SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
http://saucis.sakarya.edu.tr/
e-ISSN: 2636-8129
Publisher: Sakarya University
Vol. 7, No. 3, 2024, 325-337
DOI: 10.35377/saucis...1503018

RESEARCH ARTICLE

Improving Deep Learning Forecasting Model Based on LSTM for Türkiye's Hydro-Electricity Generation

Mehmet Bulut

Electricity Generation Co. Inc., General Management Consultant, Ankara, Türkiye

Corresponding author:
Electricity Generation Co. Inc.,
General Management Consultant,
Ankara, Türkiye,
mehmetbulut06@gmail.com

Article History:
Received: 20.06.2024
Accepted: 06.09.2024
Published Online: 30.10.2024

ABSTRACT
Electricity production in hydraulic power plants depends on the amount of water coming into the basin. This varies depending on precipitation such as snow and rain during the year, but when looking at the years, production is shaped according to the years when meteorological data are similar to each other. LSTM (Long Short-Term Memory) plays an important role in hydropower forecasting, as it is a special artificial neural network designed to model complex relationships on time series data, which is affected by various meteorological factors such as precipitation, temperature, and hydrological data such as water level, such as hydroelectric power production. Therefore, in this study, a forecast system based on the LSTM network model which is one of the deep learning methods was proposed for monthly hydropower-based electricity production forecast in Türkiye. The developed deep learning-based hydropower forecast model provides future production planning based on time series based on actual hydropower production data. Using real production data and LSTM learning models of different structures, monthly hydraulic electricity production forecasts for the next year were made and the models' performances were examined. As a result of this study, RMSE 32.4245 and MAPE 16.03% values and 200-layer LSTM model trained with 12-year data with 144 monthly data points containing hydroelectric generation information was obtained as the best model, and the performance values of the model showed that it was the correct forecasting model. The overall efficiency parameters of the found LSTM model were checked with NSE 0.5398 and KGE 0.8413 values. The performance of the model was found to be a high-accuracy model within acceptable limits and with a correlation value of R2 0.9035 to be very close to reality. The results obtained from this study have shown that deep learning models developed based on many years of production data give successful results in hydroelectric production prediction and can be used as a basis for electricity production planning.

Keywords: Hydroelectric power, Electricity production forecasting, deep learning, Long short term memory

1. Introduction

Research shows that renewable energy sources have the capacity to meet two-thirds of the world's total energy demand. They can also contribute to the drastic reduction of greenhouse gas emissions needed to keep the average global surface temperature rise below 2°C by 2050 [1]. Examples of major renewable energy sources are solar, wind, hydrogen, hydroelectric, wave and geothermal. The biggest environmental advantages of renewable energy sources are that they are renewable due to their continued existence in nature, they do not harm nature by less carbon emissions, and they are clean and sustainable energy sources [2]. Sunlight, wind speed and precipitation parameters, which form the basis of renewable energy sources, change seasonally

Discontinuous and intermittent renewable energy sources depend on meteorological weather conditions and parameters such as temperature and rainfall [3]. Estimating the amount of electricity to be obtained from hydroelectric power plants is of great importance in terms of planning production according to resources and ensuring the continuity of production [4]. The fact that renewable energy resources vary throughout the year and even during the day and depend on seasonal conditions from year to year makes it important to estimate production in the electricity grid to meet the supply. With the development of deep learning algorithms, it has achieved successful results in many areas, especially computer learning and prediction studies. In applications, deep learning models are created based on time series data, many of which are applied to real life, such as precipitation-flow modeling [5]. Deep learning network structures allow learning the complex relationships between the input and output sets of the learning structure and the complex relationships between the data [6]. Today, Long Short-Term Memory (LSTM) networks attract widespread attention and have many practical applications for time-series-based forecasting systems [7]. Wang et al. examined the methods used in estimating production based on renewable energy sources

Cite as: M. Bulut, "Improving Deep Learning Forecasting Model Based on LSTM for Türkiye's Hydro-Electricity Generation," *Sakarya University Journal of Computer and Information Sciences*, vol.7, no. 3, pp. 325-337, 2024. Doi: 10.35377/saucis…1503018

and the degree of accuracy in the predictions of these methods [8]. Cheng et al. focused on estimating the power demand needed in the network to accurately meet the demand in a network [9]. Other deep learning-based studies include predicting the amount of energy demanded monthly 3 days in advance [10], estimating the short-term electricity demand curve [11], design a new specific power demand prediction algorithm based on LSTM Deep Learning method with respect to end-user power demand patterns[12], using meteorological data of the last 24 hours. Methods have been developed for estimating the demand amount based on [13], and for predicting the next 3-day production based on meteorological data in a wind power plant [14]. They propose a power demand forecasting model based on a neural network. Li et al. tried improving the prediction performance based on deep learning by fusing different production time data components such as daily, weekly and long time of a hydraulic power plant [15]. Del Real et al., on the other hand, used a mixed deep learning architecture consisting of a convolutional neural network (CNN) combined with an artificial neural network (ANN) to perform energy demand estimation [16].

Hydroelectric generation is one of Türkiye's leading renewable energy sources [17]. The amount of electricity production from hydroelectric power plants depends on the amount of water coming to the dam basin. Production values change on a monthly basis depending on the amount of precipitation during the year, and the closer the water level is to the design values, the more efficiency in production increases. Accurate prediction of electricity supply based on data on generation resources connected to an electricity system provides a solution to meet the demanded power [18]. However, methods other than resource-based prediction can be used by using historical electrical load flow data in the electrical system [19, 20]. To create a prediction model with Türkiye’s monthly electricity generation dataset, methods such as LSTM based on a deep learning algorithm were used using time series [21]. Renewable energy sources such as solar, wind and hydroelectricity mainly depend on local environmental and meteorological conditions such as temperature and precipitation-runoff rates. In many studies in the literature, local environmental and meteorological conditions such as temperature and precipitation-flow are used in the medium-term monthly forecasting of the amount of energy to be produced by Hydroelectric Power Plants (HPP). In this article, a deep learning model is applied/trained using time series to create a forecasting model using Turkey's monthly hydroelectric power plant production data set. Monthly Hydroelectric Generation Forecasting System was developed for forecasting hydroelectric production based on the LSTM Network-Based Deep Learning Model. Besides, as a result of global warming, increasing droughts and decreasing precipitation directly affect hydroelectric production, creating a decreasing effect on the generation of electricity. Hydroelectric energy production is based on converting the potential energy of water into kinetic energy, and the decrease in water resources due to drought will reduce the amount of energy produced by decreasing the water level in dams. Also, droughts make the flow regimes of rivers irregular, then Seasonal flow differences increase and water levels may drop unexpectedly. This negatively affects the efficiency of hydroelectric power plants [22]. In order to use water more effectively in drought conditions, water management strategies come to the fore, and it is necessary to ensure the sustainability of hydroelectric energy [23]. In this sense, the estimation studies of electricity production in the country's hydroelectric power plants play an important role in determining plans and strategies.

In the study, LSTM algorithm from artificial neural networks was used to predict the hydraulic electricity production of Türkiye. Compared to other artificial neural networks, LSTM's ability to hold data in memory increases its predictive power. By using this advantage, it is aimed to predict with high accuracy. As a result of the study, the amount of electricity that Türkiye will produce from hydraulic power plants is forecasted on a monthly basis annually, one year in advance, and production planning can be done more consistently in line with these estimated production values. When we look at the setup of this study, firstly, information about Turkey's hydroelectric production to be used in the study is given in the second section. Then, information about the creation of the model to be used in the model and the methods to be used to measure the performance of the models are shared. In the third section, the estimation results obtained from the models are given and the results are evaluated.

2. Material and Methods

2.1. Hydroelectricity Generation in Türkiye

Electricity generation from renewable energy sources in Türkiye shows an increasing trend every year in terms of hydroelectric and geothermal resources, especially wind and solar. Türkiye's installed electricity capacity reached 96,270 MW at the end of 2020. Türkiye's installed capacity has increased 3.5 times in 20 years. Renewable energy sources make up 51% of this total installed power, and the total installed power capacity of the plants based on renewable energy sources has reached 49,111 MW. By the end of 2020, a total of 305.4 billion kWh electricity was produced in Türkiye. As the share of resources in electricity generation, 45.90 billion kWh, which corresponds to 15%, was obtained from solar, wind and geothermal power plants, and 78.12 billion kWh, which corresponds to 25.6%, was obtained from hydraulic power plants. In 2020, the share of renewable energy sources in electricity generation was 42.4% [24, 25]. For Türkiye the hydraulic energy installed power as shown in Figure 1.

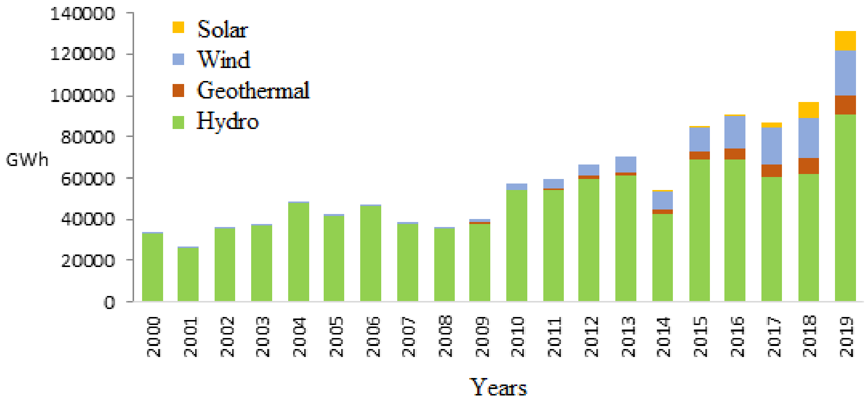


Figure 1. Electricity Generation from Various Renewable Sources Including Dam Type Hydro

Türkiye's hydroelectric production varies in some years, showing a continuous upward trend in installation and generation, consistent with total electricity generation [24]. Dataset used in this study for forecasting of hydro electrical generation is taken from the statistics section of the publicly available Load Dispatch Information System (YTBS) web portal [26]. Considering the 12-year production covering the hydroelectric production data of Türkiye in between 2007–2018, the minimum, maximum and average production values of the productions corresponding to the months between January and December are given in Figure 2. It is seen that the highest production are realized in May, and the lowest production was in October.

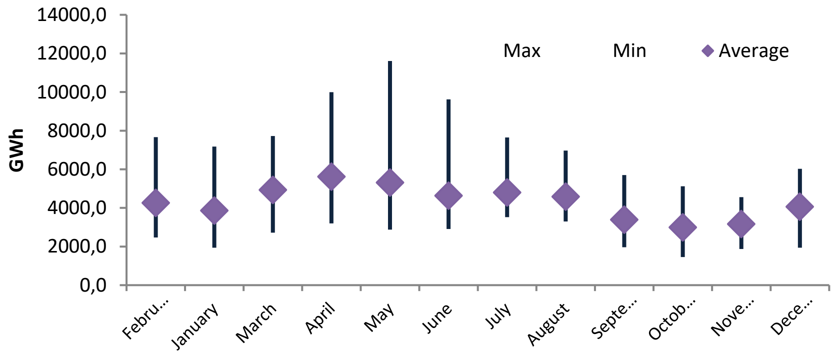


Figure 2. Monthly Maximum, Minimum and Average Generation Values of Hydroelectric Power Plants on a Long-Term Basis

2.2. Structure of Long Short-Term Memory (LSTM) Networks

Artificial Neural Networks (ANNs) represent learning algorithms developed based on the structure and functioning of human neural networks based machine learning. Deep learning (DL) networks, on the other hand, have been presented as a solution to many complex artificial intelligence (AI) problems that have existed for many years. In fact, DL models are described as deeper variants of linear or nonlinear multilayer artificial neural networks (ANNs).

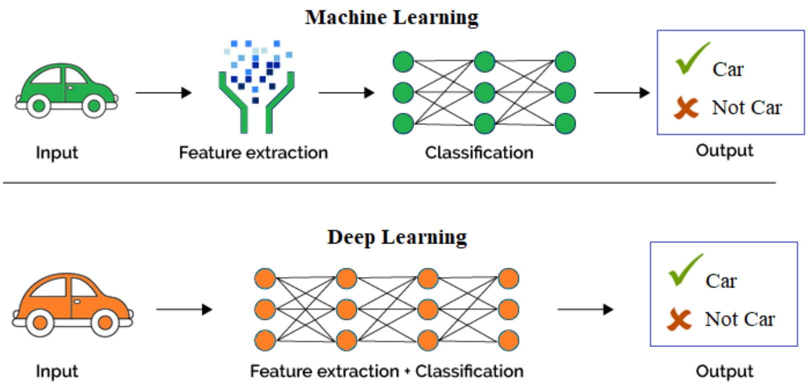


Figure 3. Difference Between Machine Learning and Deep Learning Approximation

Difference between machine learning and deep learning approximation is shown Figure 3. The ability of DL models to learn hierarchical properties from various data types makes them powerful in recognizing, regressing and solving semi-supervised and unsupervised problems [27]. ANNs can predict whether the image is a square or an equilateral rectangle. However, due to the fact that ANNs do not carry memory, they are insufficient in many cases where prior data information is important due to the nature of some problems. LSTMs can remember information for a long time. Many variations have been developed to solve the problem posed in a deep Recursive Neural Network (RNN). To solve this problem, LSTM uses gate units to decide what information to keep or remove from the previous state [28, 29]. A LSTM Network consists of three different Gates, they are Forget gate, Input gate and Output gate. Gates control the flow of information from memory to memory [30].

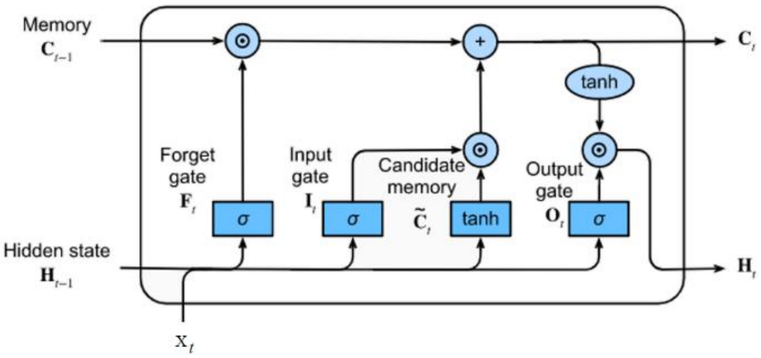


Figure 4. Gate Connections of an LSTM Cell [30]

The core component of LSTMs is a memory cell that can hold information over time controlled by gates. It can maintain its state over time, consisting of an open memory (also called a cell state vector) and gate units. Gate units regulate the flow of information entering and leaving the memory [31, 32]. The input gate is used to check whether the state in the current cell allows it to be overridden by external information as shown in Figure 4. The entrance gate in the LSTM cell structure used in the study is shown in Equation 1.

$$i_t = \sigma_k(W_i x_t + U_i h_{t-1} + b_i) \tag{1}$$

Here;
i_t : inlet gate vector,
σ_k : sigmoid function,
x_t: input vector,
W_i and U_i : matrices to parameter
b_i: bias vector.

The output vector decides whether to keep the state in the current cell, which will affect other cells, and is defined as in the Equation 2.

$$o_t = \sigma_k(W_o x_t + U_o h_{t-1} + b_o) \tag{2}$$

Another cell gate defined in the LSTM memory cell is the forgotten container, which enables the state of the LSTM to be reset and is defined as follows.

$$f_t = \sigma_k(W_f x_t + U_f h_{t-1} + b_f) \tag{3}$$

As a result, Equations 4 and 5 shows how the cell state and the output vector is revealed from the input gate, forget gate and output gate.

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh\left(W_f x_t + U_c h_{t-1} + b_c\right) \tag{4}$$

$$h_t = o_t \odot \sigma_h(c_t) \text{ or } h_t = o_t \odot tanh(c_t) \tag{5}$$

Here, ⊙ represents the Hadamart product, σ_c and σ_h represents the hyperbolic tangent functions.

2.3. Forecasting Time Series of Hydroelectric Generation Model

The hydroelectric generation estimation study consists of the following steps. In this section, the hydroelectric generation estimation framework made with LSTM is given in Figure 5.

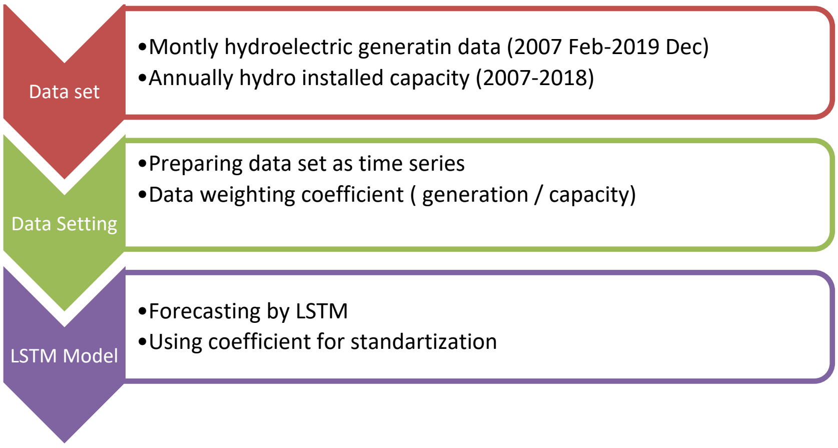


Figure 5. LSTM Hydropower Generation Forecasting Framework

It can be express the definition of the hydroelectric production estimation problem considered in this study as follows:
1- Obtaining a hydroelectric generation time series by using the monthly total electricity generation values realized in hydroelectric power plants for many years.
2- Using this obtained time series, forecasting annual hydroelectric generation monthly in the future.
3- Extract the internal feature to help better predict hydropower generation: Using the 12-year hydropower generation, the monthly correction coefficient is obtained by obtaining the average hydropower generation for each month of the year during January-December and standardizing them in a max-min range.
4- Historical observations of previous t - 1 time steps $Xt - 1 = \{X1, X2, \cdots, Xt - 1\}$ and internal features D(y): January;...;December, considering the hydroelectric generation forecast, is to learn a model that predicts electricity generation value. Time t, that is, in time step t of the electricity generation value X t {Xt | Estimating Xt <1, Ft}.

$$Xn(t)=Xn(1), X n(2),...., Xn(t-1),..., X n+1(1), X n+1(2),....,X n+1(t-1),...XN(1), XN(2),...,XN(t-1) \quad (5)$$

Here;

n=1..N, N=12 (January,…,December)

t = number of past years,

X(n+1)(t)=n in year t. hydroelectric production of the month.

The production of a dam hydroelectric power station is affected by keeping the water level too high or too low due to the change in water flow. Similarly, the generation of river or canal hydroelectric power plants depends on the water regime affected by precipitation. The most important parameter showing the generation characteristics of hydroelectric power plants is the capacity factor. The capacity factor shows how many hours the hydraulic plant produces in 8765 hours of the year and is used for performance comparison. Net plant capacity factor (CF) is referred to as full capacity plant can produce energy part of the total energy generated in a given period.

$$CF = ANP / (IP *365\ days*24\ hours/day) \quad (6)$$

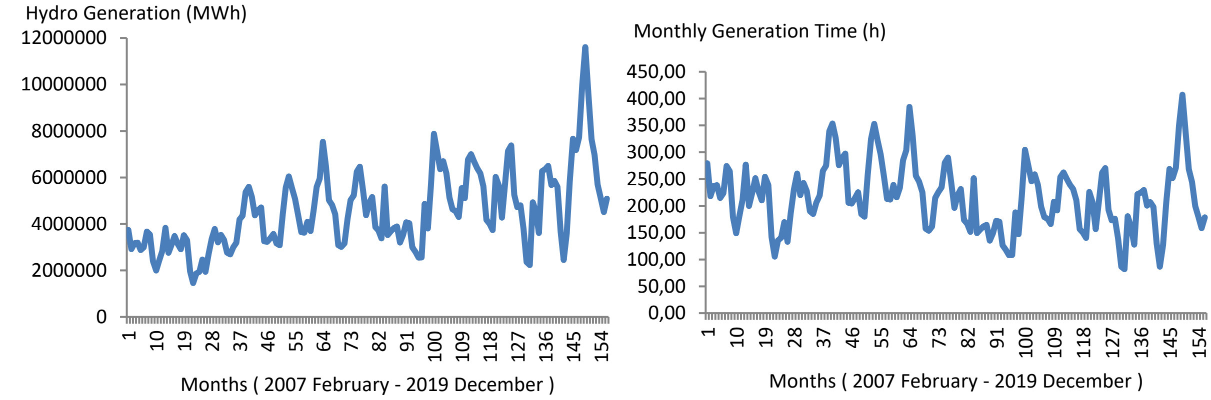
Here, ANP is annual electricity production and IP is installed capacity.

In hydroelectric power plants, the average capacity factor is around 40%. In other words, when the annual maximum operating time is taken into account as 8760 hours, hydroelectric power plants can only produce for 3500-4000 hours per year because the water coming to the reservoir is not always continuous. This means that hydroelectric power plants can only produce for 40% of the annual time period.

In this situation; Since the hydroelectric production data used in this study covers a period of 13 years, the installed capacity value that provides the production for each year is not the same due to each hydraulic installed capacity increase. To save the hydroelectric generation time data from this installed capacity value change, the time series based on monthly generation performance are obtained using the following formula (Figure 6. a, b).

$$GP = MEP/IP \quad (7)$$

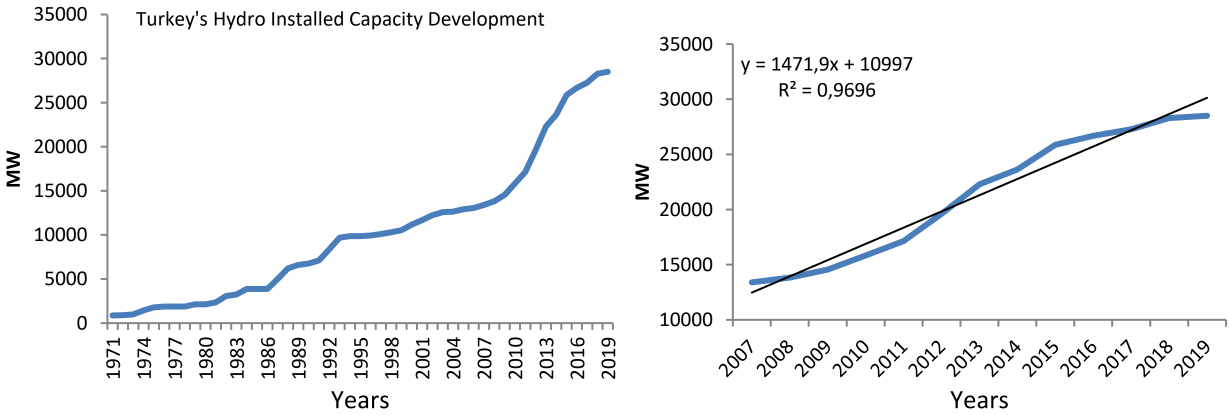
Here, GP is generation performance and MEP is monthly electricity production. Two separate data inputs explained below are fed into the LSTM Based hydroelectric system.



(a) (b)

Figure 6. a) Monthly Hydroelectric Production B) Data Set Standardization as Monthly Hydro Generation Time

One is monthly hydroelectric generation covering a 12-year time interval (Figure 6-a) and second is hydroelectric installed capacity value for these years as shown in Figure 7-a. The monthly estimated standardized hydroelectricity production values as generation duration (Figure 6b) are combined with the annual hydroelectricity estimation data, using the installed power values obtained from the regression estimation model in Figure 7-b.



(a) (b)

Figure 7. a) Graphic of Hydro Installed Capacity Data Set, B) Regression Model for Hydro Installed Capacity Data Set

Data preparation and calculation of final forecast production value;

1- Since the total installed capacity providing hydraulic electricity generation is different for each year, the annual hydraulic electricity production monthly per MW capacity must be standardized. To this end, monthly hydroelectric production (MWh) is proportioned to the total installed capacity value (MW) of the taken year. Then, monthly time curves are produced based on the number of hydroelectric production hours for each month.

2- Since the annual hydroelectricity are standardized using the annual installed capacity value, the estimated year's values are multiplied by the estimated year's total installed capacity value to obtain the monthly hydroelectricity production (MWh) estimated values

$$EMP = EPT * IP \tag{8}$$

Here, EMP is estimated monthly production as unit of MWh and EPT is estimated production time as unit of hours.

2.4. Performance Measurement and Evaluation of the Prediction Models

Hydroelectric production estimation is of great importance in terms of efficiency and reliability of energy systems, and the accuracy of these estimations allows more accurate decisions to be made in energy planning. Therefore, various criteria are important to evaluate the estimation performance. Since hydroelectric production estimation performance criteria are important for evaluating the accuracy and reliability of the model, choosing the right criterion is of great importance for developing the model and managing energy systems more effectively.

In this study, multi-criteria evaluation are performed using more than one criterion instead of a single criterion for performance evaluation. As Hydroelectric Production Estimation Performance Criteria in the study; Mean Absolute Percentage Error (MAPE), which is useful when making comparisons for data sets of different scales, Root Mean Squared Error (RMSE), which gives more weight to large errors and is used to evaluate the overall performance of the model, Coefficient of Determination (R), which shows how well the model explains the data, where the R-squared value varies between 0 and 1. As it approaches the value of 1, the explanatory power of the model increases. In addition, the Nash-Sutcliffe Efficiency Coefficient (NSE), which shows how well the model simulates the observed data, varies between -∞ and 1. The performance of the model increases as the value approaches 1, the KGE (Kling-Gupta Efficiency) method, which is a statistical measure frequently used in hydrological modeling and can evaluate how well the model fits the observed data, especially in time series data such as streamflow forecasting, is used to evaluate how well the model performs in terms of both its mean value and variance and shape.

Average absolute Percent Error (MAPE) and Root Mean Square Error (RMSE) metrics are used to select the best prediction model with the smallest estimation error by using the hydroelectric monthly generation data time series. RMSE and MAPE values were used to compare the prediction accuracy performances of different long short-term memory structures of models belonging to different time series and to measure the results obtained. RMSE is a quadratic scoring rule that also measures the mean magnitude of error and is defined as follows.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(v_{estimated} - v_{real})^2} \tag{9}$$

MAPE is a widely used statistical method that measures how close the forecast result made by a forecasting system is to the truth. It measures accuracy as a percentage. It is defined as follows

$$MAPE = \left(\frac{1}{N}\sum\frac{|v_{real}-v_{estimated}|}{|v_{real}|}\right) * 100 \tag{10}$$

And, to check the overall efficiency of the LSTM models which are used for the prediction of hydro generation, Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE) and coefficient of determination (R^2) metrics are used. NSE coefficient is used to measure the accuracy of many hydrological predictions, determining the relative magnitude of the persistent variance compared to the variance of the observation data. NSE is calculated by Equation 11,

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(X_{i=1}^{obs}-X_{i=1}^{cal}\right)^2}{\sum_{i=1}^{N}\left(X_{i=1}^{obs}-X^{avg}\right)^2} \tag{11}$$

Where; X_i^{obs} is the ith value of the observed monthly flows, X_i^{cal} is the ith value of the calculated monthly flows, X^{avr} is Average of observed monthly flows and N represents the total number of observations. NSE ranges from −∞ to 1. Here, NSE=1 proves that the method is physically excellent. A value between 0 and 1 for NSE generally indicates that the method performance is acceptable. The value is less than 0 indicates that the method performance is insufficient.

KGE is originally developed to compare the modelled and observed time series. KGE is a model evaluation criterion that can be differentiated in the contribution of mean, variance and correlation to model performance. KGE range from −∞ to 1. KGE = 1, indicating excellent agreement between simulations and observations. The KGE score for mean flow comparison in hydrological models is KGE ≈−0.41. The closer the KGE value is to 1 in performance measurement, the more accurate the model will be.

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{cal}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{cal}}{\mu_{obs}} - 1\right)^2} \tag{12}$$

Where; $\alpha = \frac{\sigma_{cal}}{\sigma_{obs}}$, $\beta = \frac{\mu_{cal}}{\mu_{obs}}$ and where $(\mu_{cal}, \sigma_{cal})$ and $(\mu_{obs}, \sigma_{obs})$ are the mean and standard deviation of estimation and observation. KGE calculates the Euclidean distance (ED) of the three components from the ideal point. This ideal point avoids underestimation of variability and enables comparison of the term bias across monthly. Like NSE, KGE=1 indicated perfect connection between estimations and observations.

R^2 measures the strength of the linear relation between x and y pairs, and the results are expected to be between 0 and 1. The closer the result is to 0, the more the model diverges from reality. The Equation 13 of R^2 is given below.

$$R^2 = \left(\frac{(n\sum xy) - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}\right)^2 \tag{13}$$

3. Results and Discussions

In this study, a data set covering monthly hydroelectric production values of Türkiye for the period January 2007-December 2018 was used. Using long-term monthly hydroelectric generation information, a forecasting model based on LSTM networks has been developed that predicts 12-month hydraulic production annually. The block diagram of the LSTM-based deep learning hydroelectric generation system aimed in this study is given in Figure 8. When estimating hydroelectricity generation, first, the generation dataset is standardized by proportioning the annual installed capacity value to the boxed capacity, and after training the LSTM deep network with this generation-based time series, to convert the monthly-based generation forecast values of the next year to the generation values in MWh, the hydroelectric generation board of that year. The capacity value was estimated by the regression model. To make an accurate estimation of the regression model used here high R^2 monovalent annual capacity value of the last five years to achieve linear function it is used.

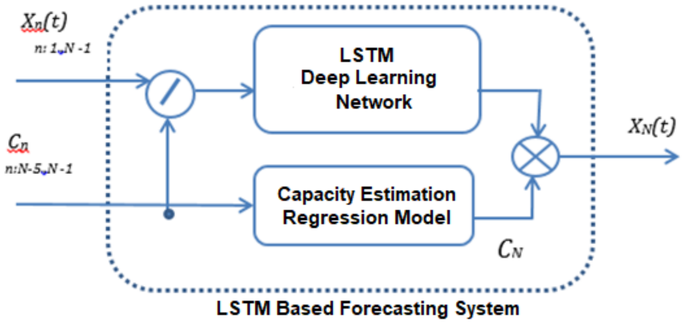


Figure 8. Block Diagram Structure of Proposed the Forecasting System

The data used in this study are taken from the public web page of Türkiye's Transmission System Operator. The results obtained in this study were obtained by running the LSTM algorithm under the Matlab program. Data set used in this study, the monthly value by Türkiye's electric transmission system operator in the country (Transmission System Operator-TSO) has established is provided from open sources [25-27]. This data set, starting from January 2007, includes a total of 144 months of hydroelectric production information, covering the month of December 2018. This data set is divided into three parts and designed to work with three LSTM estimated models. Detailed description of the data set used in study are shown in Table 1.

Table 1. Characteristics of the Dataset

Data set	Time Range/Feature	Parameter
Hydropower generation time series (MWh)	01.2007 – 12.2018	n*t number of months 144 pieces
Data time series 72 months: 120 months: 144 months:	01.2007 – 12.2012 01.2007 – 12.2016 01.2007 – 12.2018	*Xn(t)*, production value in t. month of n. year
Installed Capacity (MW)	2007 - 2018	Annually: 12 pieces

In order to find the best LSTM model, LSTM models with different number of layers were used on data with different lengths of time intervals. The effect of the number of layers on the estimation of the model was measured by choosing the number of layers starting from 25 and increasing up to 400. In the estimation system, three different data with monthly production values as 6 years, 10 years and 12 years were used. In the data sets used, 72 data points for 6 years, 120 data points for 10 years data set and 144 data points for 12 years data set were used. In these data sets, 1 year of data is reserved for validation. In the remaining data set, the 1-year data set was used for testing and the remaining data was used for training. These data are given in Table 2, which shows the results for each model. Although LSTM models use the same estimation system, they differ from each other according to the data set used for training. Accordingly, the LSTM-1 model was trained using a 72-month training set, the LSTM-2 model using a 10-month training set, and the LSTM-3 model using a 12-month training set. For all LSTM models, versions consisting of 25, 50, 100, 200 and 400 layers were used and results were obtained. In the study, the graphs of the estimation results with the number of layers that give the best result out of the estimation results obtained for the LSTM estimation models with 6, 10 and 12 year production time series in the estimation system are given. LSTM networks with 25, 50, 100, 200 and 400 hidden layers were used in 3 different models, each of which was operated with monthly production data of 72 months, 120 months and 144 months. An LSTM layer includes an RNN layer that learns the long-term dependencies between the time steps in the time series and the sequence data.

As the first LSTM-1 Model; covering the year 2007-2012 and 72 monthly time series of hydroelectric production was discussed. According to the results obtained from the LSTM-1 model and given in Table 2, the model obtained using the 400-layer long-short-term memory structure gave the best results and is presented in Figure 9.

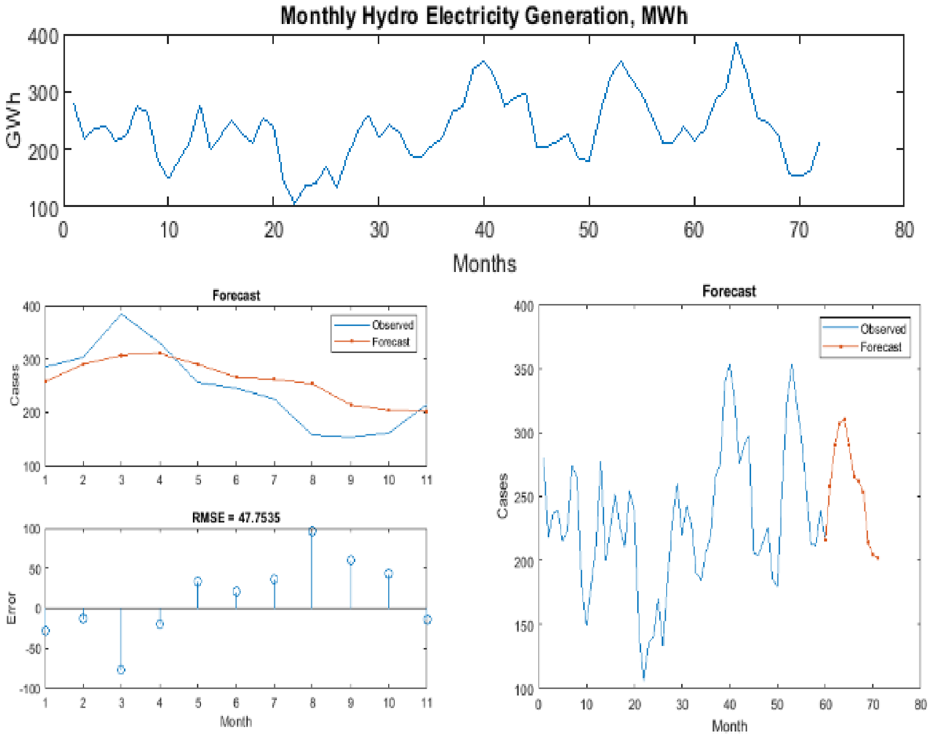


Figure 9. Forecast Results, Error Graph and RMSE Values of 400 Layers Using 72-Month Production Data Series

As the second LSTM-2 Model; the 120-month hydroelectric generation time series covering the years 2007-2016 is discussed. According to the results obtained from the LSTM-2 model and given in Table 2, the model obtained using the 400-layer long-short-term memory structure gave the best results and is presented in Figure 10.

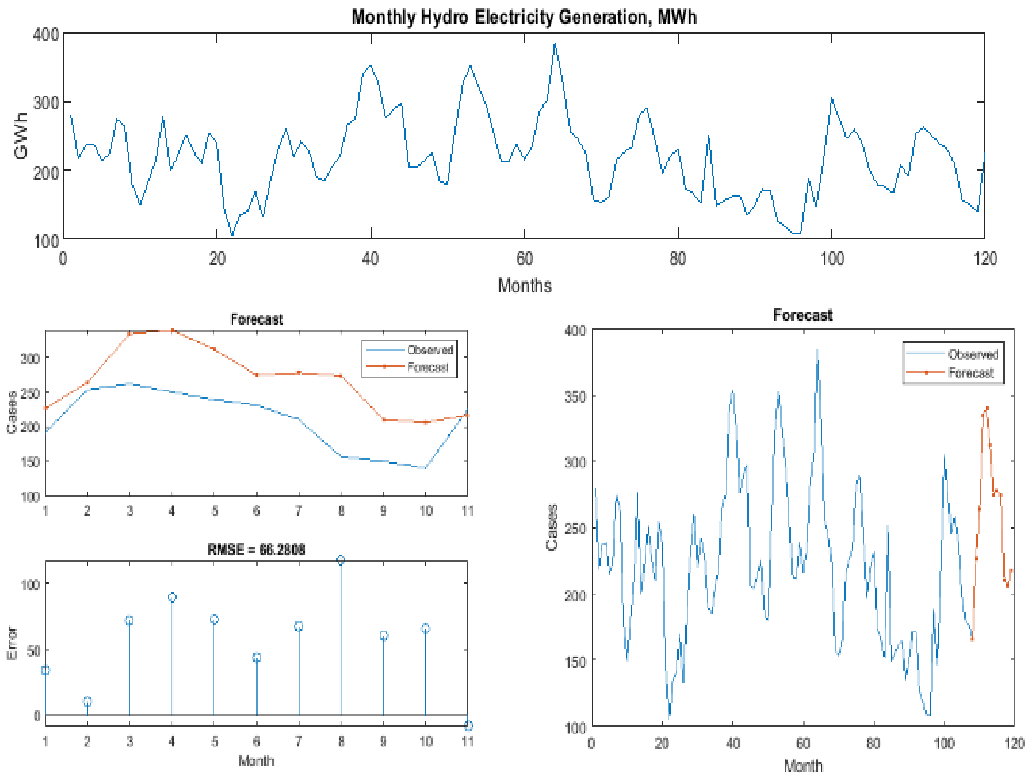


Figure 10. Forecast Results, Error Graph and RMSE Values of 400 Layers Using 120-Month Production Data Series

As the third LSTM-3 Model; the 144-month hydroelectric generation time series covering the years 2007-2018 is discussed. According to the results obtained from the LSTM-3 model and given in Table 2, the model obtained using the 200-layer long-short-term memory structure gave the best results and is presented in Figure 11.

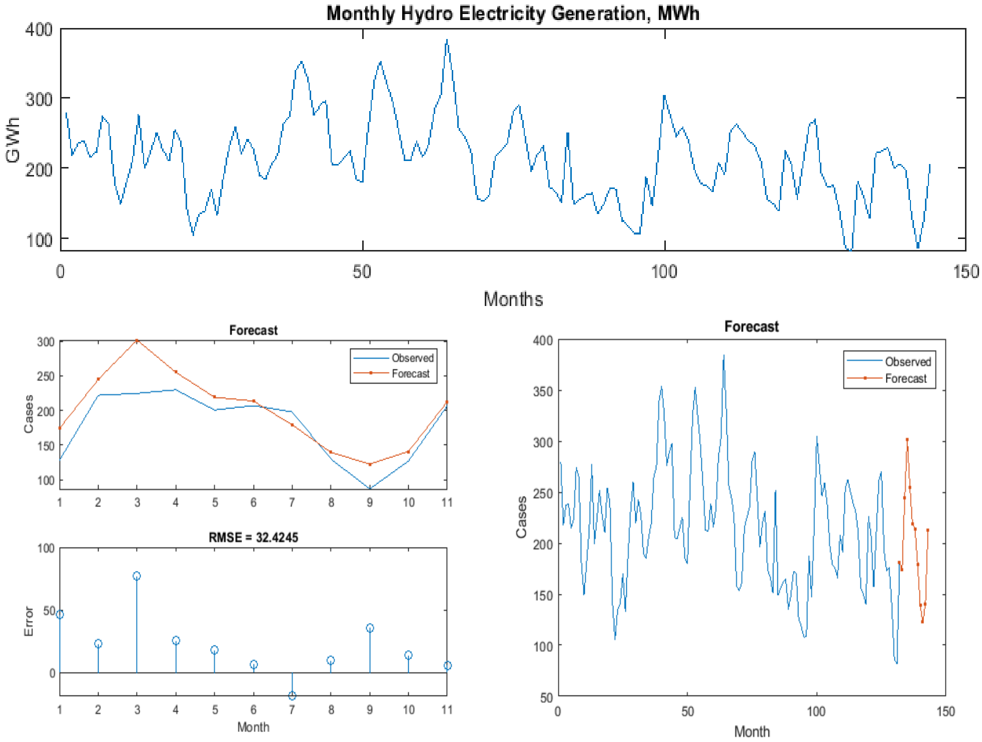


Figure 11. Forecast Results, Error Graph and RMSE Values of 200 Layers Using 144-Month Production Data Series

When the performance values and model efficiency of these models developed to estimate hydraulic production are evaluated; according to the results obtained in Table 2, which includes the RMSE and MAP values of the results obtained in the study; It is seen that the 200-layer LSTM-3 model, which includes 12 years of hydroelectric time data and 144 data points on a monthly basis, is the most predictive model with an annual RMSE of 32.4245 and an annual MAPE of 0.1603. Looking at the results in the Table 2, it is seen that the number of layers of the model with the best estimate for each LSTM model is close to each other. When 400 layers were used for LSTM-1 and LSTM-2, more suitable results were obtained for the prediction values compared to the others. However, the LSTM-3 model, which gave the best results, was found to have 200 layers. RMSE=32.4245 and MAPE=16.03% values and 200-layer LSTM model trained with 12-year data with 144 monthly data points containing hydroelectric generation information was obtained as the highest model, and the performance values of the model showed that it was the correct forecasting model. The overall efficiency parameters of the found LSTM model were checked with NSE=0.5398 and KGE=0.8413 values, the performance of the method was found to be a high-accuracy model within acceptable limits and with the correlation value of R2=0.9035 to be very close to reality. The results showed that the LSTM based forecasting model can be used as an acceptable hydropower generation forecasting model.

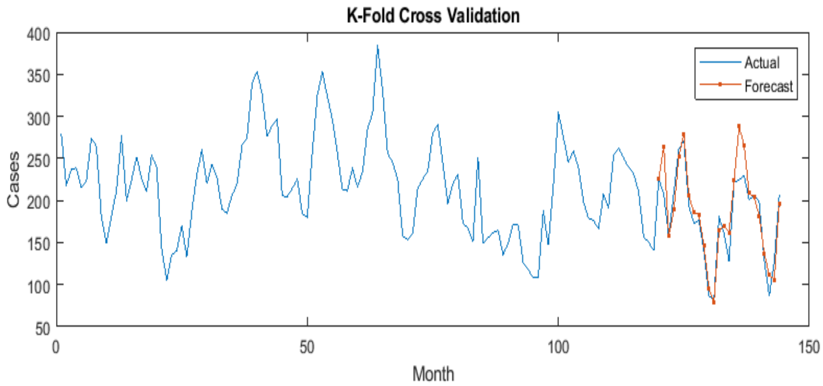


Figure 12. Results of K-Fold Validation for Best LSTM Model, which is 200 Layers Using 144-Month Production Data Series

The k-fold cross-validation method separates the data into equal parts according to the specified k number and ensures that each part is used for both training and testing, thus minimizing deviations and errors caused by scattering and fragmentation. The method was applied for the 200-layer LSTM model using a 144-month data set with the best performance and efficiency values. Statistical validation was made to analyze the predictions of LSTM, which was chosen as the most successful model for the results obtained, the data set was divided into 6 parts, 5 parts for training and one set for testing. The result obtained to show the performance of the model is shown in Figure 12 and it is seen that the selected model produces a value suitable for the prediction.

Table 2. Model Efficiency Metrics for Epoch Number=500

Model No	**Time Interval (months)**	**Training Part (months)**	**Testing Part (months)**	**Layer Number**	**Model Evaluation**		**Efficiency of Model**		
					RMSE	**MAPE (%)**	**NSE**	**KGE**	R^2
LSTM-1	72	60	12	25	57.7255	0.2265	0.3470	0.3350	0.7203
				50	70.0400	0.2450	0.0387	0.3407	0.5704
				100	58.6001	0.2025	0.3271	0.4170	0.6689
				200	52.3823	0.2124	0.4623	0.4041	0.7448
				400	**47.7535**	**0.1919**	**0.5531**	**0.5197**	**0.7944**
LSTM-2	120	108	12	25	72.2326	0.3422	-1.9366	0.5662	0.8679
				50	69.2135	0.3261	-1.6963	0.6709	0.8819
				100	68.4791	0.3108	-1.6394	0.7124	0.8062
				200	75.8111	0.3447	-2.2348	0.6012	0.8174
				400	**66.2808**	**0.3031**	**-1.4726**	**0.7141**	**0.8321**
LSTM-3	144	132	12	25	47.1694	0.2762	0.0260	0.7445	0.8368
				50	43.5890	0.2669	0.1683	0.5146	0.6305
				100	47.0654	0.2905	0.0303	0.7782	0.8367
				200	**32.4245**	**0.1603**	**0.5398**	**0.8413**	**0.9035**
				400	36.8064	0.2298	0.4070	0.6516	0.8938

RMSE) and MAPE metrics were used to analyze the performance of three learning algorithms based on 72-month, 120-month and 144-month generation hydroelectric generation data used in the study. The best prediction was obtained in the 200-layer LSTM model using 144 months (12 years) hydroelectric generation time data with the lowest MAPE percentage and lowest RMSE value. The NSE, KGE values of this model were higher than 0.5, close to the ideal 1 value, and the coefficient of correlation value (R^2) was found to be satisfactory with a value of 0.9035 in terms of estimation efficiency.

The more data available, the more accurate the predictions are, the more accurate the results are demonstrated here. However, if a similar study is conducted on a larger training set covering longer years, it is predicted that the algorithms used for the proposed estimation system will yield better results in their performance.

4. Conclusions

As a result of global warming, increasing droughts and decreasing rainfall directly affect hydroelectric production, water management strategies come to the fore in order to use water more effectively in dams, and it is necessary to ensure the sustainability of hydroelectric energy. Hydroelectric production is a process that is constantly changing over time and is changing with changing global warming. LSTMs are naturally designed to work with time series data, allowing them to better understand how data changes over time and more accurately predict future values. It was preferred in this study for hydropower forecasting due to its features such as capturing long-term dependencies and being robust to noise.

Electricity production from renewable energy sources depends mainly on meteorological conditions such as temperature, humidity, wind speed and rainfall in the geography where the facility is located. Therefore, due to the intermittent nature of renewable resources, hydroelectricity production depends on the amount of rainfall and the amount of incoming water, and due to the fluctuating nature of the production, it is important to estimate the hydroelectricity production to be provided to the electricity grid. As in many countries in the world, electricity generation from hydroelectric sources in Turkey is among the important renewable energy sources, and at times approximately 30% of the country's electricity production is provided by hydropower plants. In this sense, it is evaluated that the deep learning-based models proposed in this article will contribute to the studies on production estimation of hydroelectric power plants, which have an important share in our country's electricity production.

This study aims to choose the best LSTM for energy estimation, its performance in the overall model structure of LSTM is analyzed. For this, LSTM models with different training data sets and different number of layers are designed. The performances of the LSTM based on these criteria were given comparatively on a table and the model with the best result was tried to be determined. This study focuses on the potential of using deep learning LSTM to forecast annual hydroelectric power demand on monthly basis. For this, an estimation system based on annual hydroelectricity installed capacity development with monthly hydroelectricity production time data has been proposed. The dataset used in this study was divided into three parts 6, 10, and 12 years, estimation was conducted with three different LSTM models and the effect of the dataset lengths on the prediction was tried to be observed. Thus, three different LSTM models were created with datasets with separate time intervals and predictions were made. A the acceptability level of the dataset used to train the model was investigated to obtain satisfactory prediction results, and it was aimed to observe the short, medium and long-term prediction performance of the production dataset. According to the study, the results of the research show that LSTM provides a robust architecture for the prediction of hydropower production in medium or long-term forecasts such as at least 120 months and 144 months. It is observed that the LSTM network-based forecasting system was successful within the acceptance limits by using the time series data of hydroelectric generation and the installed power values.

References

[1] D. Gielen, F. Boshell, D. Saygin, M.D. Bazilian, N. Wagner, R. Gorini, “The role of renewable energy in the global energy transformation”, *Energy Strategy Reviews*, 24: 38-50, 2019.

[2] P.A. Owusu, S. Asumadu-Sarkodie, “A review of renewable energy sources, sustainability issues and climate change mitigation”, *Cogent Engineering*, 3(1):1167990, 2016.

[3] M.O. Abdeen, “Energy use and environmental impacts: A general review”, *Journal of Renewable and Sustainable Energy*, 1:053101, 2009.

[4] H. Alrayess, Gharbia S., Beden N., A. Ulke, “Using Machine Learning Techniques and Deep Learning in Forecasting the Hydroelectric Power Generation in Almus Dam, Türkiye”, *5th International Symposium on Dam Safety*, Istanbul,Türkiye, Oct. 27-31, 2018.

[5] F. Kratzert, Klotz D., Brenner C., Schulz K., M. Herrnegger, “Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks”, *Hydrology and Earth System Sciences*, 22: 6005–6022, 2018.

[6] K. Amarasinghe, Marino, D.L., Manic, M. “Deep neural networks for energy load forecasting”, *IEEE 26th Int. Symposium on Industrial Electronics (ISIE)*, Edinburgh, 1483-1488, 2017.

[7] Choi, J.Y., Lee, B, ”Combining LSTM Network Ensemble via Adaptive Weighting for Improved Time Series Forecasting”, *Hindawi Mathematical Problems in Engineering*, 2470171:1-8, 2018.

[8] Wang, H., Lei, Z., Zhang, X., Zhou, B. and Peng, J. (2019). A review of deep learning for renewable energy forecasting”, *Energy Conversion and Management*, 198:111799.

[9] Y. Cheng, Xu C., Mashima, D., Thing, V., and Y. Wu, “PowerLSTM: Power Demand Forecasting Using Long Short-Term Memory Neural Network”, In: Cong G., Peng WC., Zhang W., Li C., Sun A. (eds) *Advanced Data Mining and Applications*. ADMA, 727-740, 2017.

[10] N. Al Khafaf, Jalili M., P. Sokolowski, “Application of Deep Learning Long Short-Term Memory in Energy Demand Forecasting”, in: Macintyre J., Iliadis L., Maglogiannis I., Jayne C. (eds) *Engineering Applications of Neural Networks*. EANN, 31-42, 2019.

[11] J. Zheng, Cencen, Xu, Zhang, Z., and L. Xiaohua, “Electric load forecasting in smart grids using Long-Short-Term-Memory based Recurrent Neural Network*”, 51st Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, 1-6, 2017.

[12] E. Choi, Cho, S., D.K. Kim, “Power Demand Forecasting Using Long Short-Term Memory (LSTM) Deep-Learning Model for Monitoring Energy Sustainability”, *Sustainability*, 12:1109, 2020.

[13] D. Altunkaya, Yilmaz, B., “Multivariate Short-term Load Forecasting Using Deep Learning Algorithms”, the Eurasia Proceedings of Science, Technology, *Engineering & Mathematics (EPSTEM)*, 11:14-19, 2020.

[14] T.G. Barbounis, Theocharis J.B., Alexiadis M.C., and Dokopoulos P.S., “Long-term wind speed and power forecasting using local recurrent neural network models”, *IEEE Transactions on Energy Conversion*, 21(1):273-284, 2006.

[15] Li, L., Yao, F., Huang, Y., and Zhou, F., "Hydropower Generation Forecasting via Deep Neural Network", *2019 6th International Conference on Information Science and Control Engineering (ICISCE)*, Shanghai, China, 324-328, (2019).

[16] R.Torres, Dorado, F., and Durán, J., “Energy Demand Forecasting Using Deep Learning: Applications for the French Grid”, *Energies*, 13(9):2242, 2020.

[17] O. Yuksek, Komurcu, M.I., Yuksel, I., and K. Kaygusuz, “The role of hydropower in meeting Türkiye’s electric energy demand”, *Energy Policy*, 34:3093–3103, 2006.

[18] I. Kollia and Kollias, S., “A Deep Learning Approach for Load Demand Forecasting of Power Systems”, *IEEE Symposium Series on Computational Intelligence (SSCI)*, India, 912-919, 2018.

[19] He, W., “Load Forecasting via Deep Neural Networks”, *Procedia Computer Science*, 122:308-314, 2017.

[20] M. Bulut, Tora H. and B. Magdi, "Comparison of Three Different Learning Methods of Multilayer Perceptron Neural Network for Wind Speed Forecasting", *Gazi University Journal of Science*, 34.2: 439-454, 2021.

[21] R. Ünlü,“A Comparative Study of Machine Learning and Deep Learning for Time Series Forecasting: A Case Study of Choosing the Best Prediction Model for Türkiye Electricity Production”, *Suleyman Demirel University Journal of Natural and Applied Sciences*, 23:635-646, 2019.

[22] S. Bernstein, J. Spring and D. Stanway, Insight: Droughts shrink hydropower, pose risk to global push to clean energy, August 14, 2023, https://www.reuters.com/business/sustainable-business/inconvenient-truth-droughts-shrink-hydropower-pose-risk-global-push-clean-energy-2021-08-13/ [Accessed : 05.08.2024].

[23] F. A. Kaplan, The Effect of Climate Change On Hydroelectric Power Plants and Hepp, Ges Hybrid Model, Dsi Technical Bulletin Issue: 141, January 2023

[24] M. Bulut, İzzet A. "Prospects of renewable energy penetration to Türkiye's electric power grid", *International Journal of Global Warming* 25:3-4, 366-377, 2021.

[25] Türkiye Electricity Transmission Inc., (TEİAŞ) https://www.teias.gov.tr/turkiye-elektrik-uretim-iletim-istatistikleri. [Accessed : 22.09.2020].

[26] Electricity Market Regulatory Authority (EMRA) of Türkiye, https://www.epdk.gov.tr/Detay/Icerik/3-0-167/resmi-istatistikleri. [Accessed : 23.09.2020].

[27] Turkish Load Dispatch Information System (YTBS), https://ytbsbilgi.teias.gov.tr/ytbsbilgi/frm_istatistikler.jsf. [Accessed : 22.09.2020].

[28] G. Montavon, Samek, W., K-R. Müller. “Methods for interpreting and understanding deep neural networks”, *Digital Signal Processing*, 73:1–15, 2018

[29] A. Sherstinsky, “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network, *Physica D: Nonlinear Phenomena*, 404:132306, 2020.

[30] S. Hochreiter., J. Schmidhuber,“Long short-term memory”, *Neural Computing*, 9(8):1735–1780, 1997.

[31] X. Wang, Y. Zhao and F. Pourpanah, “Recent advances in deep learning”, *Int. Journal of Machine Learning and Cybernetics*, 11:747–750, 2020.

[32] V. Houdt, V., Greg & Mosquera, G. Carlos & Nápoles, ”A Review on the Long Short-Term Memory Model”, *Artificial Intelligence Review*, 53. 10.1007/s10462-020-09838-1, 2020.

Article Information Form

Author(s) Contributions

All procedures of the article were designed by Mehmet Bulut, data was prepared, results were obtained and the article was written.

Conflict of Interest Notice

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

Ethical Approval and Informed Consent

It is declared that during the preparation process of this study, scientific and ethical principles were followed, and all the studies benefitted from are stated in the bibliography.

Availability of data and materials

Not Applicable

Plagiarism Statement

This article has been scanned by iThenticate ™.