Research Article
Journal of Intelligent Systems: Theory and Applications 8(1) (2025) 25-34
DOI: 10.38016/jista.1447980

Renewable Energy Forecasting in Turkey: Analytical Approaches

Mehmet Berke Colak[1], Erkan Özhan[2*]

[1, 2] Department of Computer Engineering, Çorlu Faculty of Engineering, Namık Kemal University, Tekirdağ, Türkiye

berkecolak95@gmail.com, eozhan@nku.edu.tr

Abstract

The growing population and industrialization have resulted in an increased demand for energy, which has worsened environmental problems such as pollution and climate change. Renewable energy sources are considered a promising solution due to their environmental benefits and limited potential. This study examines the use of neural networks and time series analysis to predict electricity generation rates from renewable energy sources in Turkey. We use the LSTM, NNAR, and ELM models, all of which utilize the backpropagation algorithm for neural network forecasting. Additionally, we apply ARIMA, Holt's trend, linear regression, mean, and exponential smoothing models for time series analysis. We evaluate the performance using the mean absolute error and root mean square error on the training and test data. The study showed that LSTM models outperformed the ARIMA (1,2,1), ARIMA (2,2,1), ARIMA (3,2,1), and NNAR methods in forecasting accuracy. Although the NNAR model initially had the lowest error, its linear predictions made it less suitable for practical applications. This study highlights the effectiveness of neural networks and time series analysis in predicting renewable energy sources. The ARIMA (1,2,1), LSTM and ARIMA (3,2,1) modeling methods are useful for optimizing the planning and management of Turkey's renewable energy future, contributing to a more sustainable energy landscape.

Keywords: Renewable energy, Turkey, time series, neural networks, climate change, ARIMA, LSTM

Türkiye'de Yenilenebilir Enerji Tahmini: Analitik Yaklaşımlar

Öz

Artan nüfus ve sanayileşme, enerji talebinin artmasına neden olmuş, bu da kirlilik ve iklim değişikliği gibi çevre sorunlarını daha da kötüleştirmiştir. Yenilenebilir enerji kaynakları, çevresel faydaları ve sınırsız potansiyelleri nedeniyle ümit verici bir çözüm olarak değerlendirilmektedir. Bu çalışma, Türkiye'de yenilenebilir enerji kaynaklarından elektrik üretim oranlarını tahmin etmek için sinir ağlarının ve zaman serisi analizinin kullanımını incelemektedir. Sinir ağı tahminleri için her ikisi de geri yayılım algoritmasını temel alan LSTM, NNAR ve ELM modellerini kullanıyoruz. Ayrıca zaman serisi analizi için ARIMA, Holt trendi, doğrusal regresyon, ortalama ve üstel düzeltme modellerini kullanıyoruz. Performansı, eğitim ve test verilerinde ortalama mutlak hata ve kök ortalama kare hata kullanarak değerlendiriyoruz. Çalışma, LSTM modellerinin tahmin doğruluğunda ARIMA (1,2,1), ARIMA (2,2,1), ARIMA (3,2,1) ve NNAR yöntemlerinden daha iyi performans gösterdiğini göstermiştir. NNAR modeli başlangıçta en düşük hataya sahip olmasına rağmen doğrusal tahminleri onu pratik uygulamalar için daha az uygun hale getirdi. Çalışma, yenilenebilir enerji kaynaklarının tahmin edilmesinde sinir ağlarının ve zaman serisi analizinin etkinliğini vurguluyor. ARIMA (1,2,1), LSTM ve ARIMA (3,2,1) modelleme yöntemleri, Türkiye'nin yenilenebilir enerji geleceğinin planlanması ve yönetimini optimize etmek ve daha sürdürülebilir bir enerji ortamına katkıda bulunmak için kullanışlıdır.

Anahtar kelimeler: Yenilenebilir enerji, Türkiye, zaman serileri, sinir ağları, iklim değişikliği, ARIMA, LSTM

1. Introduction

Energy, in its basic form, is a system's ability to perform work or generate heat, while renewable energy refers to naturally replenished sources of energy (Coburn and Farhar, 2004). Renewable energy is a source of energy continually replenished by natural processes (Hersh, 2006). Renewable energy can be obtained from sources with a nearly limitless supply, ensuring sustainability over time ("Renewable energy explained - U.S. Energy Information Administration (EIA)," 2023). Renewable energy sources may vary among countries. Turkey possesses a diverse range of renewable energy sources, including hydropower, wind, and solar energy. According to information from the YTBS website,

*Corresponding Author.
E-mail: eozhan@nku.edu.tr

Recieved : 7 Mar 2024
Revision : 30 Apr 2024
Accepted : 16 Oct 2024

Turkey's renewable energy usage rates in 2023 were as follows: biomass accounted for 2.65%, solar for 5.83%, geothermal for 3.43%, and wind for 10.52%. To increase energy production, it is important to utilize different sources and invest in renewable energy. In this regard, energy prediction studies are of critical importance in accurately forecasting future energy demand and planning necessary investments. Many studies have demonstrated the success of methods such as artificial neural networks (ANNs) in numerical estimation problems. In this study, various prediction models were employed, including neural network autoregression (NNAR), extreme learning machines (ELM), exponential smoothing, Holt's trend, linear regression, mean model, autoregressive integrated moving average (ARIMA) and long short-term memory (LSTM). ANNs can learn dispersed relationships in data (Mossalam and Arafa, 2018). ANNs models possess a structure that mimics the learning functions of the human brain. Information is transmitted through connections between neurons, and the network is trained with preexisting datasets. The goal is to find the network configuration that will produce the most accurate prediction. Various components, such as the learning algorithm and the activation function, contribute to achieving this configuration. Determining the optimal network structure can be considered an optimization problem. Once this structure is established, the developed network model can be used to make future predictions with minimal inaccuracy. Typical ANN models use simplified neuron models similar to human neurons (Nastos et al., 2013). ANN, ARIMA, NNAR, and LSTM models are widely used and effective methods for energy prediction. ANNs have the ability to learn from complex datasets and forecast future trends, enabling the prediction of energy production and consumption levels, price fluctuations, and other factors. Many of these methods have been tested to establish standards, and the selection of the estimation model is based on error rates.

A time series consists of observations produced sequentially over time. A set is considered continuous if it is continuous; otherwise, it is discrete (Baskan, 2008). The primary objective of time series analysis is prediction. The fundamental idea is to utilize past observations to forecast the future, and the model that best describes the data is then employed to predict future outcomes based on historical records (Baccar, 2019).

This study investigated the usability of current methods such as artificial neural networks, ARIMA, ELM, NNAR and LSTM in estimating Turkey's renewable energy production rate.

This study aims to guide Turkey's decision-making process for its renewable energy future and identify the most effective methods. The use of new technologies such as ANNs is important for improving Turkey's energy production capacity and ensuring energy security. Forecasts regarding Turkey's energy production capacity in the future are important for planning investments and ensuring energy security. In this study, the findings of these predictions obtained from the most up-to-date methods are presented.

2. Literature Review

The literature has been surveyed to provide brief summaries of studies conducted chronologically in Turkey and around the world that employ forecasting methods related to renewable energy sources.

Paoli et al. employed neural network (MLP), ARIMA, the k-nearest neighbors algorithm, Bayesian decipherment, and Markov chain estimation methods to predict preprocessed daily solar radiation time series (Paoli et al., 2010a). Hocaoglu and Karanfil utilized Granger causality and impulsive response analysis estimation methods in their study, adopting a time series-based approach to renewable energy modeling (Hocaoglu and Karanfil, 2013a). Golestaneh et al. employed the ELM estimation method (Golestaneh et al., 2016). Jiang et al. utilized the ELM estimation method. Additionally, the investigation incorporated the bacterial-foraging optimization algorithm (BFOA) and empirical mode decomposition (EMD) as estimation methods. This study applied empirical mode decomposition and an advanced ELM optimized with the BFOA to estimate China's renewable energy terminal power consumption (Jiang et al., 2019). Tharani et al. utilized various machine learning techniques to comprehensively examine and project renewable energy trends (Tharani et al., 2020). Goncalves et al. utilized vector autoregression, privacy preservation, and distributed learning prediction methods. Additionally, they conducted a study on confidentiality-preserving distributed learning for renewable energy prediction (Goncalves et al., 2021). In their research, Gullu and Kartal focused on the electricity production objectives derived from renewable energy sources, including solar, wind, and hydroelectric power. They adopted the Box-Jenkins ARIMA methodology to estimate the individual installed capacities of these various renewable energy types. (Güllü and Kartal, 2021). In another study conducted by Cetin et al., future energy production was predicted using real data from a solar power company and employing machine learning algorithms. This study utilized the LSTM method, a type of ANN, to conduct predictions and analyses. The results revealed an error rate ranging from 1% to 15%. Future studies will focus on other renewable sources like wind, geothermal, and hydro energy (Çetin and Işık, 2021).

Erturk et al. developed ANN models using MATLAB for four provinces located in different climatic zones of Turkey (Kayseri, Rize, Hakkari, and Izmir) to accurately determine the amount of solar radiation. The solar radiation predictions made by the model yielded the best results for the province of Hakkari, with an R^2 value of

0.93, followed by Izmir, Kayseri, and Rize. There was a consistent agreement between the values predicted by the ANN models and the measured values for each province (Ertürk et al., 2023).

Kaysal et al. conducted a study using data from a wind farm located in the Mediterranean region, spanning the years 2018 to 2020. They employed convolutional neural network (CNN) and binary long short-term Memory (BLSTM) algorithms for prediction purposes (Kaysal et al., 2023). Çakir (2023), highlighted the increasing challenge of forecasting REG for effective energy management. Various time series models, encompassing physical models, statistical techniques, and artificial intelligence algorithms, have been proposed to address this challenge. Notably, fuzzy time series (FTS) models were applied to forecast Turkey's REG between 2000 and 2020. The results indicate that the proposed integrated model demonstrates high accuracy and serves as a valuable tool not only for REG forecasting but also for addressing other time series forecasting problems. Cakir suggested further exploration of this integrated model (Cakir, 2023).

Rajni et al. examined monthly energy production data spanning from January 1973 to December 2019. They conducted a study on renewable energy production in the United States from January to December 2020. This investigation employed ARIMA time series analysis techniques to forecast ten future time periods (months), specifically considering total renewable energy production (Rajni et al., 2024).

In a 2024 study, Bquet et al. developed an AI-based framework at the Swiss Federal Institute of Technology (EPFL) using a Long Short-Term Memory (LSTM) model to predict solar energy for different time horizons. The dataset consisted of 17,297,280 Global Horizontal Irradiance (GHI) measurements taken every 10 seconds between January 1, 2016, and November 1, 2021. The LSTM model proved highly effective for short-term forecasts, particularly for horizons a few hours ahead (Bouquet et al., 2024).

Solano et al. used Support Vector Regression (SVR), Extreme Gradient Boosting (XGBT), Categorical Boosting (CatBoost) machine learning algorithms for solar radiation prediction and proposed an ensemble feature selection method to select the most relevant input parameters and their past observations. The method called Voting Average (VOA) is an ensemble learning method that includes SVR, XGBT and CatBoost. As a result of the study, they proved that VOA outperformed the other algorithms (Solano et al., 2022).

In this study, in comparison to other studies identified through a literature review, the commonly utilized forecasting methods were ARIMA, employed by Paoli et al., Güllü and Kartal, Rajni et al. Golestaneh et al. and Jiang et al. utilized ELM in their studies. Additionally, LSTM was utilized by Çetin and Işık, Kaysal et al... Furthermore, Çetin and Işık, Erturk et al. and Han et al. employed ANNs.

Previous studies by Çetin and Işık and Ertürk et al. had a narrower focus, examining a single company's production capacity or a localized region, respectively. Güllü and Kartal used the ARIMA methodology to predict solar, wind, and hydroelectric power. In contrast, this study estimates the share of renewable energy in total energy production, excluding hydroelectric power. Unlike previous studies, this research evaluates the most recent data from 1960 to 2023, encompassing all renewable energy sources in Turkey. Additionally, it compares several established algorithms using the most recent data. This study stands out by testing eight different forecasting methods on the most comprehensive dataset and demonstrating the suitability of the ARIMA and LSTM algorithms predictive models by focusing on the top five results.

Upon reviewing the literature, it is evident that studies have focused on forecasting solar energy alone (Çetin and Işık, 2021; Ertürk et al., 2023; Goncalves et al., 2021; Paoli et al., 2010b), wind energy (Kaysal et al., 2023), both solar and wind energy (Cakir, 2023), and a combination of solar, wind, and hydropower (Golestaneh et al., 2016; Güllü and Kartal, 2021; Jiang et al., 2019). Additionally, numerous studies explore renewable energy consumption (Jiang et al., 2019), trends (Tharani et al., 2020), and the relationships between various parameters of renewable energy (Hocaoglu and Karanfil, 2013b).

When examining studies on renewable energy sources in Turkey, Cetin et al. employed the LSTM algorithm, but their work was limited to a specific region within Turkey, and similar to Erturk et al., they only forecast solar energy. Our study, on the other hand, covers the entire country and focuses on predicting energy generation from all renewable sources, excluding hydropower. Hocaoglu et al. investigated the relationships between parameters of renewable energy in their study, while Güllü and Kartal included hydropower among renewable energy sources. Çakır aimed to forecast solar and wind energy generation using a dataset spanning 2000 to 2020. An analysis of Çakır's predictions reveals that the aim was to estimate total production volume. However, our study focuses on predicting the share of renewable energy in Turkey's total energy production as a percentage. When comparing the trends, both studies indicate an upward trend. Moreover, while Çakır's study produced forecasts only up to 2020, our work extends the predictions to 2028, offering continued insights beyond 2020, thus contributing to a more sustained forecasting approach.

3. Method

3.1. Methods Used for Prediction

3.1.1. Extreme Learning Machine (ELM) Method

The single hidden layer feed-forward neural network consists of three neural layers. Its name derives from the nonlinear hidden layer in the model, which processes input layer data features. When the hidden layer performs no computations, the output layer becomes linear and is devoid of any transformation function or bias (Akusok, 2016). The single-layer hidden-layer neural network model with L, a number of training samples in with N, a weight vector between the input and hidden layers with w, an activation function with g, a vias vector with b and an input vector with x. The ELM method is formulated as equation g(x) as shown in Equation 1 (Erdem, 2020), which includes hidden neurons and an activation function.

$$\sum_{j=1}^{L} \beta_j \, . \; g\left(\left(w_{j,}x_i\right) + b_j\right) = y_i \, , \;\; i = 1,2,..,N \qquad (1)$$

Extreme Learning Machine (ELM) is a fast algorithm for Single-Layer Feedforward Networks (SLFN), randomly assigning input weights and biases, while analytically determining output weights using the generalized inverse of the hidden layer's output, thus greatly improving learning speed and generalization performance (Guang-Bin Huang et al., 2004).

3.1.2. Neural network autoregression (NNAR) model method

The neural network autoregression (NNAR) model is a three-layer feedforward neural network (Maleki et al., 2018), as illustrated in Figure 1.

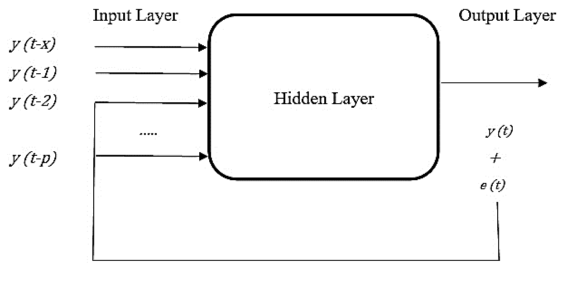

Figure 1. Schematic representation of the NNAR model (Gibson, 2020).

The NNAR (Neural Network Autoregression) model is a data-driven, feedforward neural network that uses the backpropagation algorithm for training and parameter estimation (Sadia et al., 2022). The input layer of the network receives the lagged values of the time series, representing past observations and in the hidden layer, the model learns complex nonlinear relationships between these past values (Daniyal et al., 2022). The backpropagation algorithm is used to minimize the error between the predicted and actual values by adjusting the network's weights. This error is propagated backwards from the output layer through the network, updating the weights to improve future predictions.

3.1.3. Autoregressive Integrated Moving Average (ARIMA) Modeling Method

The ARIMA model was developed according to the methodology described by Box and Jenkins (Box et al., 1994). There are three basic types of ARIMA models: the moving average (MA) model, autoregressive (AR) model, and integrated (I) model (Yang et al., 2020). The nonseasonal model is one of the various ARIMA models. The general equation for the ARIMA (p, d, q) model is formulated in Equation 2 and approximated as shown in Equation 3. In these equations, d represents the number of differences, t denotes the discrete time, $\emptyset_p$ is the autoregressive parameter, $\varepsilon_t$represents the residual and θ_q is the moving average parameter. Additionally, X_t and U_t are both reliable variables. The symbol d represents the difference (Nyatuame and Agodzo, 2018).

$$U_t = \emptyset_1 U_{t-1} + \emptyset_2 U_{t-2} + .. + \emptyset_p U_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .. - \theta_q \varepsilon_{t-q} \qquad (2)$$

$$U_t = X_t - X_{t-d} \qquad (3)$$

During the tuning of the ARIMA models, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were constructed to identify the optimal ARIMA parameters based on the lag observations. An automatic configuration was preferred for the lambda parameter. Additionally, the ACF graph is shown in Figure 2, and the PACF graph is shown in Figure 3.

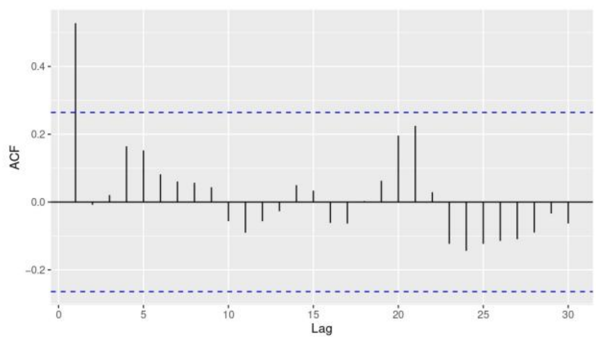

Figure 2. Representation of the ACF Graph

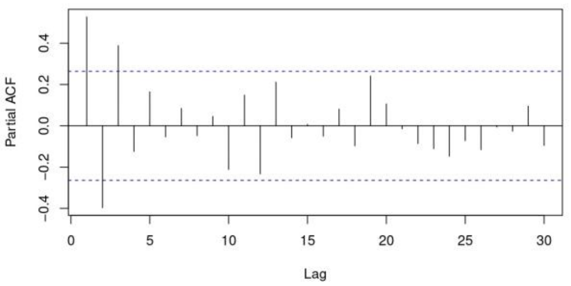

Figure 3. Representation of the PACF Graph

3.1.4. Holt's trend method

Holt's trend method estimates the parameter for trend correction using the Holt-Winters method. The multiplicative model for Holt's trend method is formulated in Equation 4, while the additive model is expressed in Equation 5. In Equation 5, when examining the equation symbols, the symbol S_t represents exponential correction at time t, S_{t-1} signifies exponential correction at time t-1 corresponding to seasonal elements, b_t denotes trend elements at time t and b_{t-1} represents trend elements at time t-1 (Nurhamidah et al., 2020). In addition, β is the smoothing factor for the trend and F is the forecast at steps ahead (Mrutyunjaya, 2020).

$$B_{t-1} = \beta\,(F_{t-1} - F_{t-2}) + (1-\beta)B_{t-2} \quad (4)$$

$$b_t = \beta\,(S_t - S_{t-1}) + (1-\beta)\,b_{t-1} \quad (5)$$

3.1.5. LSTM neural network model

The LSTM model (Hochreiter and Schmidhuber, 1997) is a powerful recurrent neural system specially designed to overcome the exploding/vanishing gradient problems that typically arise when learning long-term dependencies, even when the minimal time lags are very long (Van Houdt et al., 2020).

LSTMs are a type of recurrent neural network (RNN) designed to capture long-term dependencies in time series data, utilizing "cell states" to store important information and "gates" to determine which information should be remembered and which should be forgotten (Olah, 2015)

An LSTM neural network model typically comprises an input sequence layer, one or more LSTM layers arranged sequentially to capture time dependencies in the data, a fully connected layer to transform the output size of preceding layers into the number of classes to be recognized, and a softmax layer to compute the probability of belonging to each class. Additionally, it includes a classification output layer to calculate the cost function (Ghislieri et al., 2021). LSTM networks offer several advantages, including dynamic system modeling capabilities in diverse application domains such as image processing, speech recognition, manufacturing, autonomous systems, communication, and energy consumption (Lindemann et al., 2021).

The adaptive moment estimation (Adam) stochastic gradient descent method, which is based on the adaptive estimation of first and second-order moments, was employed for optimizing the LSTM algorithm.

3.2. Criteria used in the analysis of the results

3.2.1. Mean absolute error (MAE)

The MAE is calculated as the average of the absolute differences, referred to as errors, between the expected and actual observations. As shown in Equation 6, the MAE is calculated as the average of the absolute differences (errors) between the expected and actual observations a_n represents the actual value, is the observed value, n is the number of observations and N is the number of observations.

$$\mathrm{MAE} = \frac{\sum_{n=1}^{N}|o_n - a_n|}{N} \quad (6)$$

3.2.2. Root mean square error (RMSE)

As indicated in Equation 7, the RMSE is employed to calculate the average magnitude of the differences between the predicted and observed values. In this equation, o_n represents the observed value, a_n represents the actual value, n is the number of observations and N represents the number of tuples in the test dataset.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{n=1}^{N}(o_n - a_n)^2}{N}} \quad (7)$$

3.3 Dataset

In the experiments, we utilized a dataset encompassing the total rate of electrical energy obtained by Turkey from renewable energy sources, excluding hydroelectricity, for the years 1960–2023. The data for the years 1960 to 2015 were sourced from the World Bank website, while the data for the years 2015 to 2019 were obtained from the Ministry of Energy. Subsequently, data for the years 2019 to 2023 were acquired from the YTBS-TEIAS website ("Yük Tevzi Bilgi Sistemi (YTBS)-Türkiye Elektrik İstatistikleri," 2023).

The dataset contains two attributes: "Date" and "Electricityproduction". Since the dataset was provided as structured Excel and .csv files by the relevant institutions, no additional data cleaning was performed. However, due to the absence of records for the year 1982, the Kalman Filter method was employed to impute the missing data in the time series. Using this method, the data for 1982-1983 was automatically filled in without disrupting the integrity of the series.

The R programming language, with the R-Studio application, was used for the data analysis. In the experimental results section, the outcomes of the estimation processes are discussed in depth. The R programming language is employed for statistical analysis, particularly by data scientists, academics, and health researchers (Lanovaz and Adams, 2019).

4. Experimental Design

The available data were partitioned into training and testing sets, and the dataset underwent time series estimation using the most preferred methods. Among the time series analysis estimation methods, we employed the ARIMA, Holt's trend, linear regression, mean, and exponential smoothing models. On the other hand, artificial neural network estimation methods include the use of the ELM method, the LSTM neural network model and the neural network autoregression (NNAR) method. The methods employed in the study were implemented using the R programming language in the R-Studio application. The 'Keras3' library and the 'TensorFlow' library were utilized within the R software environment to employ the LSTM algorithm. The data used in the analysis from 1960 to 2018 are labeled training data, while the data from 2019 to 2023 are designated test data. The reason for selecting the data from the last few years as the test set is to evaluate the model's predictive ability based on recent trends and variables. This approach may help the model to better predict future trends. We developed several prediction models using training data from 2019 to 2023. The performance of these models was then evaluated using the MAE and the RMSE. This sequence represents a widely accepted approach to solving this type of problem. Furthermore, these values were compared to determine which method yielded better results. Five results were obtained from each estimation method for the years 2019-2023. Finally, future projections were made for the years 2024 to 2028. As a result of this process, 5 outcomes were obtained, contributing to the overall assessment.

5. Experimental Results

The dataset used is displayed in Figure 4. Upon examination of the graph, it becomes evident that certain irregular increases and decreases occurred in the percentage share lines. Consequently, the observed dataset is identified as having a trend component. The absence of a seasonality component is attributed to the uneven distribution of lines in the graph.

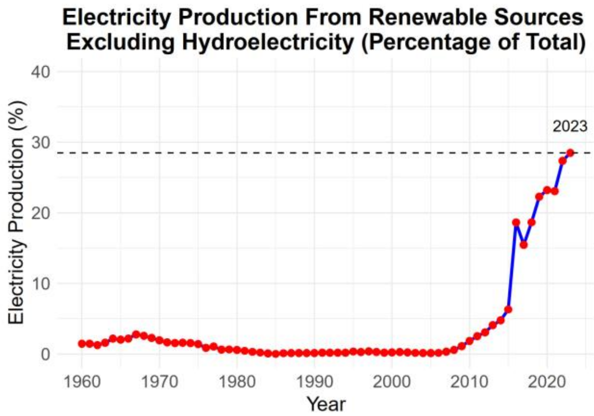

Figure 4. Graphical Representation of the Data Set

Forecasting for the years 2019-2023 was conducted using the training data. First, we tested popular models, and the results are presented in Table 1. Table 2 lists the five estimation methods that yielded the best results. Among the five methods, NNAR demonstrated the best performance when examining the test data segment of the estimation process.

The second-best method was the LSTM modeling method, followed by the ARIMA (1,2,1) modeling method as the third-best estimation approach. The fourth-best estimation method was the exponential smoothing method, followed by the ARIMA (3,2,1) as the fifth best.

Table 1. Performance Values of All Methods

Forecasting Method	MAE	RMSE
NNAR	0.99	1.48
ELM	7.54	8.49
Exponential Smoothing	6.52	7.00
Holt's Trend	7.54	8.50
Linear Regression	20.27	20.42
Mean (constant) Model	22.91	23.05
ARIMA (1,2,1)	1.65	1.84
ARIMA (2,2,1)	7.24	8.28
ARIMA (3,2,1)	6.52	7.18
LSTM	1.40	1.73

Table 2. The Accuracy Values of the Test Data of the Five Methods That Give the Best Results

Forecasting Method	MAE	RMSE
NNAR	0.99	1.48
LSTM	1.40	1.73
Exponential Smoothing	6.52	7.00
ARIMA (1,2,1)	1.65	1.84
ARIMA (3,2,1)	6.52	7.18

Table 3 displays the estimated values and the actual values for the methods that produced the best results.

The table clearly shows that the values obtained using the ARIMA (3,2,1) and ARIMA (2,2,1) modeling methods are closely aligned. Although they exhibit an increase compared to the exponential smoothing method, they do not align well with the actual values. Conversely, it has been revealed that the values obtained through the NNAR method and the LSTM method are more congruent with the actual values than those obtained through other methods.

The actual values in Table 3, along with the estimated values, are depicted in Figure 5. While the estimated values of the ARIMA (3,2,1) and ARIMA (2,2,1) models closely align on the graph, an unrelated pattern is evident in the graph line representing the actual values. The graph shows that the estimation values of the NNAR method, the LSTM method and the ARIMA (1,2,1) modeling method are closer to the graph line of the actual values than those of the other estimation methods.

Table 4 displays the estimated energy percentage values projected for the years 2024–2028, utilizing the five estimation methods that demonstrated the best results on the test data by leveraging the entirety of the training dataset**.**

Table 3. Estimated and actual values of the models

Model / Year	2019	2020	2021	2022	2023
NNAR	21.639	24.182	26.143	27.547	28.499
LSTM	22.79	24.466	26.435	28.249	29.393
Exp. Smooth.	18.344	18.344	18.344	18.344	18.344
ARIMA (1,2,1)	20.029	21.551	23.166	24.879	26.697
ARIMA (2,2,1)	23.906	27.488	31.132	36.14	41.859
ARIMA (3,2,1)	24.905	27.29	30.047	34.945	39.765
Actual Values	22.27	23.194	23.057	27.335	28.482

While the percentile values estimated in the NNAR model and the exponential smoothing model, as shown in Table 4, remained relatively stable, those calculated using the LSTM, ARIMA (1,2,1), and ARIMA (3,2,1) modeling methods increased over the years. Notably, the increase observed in the ARIMA (1,2,1) modeling method surpassed that of ARIMA (3,2,1).

In addition, upon examining the estimated values on the graph shown in Figure 6, it is evident that the graph line representing the percentile values obtained in the neural network autoregression (NNAR) model remains constant. In contrast, the values obtained through the LSTM, ARIMA (1,2,1) and ARIMA (3,2,1) modeling methods increase.

Table 4. Forecasts of the Models for the Next 5 Years

Year	NNAR	LSTM	Exp. Smooth.	ARIMA (1,2,1)	ARIMA (3,2,1)
2024	26.072	34.087	28.453	31.875	31.245
2025	25.303	37.958	28.453	35.422	35.306
2026	25.029	40.576	28.453	39.254	38.903
2027	24.928	43.130	28.453	43.377	42.874
2028	24.890	45.017	28.453	47.806	47.412

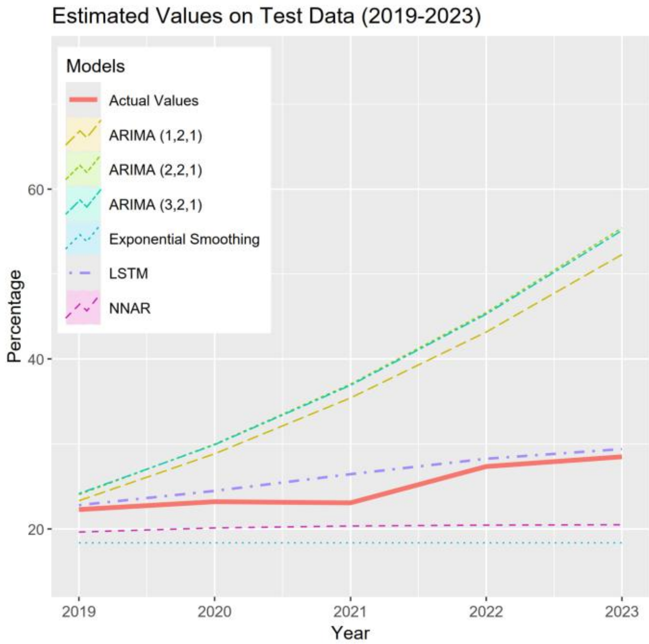

Figure 5. Estimated Values of the Models versus the Actual Values

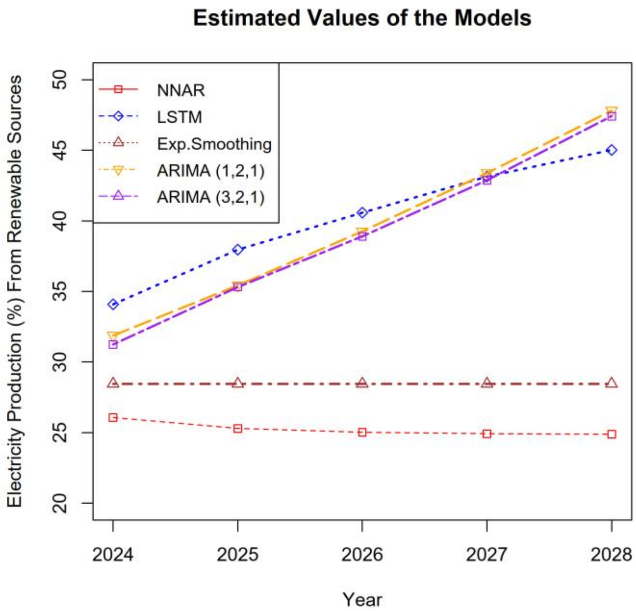

Figure 6. Estimated Values of the Models versus the Actual Values

6. Conclusions and Discussion

We are witnessing the escalating impacts of global warming each day, and it is evident that carbon dioxide emissions, in particular, are the primary cause of this situation. One of the most effective methods for mitigating this effect is to accelerate the adoption of renewable energy resources. Understanding the potential of existing renewable energy sources and assessing the degree to which this potential is already being harnessed will provide valuable insights into the broader landscape. Additionally, for planners and strategists, envisioning the future utilization of renewable energy sources is a crucial consideration. Examining our renewable energy potential and strategizing its future utilization allows us to formulate a comprehensive plan. In this study reviews various estimation approaches for forecasting the future utilization of renewable energy sources and these approaches can be applied to other countries and regions.

The dataset includes information on the electricity generated from renewable sources in Turkey spanning 64 years (1960–2023). For modeling purposes, the initial 59 years of data were used as training data, while the remaining 5 years served as test data. Various time series estimation methods were applied to the dataset, revealing that the NNAR method, a type of artificial neural network method, demonstrated the best performance. However, NNAR consistently iterated a constant value for its future predictions. In contrast, the series exhibited an upward trend, which NNAR failed to capture. The NNAR model makes predictions by establishing a linear relationship from past observations. This may have resulted in an inability to adequately capture the non-linear dynamics inherent in complex and variable processes such as energy production, which is the focus of this study. As a result, NNAR's linear predictions may fail to account for these complexities, making them less suitable for long-term and precise forecasting. Therefore, LSTM algorithm, which had the second-best prediction results, was considered more suitable for forecasting.

Considering the upward trend, the LSTM, ARIMA (1,2,1) and ARIMA (3,2,1) modeling methods produced the best results. According to our findings, it can be predicted that the share of renewable energy in Turkey's total energy production (excluding hydroelectric) from 2024 to 2028 will fall within the range of 34.09% to 45.02%. The techniques employed in this study can be tested for the quantitative estimation of both underground and surface resources. In similar tests, researchers may opt for LSTM and ARIMA modeling methods.

Unlike previous studies that focused on specific regions or types of renewable energy, this study aimed to forecast the share of all renewable energy production in Turkey, excluding hydroelectric power. Furthermore, the study employs eight distinct forecasting methods, thereby offering a more comprehensive understanding of the predictive models applied in the field. Notably, the study validates the effectiveness of well-performing algorithms such as ARIMA and ELM, further contributing to the empirical knowledge base in renewable energy forecasting.

Furthermore, the fact that these forecasting methods can be applied to other resource types and geographical areas implies that comparable approaches can be applied to provide reliable energy projections on a global scale. For their forecasting requirements, researchers and practitioners might investigate the usage of LSTM and ARIMA models. Researchers by modifying the models to take into consideration local characteristics and data accessibility, they can improve future predictions.

In conclusion, by highlighting the advantages of both conventional and modern modeling approaches and offering useful data for future studies and policy formulation, this work seeks to add to the body of knowledge on renewable energy forecasting.

Acknowledgments

This study is part of the Master of Science thesis of Mehmet Berke ÇOLAK, conducted within the Institute of Natural and Applied Science at Tekirdağ Namık Kemal University, under the guidance of thesis advisor Erkan ÖZHAN. The authors would like to express their gratitude to the Institute for its support and valuable contributions. Additionally, the authors extend thanks to the World Bank, the Ministry of Energy of Türkiye, and YTBS-TEIAS for providing the dataset used in this study.

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Ethical approval and informed consent

During the preparation process of this study, scientific and ethical principles were followed, and all the studies included in the study were provided in the bibliography.

Author Contributions

The contributions of the authors to this study are equal.

Funding

This research received no external funding.

Availability of data and material

The dataset used in this study is publicly available and can be accessed and downloaded from the following GitHub repository: https://github.com/erkanozhan/renewable_energy. The repository contains the complete dataset in CSV format, along with detailed documentation and scripts for data analysis. Researchers are encouraged to explore, utilize, and contribute to the repository for further studies.

References

Akusok, A., 2016. Extreme Learning Machines : novel extensions and application to Big Data. University of Iowa Iowa Research Online,A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Industrial Engineering in the Graduate College of The University of Iowa.

Baccar, Y.B., 2019. Comparative Study on Time Series Forecasting Models. Master of Science (Data Science)Advisor: Bertrand Lamy, Jacques Doan HUU 1–92. https://doi.org/10.13140/RG.2.2.32241.02408

Baskan, S., 2008. Effect Of Ligand Binding On Protein Dynamics : A Time Series Analysis. Bogazici University 77.

Bouquet, P., Jackson, I., Nick, M., Kaboli, A., 2024. AI-based forecasting for optimised solar energy management and smart grid efficiency. International Journal of Production Research 62, 4623–4644. https://doi.org/10.1080/00207543.2023.2269565

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 1994. Time Series Analysis Forecasting and Control.

Cakir, S., 2023. Renewable energy generation forecasting in Turkey via intuitionistic fuzzy time series approach. Renewable Energy 214, 194–200. https://doi.org/10.1016/j.renene.2023.05.132

Çetin, Ö., Işık, A.H., 2021. Monthly Electricity Generatıon Forecast in Solar Power Plants with LSTM. Düzce Üniversitesi Bilim ve Teknoloji Dergisi 9, 55–64. https://doi.org/10.29130/dubited.1015251

Daniyal, M., Tawiah, K., Muhammadullah, S., Opoku-Ameyaw, K., 2022. Comparison of Conventional Modeling Techniques with the Neural Network Autoregressive Model (NNAR): Application to COVID-19 Data. Journal of Healthcare Engineering 2022, 1–9. https://doi.org/10.1155/2022/4802743

Erdem, K., 2020. Introduction to Extreme Learning Machines | by Kemal Erdem (burnpiro) | Towards Data Science.

Ertürk, S., Kara, H., Akkus, C., Genc, G., 2023. Turkiye'de Farklı İklim Kuşakları İçin Yapay Sinir Ağları Kullanılarak Güneş Isınımının Tahmini. Gazi University Journal of Science Part C: Design and Technology 11, 885–892. https://doi.org/10.29109/gujsc.1331788

Ghislieri, M., Cerone, G.L., Knaflitz, M., Agostini, V., 2021. Long short-term memory (LSTM) recurrent neural network for muscle activity detection. Journal of NeuroEngineering and Rehabilitation 18, 1–15. https://doi.org/10.1186/s12984-021-00945-w

Gibson, K., 2020. The Application Of Machine Learning For Grounwater Level Prediction In The Steenkoppies Compartment Of The Gauteng And North West Dolomite Aquifer , South Africa.

Golestaneh, F., Pinson, P., Gooi, H.B., 2016. Very short-term nonparametric probabilistic forecasting of renewable energy generation - With application to solar energy. IEEE Transactions on Power Systems 31, 3850–3863. https://doi.org/10.1109/TPWRS.2015.2502423

Goncalves, C., Bessa, R.J., Pinson, P., 2021. Privacy-preserving Distributed Learning for Renewable Energy Forecasting. IEEE Transactions on Sustainable Energy 3029, 1–10. https://doi.org/10.1109/TSTE.2021.3065117

Guang-Bin Huang, Qin-Yu Zhu, Chee-Kheong Siew, 2004. Extreme learning machine: a new learning scheme of feedforward neural networks, in: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541). Presented at the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), IEEE, Budapest, Hungary, pp. 985–990. https://doi.org/10.1109/IJCNN.2004.1380068

Güllü, M., Kartal, Z., 2021. Türkiye'nin Yenilenebilir Enerji Kaynaklarının 2030 Yılına Kadar Tahmini. 19 Mayis Sosyal Bilimler Dergisi 2, 288–313. https://doi.org/10.52835/19maysbd.849978

Hersh, M.A., 2006. The Economics and Politics of Energy Generation. IFAC Proceedings Volumes 39, 73–78. https://doi.org/10.1016/S1474-6670(17)30097-6

Hocaoglu, F.O., Karanfil, F., 2013a. A time series-based approach for renewable energy modeling. Renewable and Sustainable Energy Reviews 28, 204–214. https://doi.org/10.1016/j.rser.2013.07.054

Hocaoglu, F.O., Karanfil, F., 2013b. A time series-based approach for renewable energy modeling. Renewable and Sustainable Energy Reviews 28, 204–214. https://doi.org/10.1016/j.rser.2013.07.054

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Computation 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jiang, P., Dong, J., Huang, H., 2019. Forecasting China's renewable energy terminal power consumption based on empirical mode decomposition and an improved extreme learning machine optimized by a bacterial foraging algorithm. Energies 12. https://doi.org/10.3390/en12071331

Kaysal, K., Yurttakal, A.H., Hocaoğlu, F.O., 2023. Hibrit derin öğrenme yöntemi kullanılarak hiperparametre optimizasyonu ile yenilenebilir elektrik enerjisi tahmini. Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi 12, 770–777. https://doi.org/10.28948/ngumuh.1263782

Lanovaz, M.J., Adams, B., 2019. Comparing the Communication Tone and Responses of Users and Developers in Two R Mailing Lists: Measuring Positive and Negative Emails. IEEE Software 36, 46–50. https://doi.org/10.1109/MS.2019.2922949

Lindemann, B., Müller, T., Vietz, H., Jazdi, N., Weyrich, M., 2021. A survey on long short-term memory networks for time series prediction. Procedia CIRP 99, 650–655. https://doi.org/10.1016/j.procir.2021.03.088

Maleki, A., Nasseri, S., Aminabad, M.S., Hadi, M., 2018. Comparison of ARIMA and NNAR Models for Forecasting Water Treatment Plant's Influent Characteristics. KSCE Journal of Civil Engineering 22, 3233–3245. https://doi.org/10.1007/s12205-018-1195-z

Mossalam, A., Arafa, M., 2018. Using artificial neural networks (ANN) in projects monitoring dashboards' formulation. HBRC Journal 14, 385–392. https://doi.org/10.1016/j.hbrcj.2017.11.002

Mrutyunjaya, P., 2020. Application of ARIMA and Holt-Winters forecasting model to predict the spreading of COVID-19 for India and its states. Department of Computer and Applications, Utkal University, Vani Vihar, India 14, 1–4.

Nastos, P.T., Moustris, K.P., Larissi, I.K., Paliatsos, A.G., 2013. Rain intensity forecast using Artificial Neural Networks in Athens, Greece. Atmospheric Research 119, 153–160. https://doi.org/10.1016/j.atmosres.2011.07.020

Nurhamidah, N., Nusyirwan, N., Faisol, A., 2020. Forecasting Seasonal Time Series Data Using the Holt-Winters Exponential Smoothing Method of Additive Models. Jurnal Matematika Integratif 16, 151. https://doi.org/10.24198/jmi.v16.n2.29293.151-157

Nyatuame, M., Agodzo, S.K., 2018. Stochastic ARIMA model for annual rainfall and maximum temperature forecasting over Tordzie watershed in Ghana. Journal of Water and Land Development 37, 127–140. https://doi.org/10.2478/jwld-2018-0032

Olah, C., 2015. Understanding LSTM Networks.

Paoli, C., Voyant, C., Muselli, M., Nivet, M.L., 2010a. Forecasting of preprocessed daily solar radiation time series using neural networks. Solar Energy 84, 2146–2160. https://doi.org/10.1016/j.solener.2010.08.011

Paoli, C., Voyant, C., Muselli, M., Nivet, M.L., 2010b. Forecasting of preprocessed daily solar radiation time series using neural networks. Solar Energy 84, 2146–2160. https://doi.org/10.1016/j.solener.2010.08.011

Rajni, Banerjee, T., Kumar, P., 2024. Forecasting of renewable energy production in United States: An ARIMA based time series analysis. AIP Conference Proceedings 3010, 030014. https://doi.org/10.1063/5.0193938

Renewable energy explained - U.S. Energy Information Administration (EIA) [WWW Document], 2023. . EIA. URL https://www.eia.gov/energyexplained/renewable-sources/ (accessed 1.15.24).

Sadia, I., Mahmood, A., Binti Mat Kiah, L., Azzuhri, S.R., 2022. Analysis and Forecasting of Blockchain-based Cryptocurrencies and Performance Evaluation of TBATS, NNAR and ARIMA, in: 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET). Presented at the 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), IEEE, Kota Kinabalu, Malaysia, pp. 1–6. https://doi.org/10.1109/IICAIET55139.2022.9936798

Solano, E.S., Dehghanian, P., Affonso, C.M., 2022. Solar Radiation Forecasting Using Machine Learning and Ensemble Feature Selection. Energies 15, 7049. https://doi.org/10.3390/en15197049

Tharani, K., Kumar, N., Srivastava, V., Mishra, S., Pratyush Jayachandran, M., 2020. Machine learning models for renewable energy forecasting. Journal of Statistics and Management Systems 23, 171–180. https://doi.org/10.1080/09720510.2020.1721636

Van Houdt, G., Mosquera, C., Nápoles, G., 2020. A review on the long short-term memory model. Artif Intell Rev 53, 5929–5955. https://doi.org/10.1007/s10462-020-09838-1

Yang, Q., Wang, J., Ma, H., Wang, X., 2020. Research on COVID-19 based on ARIMA modelΔ—Taking Hubei, China as an example to see the epidemic in Italy. Journal of Infection and Public Health 13, 1415–1418. https://doi.org/10.1016/j.jiph.2020.06.019

Yük Tevzi Bilgi Sistemi (YTBS)-Türkiye Elektrik İstatistikleri [WWW Document], 2023. . Yük Tevzi Bilgi Sistemi (YTBS). URL https://ytbsbilgi.teias.gov.tr/ytbsbilgi/frm_istatistikler.jsf (accessed 1.15.24).