Journal of Artificial Intelligence and Data Science (JAIDA)
e-ISSN: 2791-8335
https://dergipark.org.tr/pub/jaida

Vol. 3, No. 2, pp. 116-123, 2023
Research Article
Received: 09/10/2023 Accepted: 26/11/2023

Time Series Forecasting of Solar Energy Production Data Using LSTM

Kadriye Filiz BALBAL[1,*], Özge ÇELİK [2] , Sebahattin İKİKARDEŞ [3]

Abstract

The fact that countries have increased the use of renewable energy resources in order to meet increasing energy demands has brought to light the fact that the components and energy production amounts of the solar energy systems to be installed must be estimated accurately. With the benefits provided by developing technology, forecasting calculations of these variable natural energy resources have become much more economical by using machine learning methods. In this context, this study proposes a deep learning-based methodology that includes LSTM-based tuned models for PV power forecasting, with univariate time series estimation of the amount of power obtained from a solar energy system integrated on a factory roof. When the created models are compared, the results show that the model approach named LSTM13 provides the most accurate prediction performance with the lowest RMSE metric value of 0.1470 among the other proposed models.

Key Words: Deep Learning, LSTM, Machine Learning, Renewable Energy, Solar Power Systems.

1. Introduction

Developing technology and increasing population cause the world's energy needs to increase day by day. Today, 70% of this energy need is met by fossil fuels. "Research since 1970 shows that the use of fossil fuels increases the emission of greenhouse gasses such as carbon dioxide (CO_2) and methane (CH_4) to the atmosphere. This increase thickens the ozone layer and causes global warming. To minimize the future effects of these phenomena, it is necessary to replace the use of fossil fuels with environmentally friendly, clean, and renewable energy sources as much as possible." [1]. Moreover, the fact that countries no longer want to depend on foreign energy has accelerated their orientation toward renewable clean energy sources.

The use of renewable energy sources, solar and wind energy, has been increasing in recent years. While the amount of electricity generated from solar energy in the world was 65,631 GWh in 2011, this value increased to 192 GWh in 2019 [2]. However, since solar energy is an energy source with daily, monthly, or annual inconsistencies, it is called an unstable renewable energy source. Therefore, unstable renewable resources differ in the amount of production compared with traditional energy sources. This situation has led to an increase in the importance of artificial intelligence in determining the potential of a renewable energy source with an unstable structure such as solar energy in solar energy systems, unlike traditional energy sources, and in predicting future characteristics.

The most important prerequisite for having a reliable and economic power source is determining the appropriate system design for the establishment and operation of new-generation renewable energy systems. However, for this, the capacity of these systems, which produce electricity from variable renewable energy sources, must be optimized and each component must be appropriately sized. The right choice of sizing techniques for these power systems, which are in relationship with smart network structures, can not only meet the required

*Corresponding author

KADRIYE FILIZ BALBAL; Dokuz Eylül University, Faculty of Science, Computer Science Department, Türkiye; e-mail: kadriyefiliz.balbal@deu.edu.tr ; 0000-0002-7215-9964

ÖZGE ÇELİK; Balıkesir University, The Graduate School of Natural and Applied Sciences, Mathematics Department, Türkiye; e-mail: ozgecelik2387@gmail.com ; 0009-0003-9565-8242

SEBAHATTIN İKİKARDEŞ; Balıkesir University, Faculty of Science, Mathematics Department, Türkiye; e-mail: skardes@balikesir.edu.tr ; 0000-0003-2924-5397

energy production amount but also contribute positively to the amount of energy that can be transferred to the connected grid with maximum efficiency.

Various algorithms are used to arrive at accurate estimates of renewable energy production. These algorithms are generally divided into four categories [3]: physical methods [4], [5], statistical models [6], artificial intelligence techniques [7], [8], and hybrid methods [9].

Renewable energy systems produce electricity using only solar energy, only wind energy, or both solar and wind energy. Many studies have been conducted to estimate the amount of electricity that can be obtained from these energy sources that change hourly, daily, monthly, or annually. The contribution of solar energy, which is considered an unstable renewable energy, to electricity production can be calculated with the predictions obtained by these algorithms using the climate regions, seasons, or meteorological data of the countries. There are many studies in the literature on photovoltaic (PV) power estimation using Deep Learning (DL) and Machine Learning (ML) approaches. Chaouachi et. al. [8] built the Recurrent Neural Network (RNN) ensemble model to predict short-term solar energy production. Tokgöz and Ünal [10] used RNN, LSTM, and GRU to estimate Turkey's electricity load in their study. Nam et al. [11] used SARIMA, GRU, LSTM, and MLR forecasting models to guide Korea's sustainable energy policy. Gao et al. [12] introduced the multi-strategy CNN-LSTM model to find the hourly solar irradiance forecast. Bişkin and Çiftci [13] used LSTM and GRU networks for forecasting Türkiye’s electrical energy consumption one hour ahead and three hours ahead. Ünlü [14] found that LSTM showed the best performance in prediction of Türkiye’s daily electricity load one hour ahead.

The aim of this study is to forecast solar power production using deep learning algorithm. For this purpose, a real photovoltaic plant production dataset is used in this study. The main objectives of this study are summarized as follows:

- First, this study provides Long Short-Term Memory (LSTM) models that forecast one hour ahead of real PV power production and comparative analysis.
- This study seeks to understand the relationship between time series input and forecasting accuracy.
- This study was conducted to evaluate the performance of LSTM models with univariate time series.

This paper is presented with its main headings as follows; The data used, the models created, and the accuracy measurements used are given in the Methodology section; forecasting results are discussed in the Results and Discussion section using graphics.

2. Methodology

Türkiye is the developing country that has plenty of solar power potential to produce own electricity independently. In order to ensure the widespread use of solar energy systems in the country, accurate estimates of the energy to be obtained must be made. In line with these ideals, the data collected, the pre-processing of the data, the introduction of the models planned to make consistent predictions and how they were created, and finally the metrics used are explained in this section. Figure 1 summarizes the framework of this study.

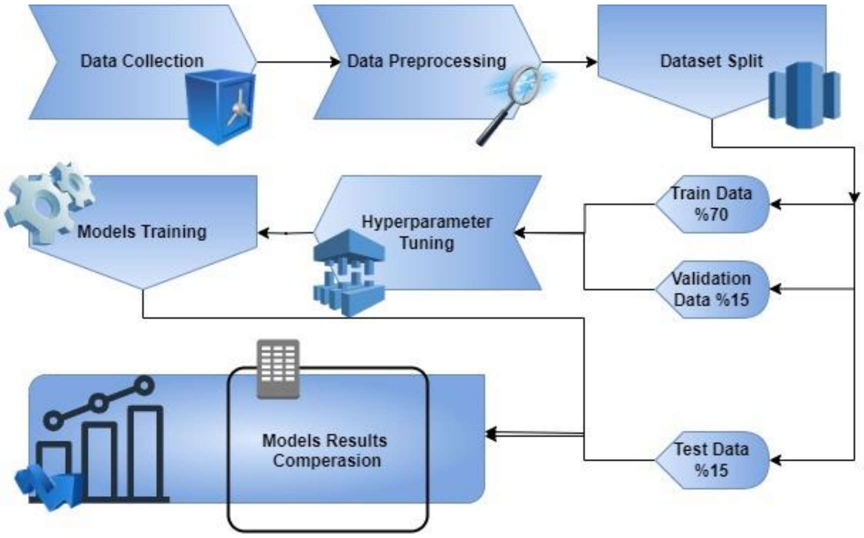


Figure 1: *Methodology of the study*

All calculations and visualizations were performed on the Colab (Collaborative) Platform [15] provided by Google. The GPU provided by the Colab Platform was also utilized to accelerate model training. For model training, TensorFlow v2.14.0 [16], Pandas v2.1.1 [17], and Keras v2.4.3 [18] and for visualizing Mathplotlib [19] libraries were used.

2.1 Data collection and preprocessing

Time series are data sets that can have a certain pattern, consisting of data with fixed time intervals. It enables the development of different forecasting approaches, depending on whether they are time-dependent or seasonal, to determine the underlying model in the data. If only one variable varies over time in the dataset, the time-series is named as “Univariate Time Series”.

The selection and collection of input data are the primary factors that affect the consistency of the prediction. Researchers have used solar radiation amount, actual power plant production, and some meteorological data, especially in studies conducted for solar energy production estimates. In this study, inverter data of a roof-integrated solar energy system, which has been operating for two years, was used as input. Raw data are univariate time series data that are collected hourly production kW values that are not used before with 16464 lines. The factory whose data were used is located in Izmir, a city on the west coast of Türkiye. The city has a Mediterranean climate, with hot and dry summers and warm and rainy winters. With data visualization processes, the seasonality of the data can be clearly seen in Figure 2.

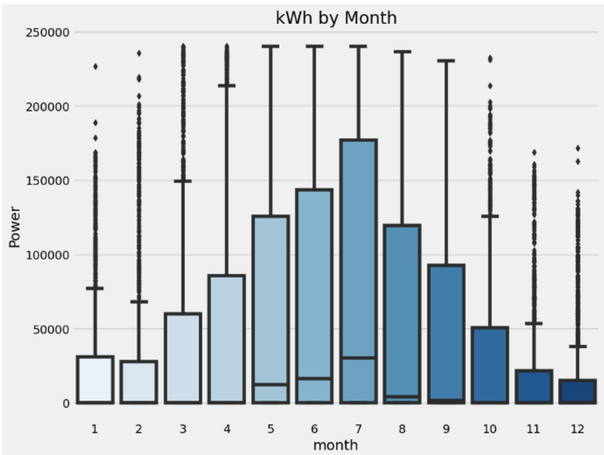


Figure 2: *Annual solar power production*

The seasonal difference in winter and summer production can be shown with one-week production amount graphs for February in Figure 3 and August in Figure 4.

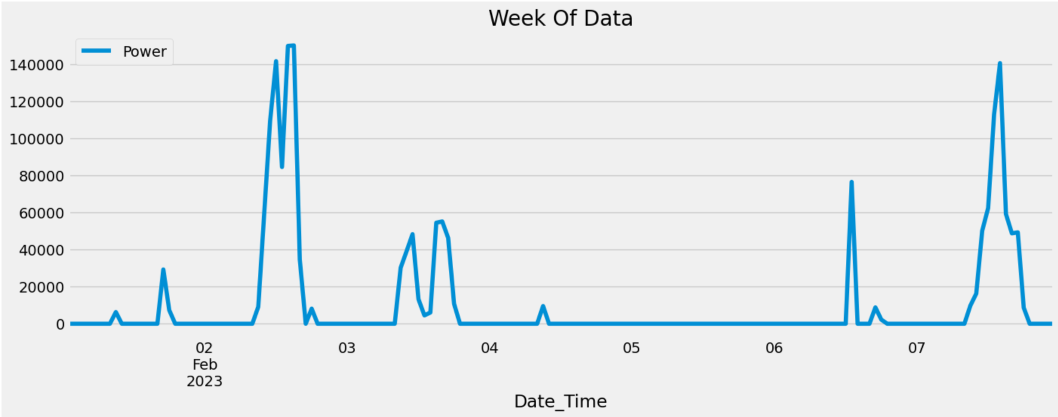


Figure 3: *Winter week of data*

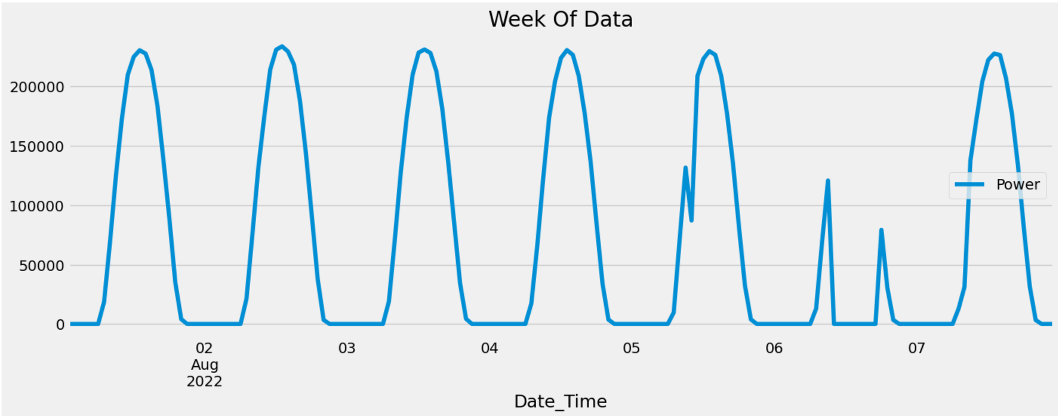


Figure 4: *Summer week of data*

For the DL method examined with the obtained time series dataset, it is necessary to configure a new dataset with transformed input and output variables. For this, after some Exploratory Data Analysis (EDA) were performed, sliding windows of size 7 were used and the dataset was transformed.

The newly transformed dataset was divided into training, validation, and test sets in the following proportions: 70%, 15%, and 15%.

2.2 Model development

Many machine learning and deep learning methods are used to forecast solar power production. For this study, Long Short-Term Memory (LSTM), which was proposed by Hochreiter & Schmidhuber [20] in 1997 and is known as a new generation Recurrent Neural Network (RNN) used in deep learning, was chosen because of its optimized architecture to easily capture patterns in time series data. Using a network such as LSTM provides benefits such as the ability to learn and remember over long sequences.

Table 1: *Model Structure*

	Hyperparameter	**Value**
	Hidden Layers	LSTM, Dense, Flatten
	Units	16, 32, 64, 128, 256
	Learning rates	0.01, 0.001, 0.0001, 0.00001
MODELS	Activation function	ReLu, Linear
	Optimizer	Adam
	Loss	Mean Squared Error
	Metric	Root Mean Squared Error
	Epochs	10, 50, 100

Table 1 demonstrates the construction of an LSTM network model for time series forecasting using Keras [18]. All models were constructed using different LSTM layers consisting of 16, 32, 64, 128, and 256 neurons. The activation function is the Rectified Linear Unit (ReLU) function used in the hidden layer, and the linear function is used in the output layer. The dropout probability was set to 0.2. The training epochs was chosen as 10, 50 and 100. To calculate the loss and weight in the compiler network, it is necessary to determine the loss function and optimization algorithm. For this purpose, in this study, Mean Squared Error (MSE) was selected as the loss function and Adaptive Moment Optimization (Adam) was selected as the optimization algorithm with learning rates as 0.01, 0.001, 0.0001, and 0.00001.

In this study, 30 different single and multilayer LSTM models were developed. Since the performances of 14 of the models created were higher than those of the others, only these fourteen models were compared in the Results and Discussion section.

3. Results and Discussion

The main purpose of this study is to investigate the consistency of LSTM models in univariate time series forecasting. For this purpose, different LSTM-based models have been developed using real PV power values obtained from a roof-integrated solar energy system for one hour ahead forecasting. Forecasting models use the sliding window method, which uses seven hours of data to learn and forecast electricity data for the next one hour. This sliding window technique has the advantage of increasing the time-dependent forecast performance of the developed models [11].

Models are trained using the 70% training set of the divided data after some EDA calculations and hyperparameter tuning transformations. After each training iteration on the 15% validation set, training errors and loss values are calculated using the MSE and RMSE metrics. The model with the lowest error in the validation set is selected for the final performance evaluation on the 15% test dataset. Table 2 lists the models' learning rate, epochs, loss, and RMSE values, respectively.

Table 2: *Statistical Test and Results of Models*

Model	Learning Rate	Epochs	Loss	RMSE
LSTM1	0.001	100	0.0323	0.1797
LSTM2	**0.001**	**100**	**0.0228**	**0.1510**
LSTM3	0.01	50	0.0303	0.1741
LSTM4	0.001	50	0.0371	0.1927
LSTM5	0.0001	50	0.0436	0.2088
LSTM6	0.001	50	0.0392	0.1981
LSTM7	0.00001	50	0.0525	0.2291
LSTM8	0.0001	10	0.0666	0.2580
LSTM9	0.001	50	0.0416	0.2039
LSTM10	**0.01**	**50**	**0.0250**	**0.1580**
LSTM11	0.01	50	0.0272	0.1650
LSTM12	**0.01**	**50**	**0.0245**	**0.1564**
LSTM13	**0.01**	**50**	**0.0216**	**0.1470**
LSTM14	0.01	50	0.0298	0.1726

All models use MSE as the loss function. Comparing the test data's loss and RMSE values in Figure 5, it is clearly noticeable how all loss and error values vary between models.

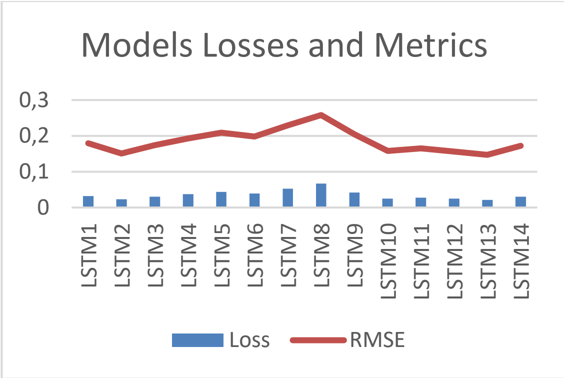
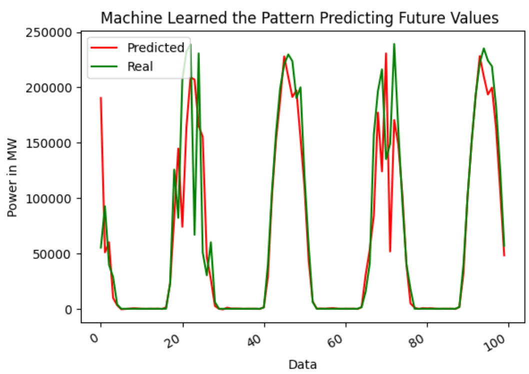


Figure 5: *Models Loss and Metrics changes*

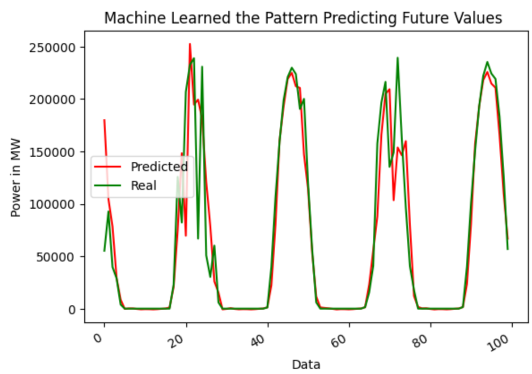
In Table 2, four models with higher performance are indicated in bold that are named LSTM2, LSTM10, LSTM12, and LSTM13. LSTM2 has hidden layers with 16, 32, 64, and 128 neurons; LSTM10 has hidden layers with two-64-neurons; LSTM12 has hidden layers with two-128-neurons; and LSTM13 has hidden layers with two-256-neurons; they were developed with two dense layers. These models were found to be the best four of fourteen proposed models.

Table 3 shows the PV power generation test set against the prediction results obtained under the conditions of using a two-year univariate time-series dataset and a single deep learning model. The results correspond to a 4-day period where actual and predicted PV power amounts perform scalability analysis.

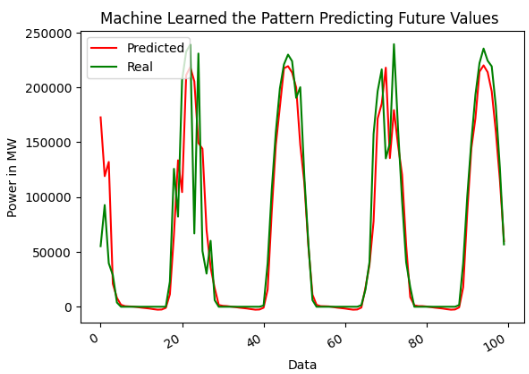
Table 3: *Visualizing the first 100 values of the test set by selected LSTM models*



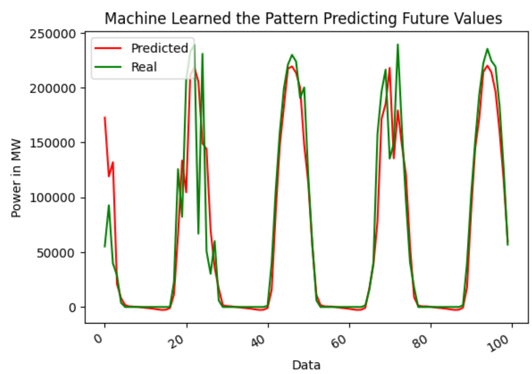
LSTM2



LSTM10



LSTM12



LSTM13

It is important to note that the choice of an LSTM model that will work using a univariate dataset will depend primarily on the characteristics of the time series data and the different requirements of the current forecasting problem [21]. Each type of univariate LSTM model has its own strengths and limitations. In particular, when forecasting horizons are considered, each model developed has different accuracy values for different forecasting horizons. The accuracy values of the model decrease as the forecasting horizons progress.

Considering the complexity of the preliminary calculations required for the widespread use of solar panels and the unstable nature of the renewable energy source to be used, research has shown many times that deep learning methods and hybrid models are the most appropriate method to eliminate these disadvantages. CEEMDAN and multi-strategy CNN-LSTM models RMSE performance improves 9.97% to 73.06% in [12], ConvLSTM1D is introduced by 0.0264 RMSE for 15 min ahead, 0.0273 RMSE for 30 min ahead and 0.0275 RMSE for 1h ahead in [22], averaged RMSE for Auto-LSTM is found 0.0713 in [23], plenty of GRU-LSTM models introduced in [24] with higher accuracies.

4. Conclusion

This article aims to develop deep learning models for forecasting on real PV output time series data. The studies argue that data-based forecasting increases the efficiency of renewable energies compared with other methods, and therefore it is emphasized that deep learning techniques should be investigated in solar energy forecasting [21]. In other words, the research underlines that each application is justified and that more efficient calculations can be made with significant improvements in the future. According to [21], a perfect estimate can save approximately 15%, while using an LSTM or a simple MLP can save approximately 12%. Therefore, the LSTM network, which is well-known for handling long sequence data effectively, is chosen to invest in accuracy while using univariate time series data. Although, it is unclear which structures could lead to the best forecasting performance. Several variations have been made to clarify the model structures. Different hidden layers, learning rates, epochs, and window sizes are used to build the model. However, using univariate time series data as a unique input has some limitations. To overcome such negative effects, using other deep learning methods or adding some independent varieties as inputs would be efficient.

For future works, studies using other deep learning methods such as GRU, CNN, and a built hybrid of these methods can be used to develop efficient power strategies to help PV power plant developers and countries to lead in dynamic energy markets.

Declaration of Interest

The authors declare that there are no conflicts of interest.

Author Contributions

Kadriye Filiz Balbal: Conceptualization, Methodology, Formal Analysis, Writing - Review & Editing, Supervision.
Özge Çelik: Conceptualization, Methodology, Data Generation, Writing - Review & Editing, Visualization.
Sebahattin İkikardeş: Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision.
All authors reviewed the manuscript.

References

[1] Z. Şen, "Solar energy in progress and future research trends," *Progress in Energy and Combustion Science 30,* p. 367–416, 2004.

[2] IRENA, "Renewable Energy Capacity Statistics 2023," International Renewable Energy Agency(IRENA), 2023.

[3] H. Wang, Z. Lei, X. Zhang, B. Zhou and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management,* no. 198, p. 111799, 2019.

[4] H. Bulut and O. Büyükalaca, "Simple model for the generation of daily global solar-radiation data in Turkey," *Applied Energy 84,* p. 477–491, 2007.

[5] M. Çınaroğlu and M. Nalbantoğlu, "Şebekeye Bağlı Üç Adet Fotovoltaik Enerji Santralinin PVsyst Programı ile Analizi; Kilis Örneği," *El-Cezerî Fen ve Mühendislik Dergisi,* vol. 8, no. 2, pp. 675-687, 2021.

[6] İ. T. Toğrul and H. Toğrul, "Global solar radiation over Turkey: comparison of predicted and measured data," *Renewable Energy ,* no. 25, p. 55–67, 2002.

[7] F. O. Hocaoğlu, Ö. N. Gerek and M. Kurban, "Hourly solar radiation forecasting using optimal coefficient 2-D linear filters and feed-forward neural networks," *Solar Energy,* vol. 82, pp. 714-726, 2008.

[8] A. Chaouachi, R. M. Kamel and K. Nagasaka, "Neural Network Ensemble-Based Solar Power Generation Short-Term Forecasting," *Journal of Advanced Computational Intelligence and Intelligent Informatics,* pp. 69-75, 2010.

[9] H. Demolli, *Makine Öğrenmesi Teknikleri Kullanılarak Yenilenebilir Enerji Güç Tahmini ve Optimum Hibrit güç sistemi Tasarımı(Doktora tezi),* Yükseköğretim Kurulu Ulusal Tez Merkezi veri tabanından erişildi (Tez No. 623617), 2020.

[10] A. Tokgöz and G. Ünal, "Türkiye Elektrik Tüketimi Tahmini ˙Için RNN Tabanlı Zaman Serisi Yaklaşımı," IEEE, 2018.

[11] K. Nam, S. Hwangbo and C. Yoo, "A deep learning-based forecasting model for renewable energy scenarios to guide sustainable energy policy: A case study of Korea," *Renewable and Sustainable Energy Reviews,* vol. 109725, no. 122, 2020.

[12] B. Gao, . X. Huang, J. Shi, . Y. Tai and J. Zhang, "Hourly Forecasting of Solar Irradiance based on CEEMDAN and Multi-strategy CNN-LSTM neural networks," *Renewable Energy, pp.* 1665-1683, 2020.

[13] O. T. Bişkin and A. Çifci, "Forecasting of Turkey's Electrical Energy Consumption using LSTM and GRU Networks," *BSEU Journal of Science,* vol. 8, no. 2, pp. 656-667, 2021.

[14] K. D. Ünlü, "A Data-Driven Model to Forecast Multi-Step Ahead Time Series of Turkish Daily Electricity Load," *Electronics* , vol. 11, no. 1524, 2022.

[15] "Colaboratory," Google, [Online]. Available: https://colab.research.google.com/.

[16] "TensorFlow v2.14.0," [Online]. Available: https://www.tensorflow.org/.

[17] "Pandas 2.1.1," [Online]. Available: https://pandas.pydata.org/.

[18] "Keras," [Online]. Available: https://keras.io/.

[19] "Mathplotlib," [Online]. Available: https://matplotlib.org/.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput. 9 (8)* , p. 1735–1780, 1997.

[21] M. Ajith and . M. Martínez-Ramón, "Deep learning algorithms for very short term solar irradiance forecasting: A survey," *Renewable and Sustainable Energy Reviews,* vol. 182, p. 113362, 2023.

[22] S. Cantillo-Luna, R. Moreno-Chuquen, D. Celeita and G. Anders, "Deep and Machine Learning Models to Forecast Photovoltaic Power Generation," *Energies,* vol. 16, no. 4097, 2023.

[23] A. Gensler, . J. Henze, . B. Sick and N. Raabe, "Deep Learning for Solar Power Forecasting – An Approach Using Autoencoder and LSTM Neural Networks," in *IEEE International Conference on Systems, Man, and Cybernetics • SMC 2016* , Budapest, Hungary, 2016.

[24] M. Neshat, M. M. Nezhad, S. Mirjalili, D. A. Garcia, E. Dahlquist and A. H. Gandomi, "Short-term solar radiation forecasting using hybrid deep residual learning and gated LSTM recurrent network with differential covariance matrix adaptation evolution strategy," *Energy,* vol. 278, no. 127701, 2023.