U.P.B. Sci. Bull., Series C, Vol. 83, Iss. 2, 2021 ISSN 2286-3540

NEURAL NETWORKS APPLICATION IN SHORT-TERM LOAD FORECASTING

Andrei TUDOSE[1], Irina PICIOROAGA[2], Dorian SIDEA[3], Constantin BULAC[4]

Short-term load forecasting (STLF) is a fundamental procedure in power systems operation that underlies the most important decision-making processes, such as economic dispatch or equipment maintenance planning. Due to the high degree of uncertainties in demand variations, advanced techniques based on artificial intelligence are needed in order to obtain an accurate electrical load forecasting. In this paper, multiple forecasting methods based on neural networks, including the multilayer perceptron (MLP), convolutional neural networks (CNN), long short-term memory (LSTM) and gated recurrent unit (GRU), are applied to solve the STLF problem, using a real dataset provided by the Romanian TSO. In this regard, the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE) are used as evaluation metrics for the day-ahead load forecasting results.

Keywords: artificial intelligence, load forecasting, neural networks, power systems

1. Introduction

The modern society development heavily relies on an appropriate power supply. Providing this service at the lowest possible cost involves the accurate covering of demand fluctuations and system losses by generating the appropriate amount of energy. The principle by which utility companies provide the energy needs to their users consists in estimating the electrical load in advance, through already known consumption patterns and various factors that may influence them (weather conditions, economic growth, social events etc.). This analysis is known as the "electricity consumption forecasting". Therefore, the proper power systems operation is centered around the power demanded by their various consumers, as generation is scheduled to follow the load, while both transmission and distribution grids need to provide a reliable connection between consumption and supply. In this context, load forecasting represents one of the most important

[1] PhD. Student, Dept. of Power Systems, Faculty of Power Engineering, University POLITEHNICA of Bucharest, e-mail: andrei.tudose1604@upb.ro

[2] Teaching assistant, Dept. of Power Systems, Faculty of Power Engineering, University POLITEHNICA of Bucharest, e-mail: irina.picioroaga@upb.ro

[3] Lect., Dept. of Power Systems, Faculty of Power Engineering, University POLITEHNICA of Bucharest, e-mail: dorian.sidea@upb.ro

[4] Prof., Dept. of Power Systems, Faculty of Power Engineering, University POLITEHNICA of Bucharest, e-mail: constantin.bulac@upb.ro

inputs in power system planning to correctly define all the scenarios under study [1]. From a technical point of view, load forecasting is essential for security assessments in power systems, as their behavior can be simulated in this manner at some future time in order to verify if a secure operation can be achieved in a variety of operating conditions. Furthermore, load forecasting is of high economic importance, since it facilitates decision-making in various activities of power systems operation, such as the allocation of certain equipment, scheduling of different maintenance operations or the economic dispatch of power between generation units [2].

The prediction interval for load forecasting studies can vary from a few minutes (for stability assessment) to a few decades (for investment planning). Among these analyses, the short-term load forecast plays a critical role in ensuring the safety, stability and efficiency in power systems operation.

Many approaches have been proposed over the years to solve the short-term load forecasting. The early load forecasting models were mostly based on statistical methods, such as the similar day approach, Box-Jenkins basic models (ARMA [3], ARIMA [4] or ARIMAX [5]) or exponential smoothing [6]. However, these traditional methods fail in adapting to the continuously changing profile of the electrical load. Therefore, recent studies focus on modern techniques using artificial intelligence, such as fuzzy logic [7] and support vector regression [8]. In this paper, several artificial intelligence techniques, including the multi-layer perceptron (MLP), convolutional neural networks (CNN), long short-term memory (LSTM) and gated recurrent unit (GRU), are applied for solving the short-term load forecasting problem.

The main contribution of the authors is the development of a general neural network-based framework capable of employing the mentioned neural network models (MLP, CNN, LSTM and GRU) to solve the STLF problem for the aggregated consumption in the Romanian power system. A comparison study is performed in order to identify the most efficient model, using the mean absolute error (MAE), the mean absolute percentage error (MAPE), the mean squared error (MSE) and the root mean square error (RMSE) as assessment indexes.

2. Neural Networks

The artificial intelligence domain has become a topic of high interest for the research community, primarily due to recent advancements of hardware components, such as graphical processing units (GPU), which allow the design of neural networks models with increased number of layers, while also ensuring faster convergence time compared to central processing units (CPU). Among these, neural networks (NN) are capable to learn complex features from the input data and attain good accuracy for various applications.

For regression problems, such as the load forecasting, many applications employ the classical MLP, as it is easier to implement, since it involves less parameters to set. More complex techniques, such as CNN, LSTM and GRU, are used as well, since they are specialized in sequence modelling. Further in this section, a brief overview of the mentioned neural networks is presented.

A. Multi-layer perceptron

The multi-layer perceptron (or feed-forward neural network) represents a fundamental class of neural networks, being able to generate good results on both classification and regression problems.

Full connectivity between consecutive layers provides MLP the capability to solve complex problems, but with the cost of a high computational effort required in the training process [9]. This leads to a slower learning process, which represents a major disadvantage of MLP, as the fine-tuning of neural networks, which implies refitting of the model several times with different parameters, is a necessary step in improving the performance of the model.

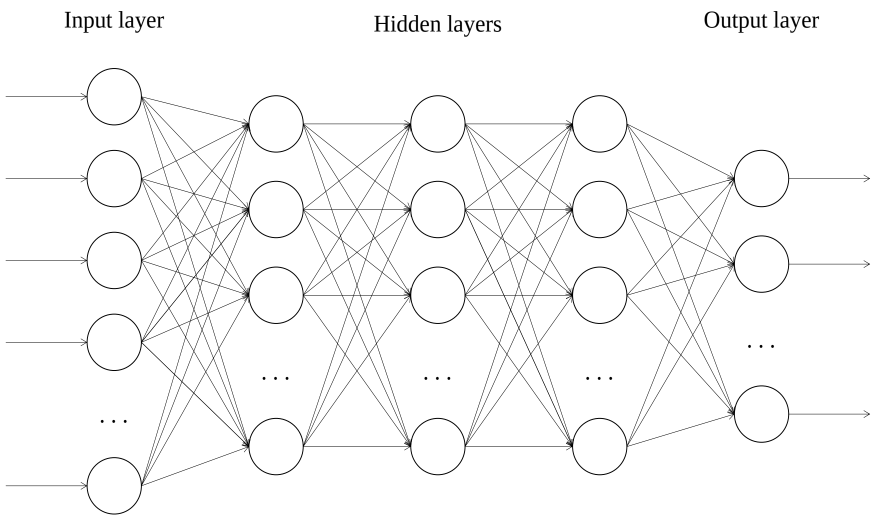


Fig. 1 General structure of the MLP

B. Convolutional Neural Networks

CNNs represent one of the main topics in the Deep Learning paradigm, mostly due to their applications in object recognition problems [10]. The main requirement of a CNN is the input defining as an array (i.e. an array of pixel intensities or a sequence of values in a time-series). Thus, CNNs can be easily adapted to solve load forecasting problems [11].

CNNs have the capability to detect and extract features from neighbouring values. Thus, in a time-series problem, the convolutional layers of a CNN emphasize the relationship between the values at consecutive timesteps. The pooling layers are used to attain a dimensionality reduction of the features, improving the training speed, while also keeping the most vital information.

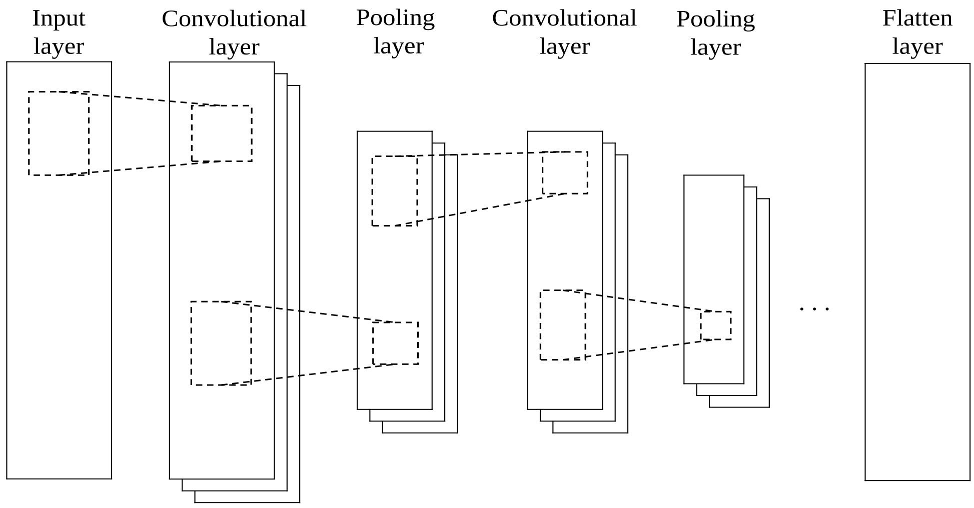
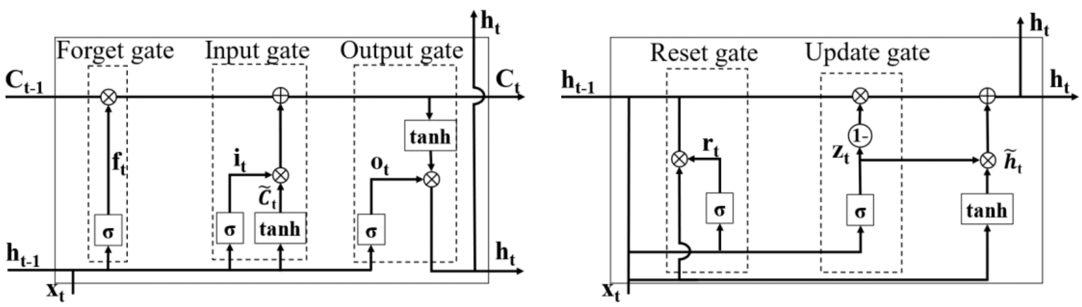


Fig. 2 General structure of a convolutional neural network

C. Recurrent Neural Networks

The capability of learning time-dependent features from a sequence makes the RNN a suitable model for time-series forecasting. Simple RNNs present the drawback of the vanishing gradient during training, issue that is overcome by the more complex units, the long short-term memory (LSTM) [12] and gated recurrent unit (GRU) [13], which are depicted in Fig. 3. LSTM and GRU cells consist in various gates, that perform multiple operations in order to establish the cell output. LSTM contains three gates: the forget gate (f_t), the input gate (i_t) and the output gate (o_t), while a GRU cell, although accomplishes a similar purpose, consists in only two gates: the reset gate (r_t) and the update gate (z_t).



a) b)

Fig.3 Illustration of: a. LSTM cell, b. GRU cell

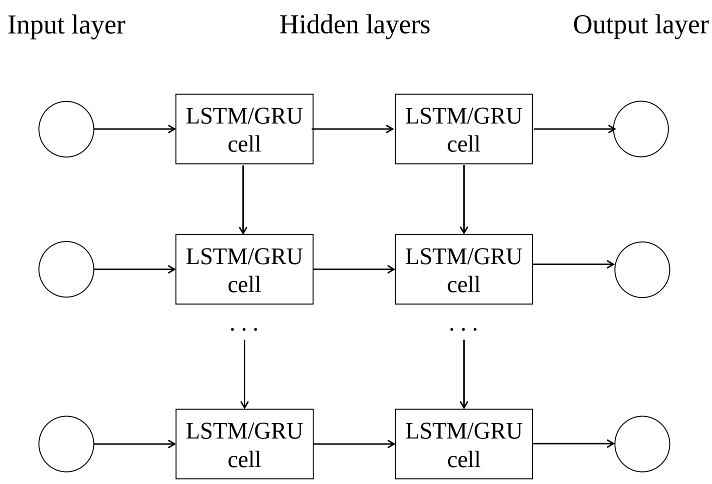


Fig. 4 General structure of the LSTM/GRU network

Most applications in the field of natural language processing, a topic of high importance in the era of digitalization, are based on RNN and its' variations, as they achieve great performances [14]. Results presented in [15] show that the efficiency in solving the STLF problem obtained by LSTM and GRU is similar, making them a reliable tool to use. Despite showing better results than the simple RNN, LSTM and GRU have more weights to update, which leads to a longer training time. In general, GRU trains faster than LSTM, as a GRU has only 3 trainable parameters for each unit, while a LSTM unit has 4. Accuracy may vary depending on the problem and the structure and size of the dataset, thus an analysis of both is required.

3. The NN model for solving the STLF problem

Good accuracy of the load forecasting model heavily relies on external variables, such as weather factors or the time of the year/week. Most of the neural networks will not work properly if these external variables are fed directly into them, as they require the input to be a sequence of values from the analyzed time-series. In this section, the general framework used for solving the STLF problem based on artificial intelligence is described.

A way to incorporate the external variables is to firstly feed the sequence of the previous load into a neural network, such as a CNN, GRU or LSTM, and secondly, concatenate the output of this neural network with the external variables and feed them into a fully-connected layer. The output layer consists in 24 neurons, representing the day-ahead hourly energy consumption forecasting. This technique is frequently used to include external variables in time-series forecasting [16], [17]. The general architecture of the implemented model is depicted in Fig. 5.

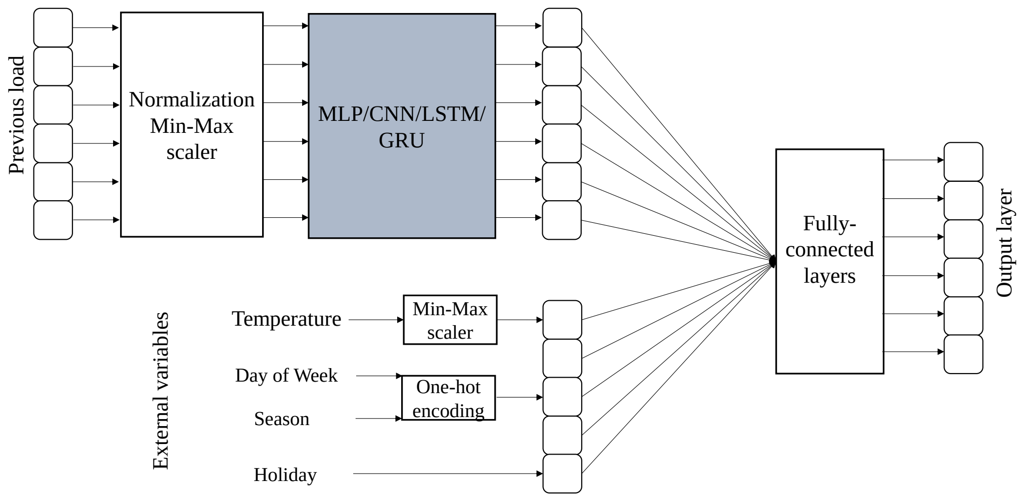


Fig. 5. The architecture of the implemented model

The ReLU (rectified linear unit) activation function, described by equation (1), is used for the MLP and CNN implementation. The fully connected layers also employ ReLU. For the LSTM and GRU networks, the sigmoid and hyperbolic tangent activation functions are used, according to the cells' structure previously presented in Fig. 3.

$$y_{\mathrm{ReLU}} = \begin{cases} x, \ if \ \mathrm{x>0} \\ 0, \ \mathrm{if\ x \leq 0} \end{cases} \tag{1}$$

Data normalization is mandatory for some neural network types, such as the recurrent types (LSTM and GRU), as they are using tanh and sigmoid activation functions, which are prone to the saturation effect. Also, neural networks' performance generally improves when the input data are scaled to the same range of values. In the developed model, the min-max normalization method is employed, which is described by the following equation:

$$x_{norm} = \frac{x - \min(X)}{\max(X) - \min(X)} \tag{2}$$

where *x* is the value to be normalized from the array *X*.

Hourly energy consumption values, as well as temperature values, are normalized using the min-max method. The day of week and the season are extracted from the calendar date and are processed using one-hot encoding technique, while the holiday influence is quantified using a binary variable. The assessment of the forecasting results is done by applying the following four frequently used evaluation metrics: mean absolute percentage error (MAPE), mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE), described by equations (3)-(6) [18]:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|\cdot 100 \tag{3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{4}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{6}$$

where n is the number of samples, y and $\hat{y}$ are the real load and the predicted load, respectively.

4. Case study

A. Dataset

In this paper, the model previously described is applied to solve the day-ahead load forecasting for the Romanian power system, with a resolution of one hour. The dataset consists in hourly aggregated energy consumption at national level, available at [19], daily minimum, average and maximum temperatures measured in Bucharest, obtained from [20] and data derived from the calendar date, collected for a period of 7 years, from 2012 to 2019.

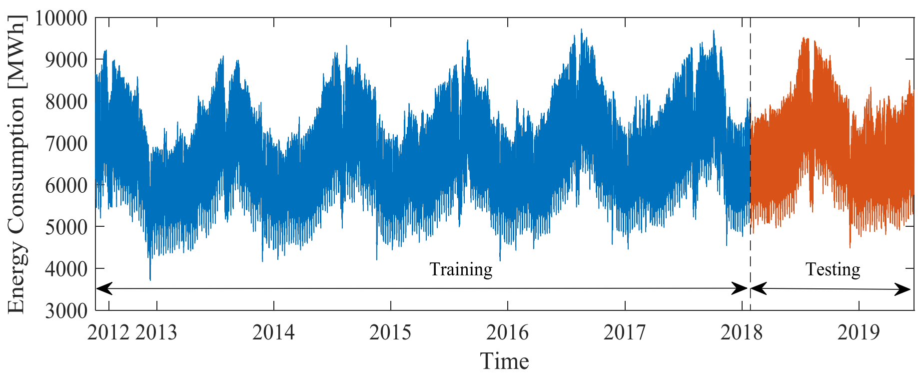


Fig. 6 Training and testing split of the dataset

As it can be observed in Fig. 6, the dataset is split into a training dataset and a testing dataset. The last 511 days from the 7-year period, which represent 20% of the dataset, are used to test the model's performance. From the remaining dataset, 30% of data are randomly selected and used for the validation process, while the other 70% of data are used for training.

B. Model performance analysis

This study focuses on evaluating the model described in Section III, by comparing four different architectures of neural networks (MLP, CNN, LSTM and GRU). The influence of number of layers that process the load data (included in the grey block in Fig. 5) is also assessed, as presented in Table 1. The output of these layers is concatenated with the external variables and fed into fully-connected layers. For each type of NN studied, two fully-connected layers are used after concatenating the first part of the model with the external variables. The model implementation was done in Python using TensorFlow and Keras libraries. For training, Nadam optimizer was applied.

Table 1

Evaluation metrics for the studied models

NN type	No. of layers	MAE [MWh]	MAPE [%]	MSE [-]	RMSE [MWh]
MLP	1	128.5	1.85	29384	179.6
	2	114.8	1.66	23945	161.8
	3	117.5	1.7	24259	166.1
CNN	1	123.6	1.78	27124	173
	2	112.5	1.63	23025	159.9
	3	122.5	1.77	26423	170.9
LSTM	1	134.3	1.94	30570	188.8
	2	132.7	1.92	33167	192.5
	3	164.2	2.35	68705	251.8
GRU	1	135.2	1.94	36118	187.3
	2	128.8	1.86	33654	181.6
	3	153.6	2.2	51992	218.6

As it can be observed, the CNN architecture with 2 convolutional and pooling layers obtained the best performance by means of all evaluation metrics considered in this study, achieving a MAPE value of 1.63% and a MAE value of 112.5 MWh, which indicate a great potential of the model. The MLP also shows good accuracy, with a MAPE of 1.66% when two layers are used.

For the analyzed dataset, it can be observed that GRU showed better results than LSTM networks. The lowest MAPE values are 1.86% for GRU and 1.92% for LSTM, respectively.

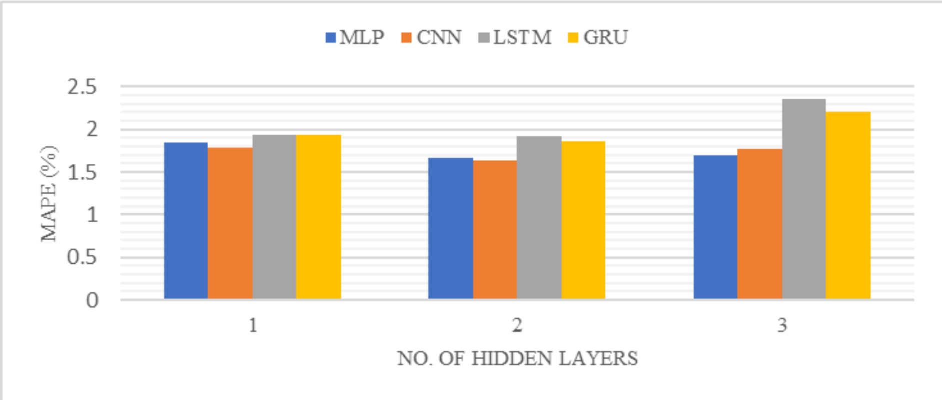


Fig. 7 The NN techniques assessment results

As depicted in Fig. 7, for every type on neural network, the best performance is achieved when two layers are used. When only one layer is used, the neural network is too simple and cannot extract the complex features from the input data. Using too many layers generally leads to overfitting, the neural network obtaining good accuracy for the training set, but with poor performance in the testing phase.

5. Conclusions

In this paper multiple neural networks-based methods were investigated in solving the short-term load forecasting for the Romanian power system. Four types of neural networks, namely the MLP, CNN, LSTM and GRU, are used within the proposed model and are evaluated based on MAPE, MAE, MSE and RMSE metrics. Among the tested architectures, the convolutional neural network obtained the best results by means of all evaluation indexes. The proposed model is able to integrate each of the discussed neural networks without additional processing, thus providing the necessary framework for comparison studies regarding the performance of the different neural networks in the load forecasting problem.

Future developments of forecasting methodologies may involve model's hybridization by combining multiple neural network architectures. Also, ensemble learning techniques are considered, as aggregating results produced by different models may reduce large errors and lead to a more consistent performance.

R E F E R E N C E S

[1] *Lopes JAP, Madureira AG, Matos M, et al.* The future of power systems: Challenges, trends, and upcoming paradigms. WIREs Energy Environ, 2019.

[2] *H. Hahn, S. Meyer-Nieberg and S. Pickl*, "Electric load forecasting methods: Tools for decision making," European Journal of Operational Research, **vol. 199**, no. 3, pp. 902-907, 2009.

[3] *L. Zhang and L. Xu*, "Forecasting of fluctuations and turning points of power demand in China based on the maximum entropy method and ARMA model," in 5th International Conference on Critical Infrastructure (CRIS), Beijing, China, 2010.

[4] *S. Noureen, S. Atique, V. Roy and S. Bayne*, "Analysis and application of seasonal ARIMA model in Energy Demand Forecasting: A case study of small scale agricultural load," in IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS), Dallas, TX, USA, 2019.

[5] *K. Goswami, A. Ganguly and A. K. Sil*, "Day Ahead Forecasting and Peak Load Management using Multivariate Auto Regression Technique," in IEEE Applied Signal Processing Conference (ASPCON), Kolkata, India, 2018.

[6] *P. Ji, D. Xiong, P. Wang and J. Chen*, "A Study on Exponential Smoothing Model for Load Forecasting," in Asia-Pacific Power and Energy Engineering Conference, Shanghai, China, 2012.

[7] *P. Mukhopadhyay, G. Mitra, S. Banerjee and G. Mukherjee*, "Electricity load forecasting using fuzzy logic: Short term load forecasting factoring weather parameter," in 7th International Conference on Power Systems (ICPS), Pune, India, 2017.

[8] *C. Feng and J. Zhang*, "Assessment of aggregation strategies for machine-learning based short-term load forecasting," Electric Power Systems Research, **vol. 184**, 2020.

[9] *Y.S. Park, S. Lek*, Artificial Neural Networks: Multilayer Perceptron for Ecological Modeling, Elsevier, 2016.

[10] *Y. Lecun, Y. Bengio and G. Hinton*, "Deep learning," Nature, **vol. 521**, pp. 435-444, 2015.

[11] *X. Dong, L. Qian and L. Huang*, "Short-term load forecasting in smart grid: A combined CNN and K-means clustering approach," in 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju, 2017.

[12] *S. Hochreiter and J. Schmidhuber*, "Long Short-Term Memory," Neural Computation, **vol. 9**, no. 8, pp. 1735-1780, 1997.

[13] *K. Cho, B. van Merriënboer, C. Gulcehre, F. Bougares and Y. Bengio*, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," 2014.

[14] *J. Chung, C. Gulcehre, K. Cho and Y. Bengio*, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling.," 2014.

[15] *G. Petneházi*, Recurrent Neural Networks for Time Series Forecasting, 2018.

[16] *A. Elvers, M. Voß and S. Albayrak*, "Short-Term Probabilistic Load Forecasting at Low Aggregation Levels Using Convolutional Neural Networks," in 2019 IEEE Milan PowerTech, Milan, 2019.

[17] *K. Amarasinghe, D. L. Marino and M. Manic*, "Deep neural networks for energy load forecasting," in 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh, 2017.

[18] *Shcherbakov et. al*, "A survey of forecast error measures." World Applied Sciences Journal. **vol. 24**. pp. 171-176, 2013.

[19] *Transelectrica*: available at: http://www.transelectrica.ro/en/web/tel/consum, accessed on 15/09/2020.

[20] *wunderground*: available at*:* https://www.wunderground.com/, accessed on 15/09/2020.