deep learning framework. In *2nd IEEE Conference on Energy Internet and Energy System Integration - EI2* (pp. 1–4). Beijing, China: IEEE <http://dx.doi.org/10.1109/EI2.2018.8582419>
- Xuan, Y., Si, W., Zhu, J., Sun, Z., Zhao, J., Xu, M., & Xu, S. (2021). Multi-model fusion short-term load forecasting based on random forest feature selection and hybrid neural network. *IEEE Access : Practical Innovations, Open Solutions*, 9, 69002-69009. <http://dx.doi.org/10.1109/ACCESS.2021.3051337>.
- Yan, K., Li, W., Ji, Z., Qi, M., & Du, Y. (2019). A hybrid LSTM neural network for energy consumption forecasting of individual households. *IEEE Access : Practical Innovations, Open Solutions*, 7, 157633-157642. <http://dx.doi.org/10.1109/ACCESS.2019.2949065>.
- Yang, H., Huang, C., & Huang, C. (1996). Identification of ARMAX model for short term load forecasting: an evolutionary programming approach. *IEEE Transactions on Power Systems*, 11(1), 403-408. <http://dx.doi.org/10.1109/59.486125>.
- Zhang, D., Tian, L., Hong, M., Han, F., Ren, Y., & Chen, Y. (2018). Combining convolution neural network and bidirectional gated recurrent unit for sentence semantic classification. *IEEE Access: Practical Innovations, Open Solutions*, 6(8), 73750-73759. <http://dx.doi.org/10.1109/ACCESS.2018.2882878>.