



Estimation of Solar Energy Production Data from Renewable Energy Sources Using ANN Method and Presenting It to the Public Benefit

Mehmet Akif YÜKSEL^{1*}, Esmâ UZUNHİSARCIKLİ², Ramazan ALDEMİR³

¹ Kayseri University, Department of Electrical-Electronics Engineering, 38280, Kayseri, Türkiye

² Yozgat Bozok University, Electrical and Electronics Engineering, Control and Command Systems, 66100, Yozgat, Türkiye

³ Kayseri University, Vocational School of Technical Sciences, Department of Electronics and Automation, 38280, Kayseri, Türkiye

*Correspondence: mehmet.akif.yuksel@hotmail.com

Received: 21.08.2025

Accepted: 18.10.2025

Final Version: 22.10.2025

Abstract

Nowadays, the use of renewable energy sources in electricity generation is gradually increasing compared to non-renewable sources, and artificial intelligence algorithms are effectively used in energy forecasting. In this study, a multi-layer feedforward ANN model was developed using real energy production data from solar panels of two different public enterprises for the years 2019-2022. Hyperparameter optimization was performed, and the Levenberg-Marquardt (LM) algorithm was selected after comparison with other algorithms due to its superior performance. The data were split into 70% for training and 30% for testing, and seasonal prediction analyses were carried out. According to the results obtained, seasonal forecasts made using the ANN method were found to be highly accurate, especially during the summer months.

Keywords: Renewable energy, Solar energy, Artificial neural network.

1. INTRODUCTION

Energy has underpinned human activity throughout history, and its strategic role in a country's economic and social development grows by the day. Beginning in the early 18th century, coal, oil, and natural gas were classified as finite fossil resources—depletion of their reserves threatens both current populations and future generations. Against this backdrop, the contribution of renewables to electricity generation has exhibited a persistent upward trend, making efficient and judicious energy use imperative. This need is also evident in Türkiye; for example, in 2021, 35.42% of the nation's total electricity production originated from renewable sources (TEİAŞ, 2023). While uninterrupted access to electricity enhances user satisfaction, efficiency and the prevention of wasteful consumption are essential for sustaining supply over the long term. As fossil reserves become scarcer and extraction costs rise, renewables—despite their apparent abundance—must be treated as indispensable clean alternatives. A review of recent studies utilizing artificial neural networks (ANNs) reveals that a variety of data types have been employed for forecasting energy production and consumption. For instance, in a study focusing on the Keban and Karakaya Hydroelectric Power Plants in Türkiye, data from 2008–2012 were analyzed, and predictions regarding the plants' energy production and efficiency were made using the ANN method based on variables such as temperature, pressure, humidity, evaporation, and precipitation (İnallı, 2018). Another study developed a demand forecasting approach for the İzmir-Urla region by applying an ANN method implemented in MATLAB, using short-term seven-day data related to total and regional loads (Biçer, 2018). At İzmir Bakırçay University, a study was conducted to forecast the hourly electricity consumption by utilizing the electricity generation data from solar power plants installed on the campus, using meteorological data collected over a total of 28 months within a 3-year period, and applying an ANN model (Orman and Köse, 2022). Additionally, short-term electricity demand in Türkiye was forecast on an hourly, daily, and weekly basis using the ANN method, by combining the country's total electricity consumption data for 2008 and 2009 with meteorological variables for Istanbul, such as temperature and daily sunshine duration (Cihat, 2009). Türkiye's wind energy potential was investigated by applying an ANN-based forecasting model to meteorological datasets gathered from 58 regions. The network was trained with data from 40 sites and validated with data from 18 sites (Aktaş, 2016). Furthermore,

electricity prices in Türkiye's day-ahead market were forecast on an hourly and daily average basis. Using data from 2010, 2011, 2012, and 2013, the hourly and daily average electricity prices for 2014 were predicted through the ANN method (Adalı and Korkmaz, 2015). In summary, recent forecasting studies utilizing the ANN method have been outlined in this manner. In the present study, real solar energy production data collected from public institutions for the periods 2019–2022 and 2020–2022 were used to conduct seasonal electricity generation forecasts using the ANN learning approach.

Table 1. Selected Studies Conducted Using the ANN Method in the Literature

Author/Year	Features Used	Data Time Interval	Method	Predicted Value
İnalı, 2018	Temperature, pressure, humidity, evaporation, precipitation	2008–2012	ANN	Energy production and efficiency
Bıçer, 2018	Total and regional load	7-day short-term	ANN	Demand forecasting
Orman and Köse, 2022	Meteorological data	28 months	ANN	Hourly electricity consumption
Cihat and Korkmaz, 2009	Temperature, sunshine duration	2008–2009	ANN	Hourly, daily, and weekly demand
Aktaş, 2016	Month, latitude, longitude, temperature, humidity, etc.	Annual	ANN	Wind power density
Adalı, 2015	Electricity consumption	2010–2013	ANN	Electricity price forecasting

The primary distinction of this study from similar works in the literature lies in its use of solar energy production data obtained from two different public institutions and the seasonal analysis of these data. In this respect, the study differs from previous research that relied solely on private sector data and presents a unique application with a focus on public benefit. The real-time and long-term nature of the data used (2019–2022) enhances the study's accuracy. The selection of the Levenberg–Marquardt algorithm also contributed to achieving high accuracy levels. However, a limitation of the study is the scarcity of data for the winter season and the absence of meteorological parameters in all models. As a contribution to the literature, this study demonstrates the feasibility of conducting ANN-based analyses using publicly available data and offers valuable input to decision-making processes that can support public energy policies.

2. METHODOLOGY

Today, technology is advancing rapidly, and its constant evolution and adaptability drive continuous research efforts among scientists.

2.1. Artificial Neural Network-ANN

Among the most widely preferred forecasting methods in academic research applications is the artificial neural network (ANN) approach. It is a machine learning method inspired by the structure and functioning of neural cells in the human brain. The complexity of the human brain, its decision-making processes, and reasoning mechanisms have guided the development of ANN architectures. ANNs resemble the human brain in their ability to model nonlinear structures, operate in a parallel distributed manner, learn and generalize from data, adapt to various problems, and tolerate errors (Es et al., 2014). The advantages of ANNs include nonlinearity, generalization, applicability, learning capability, parallelism, error tolerance, the ability to work with incomplete data, and adaptability. However, they also have certain disadvantages: dependence on hardware, the requirement of large datasets, the inability to explain the network's internal behavior, the necessity of trial-and-error to determine the appropriate network structure due to the lack of predefined rules, the absence of a clear criterion for when to stop training, and the difficulty and complexity involved in modelling (Harika and Yalpir, 2021).

2.2. TYPES OF ARTIFICIAL NEURAL NETWORKS (ANNS)

2.2.1. According to Structural Characteristics

ANN structures are classified into two types—feedforward and feedback neural networks—based on the way neurons are interconnected.

2.2.1.1. Feedforward ANN

In the ANN type shown in Figure 2.1, the cells are organized in layers, and the outputs of the neurons in each layer are transmitted as input to the next layer through weighted connections. The data from the input layer are passed through the hidden layers and are processed directly in the output layer without being fed back to the previous layers (Aydemir et al., 2020).

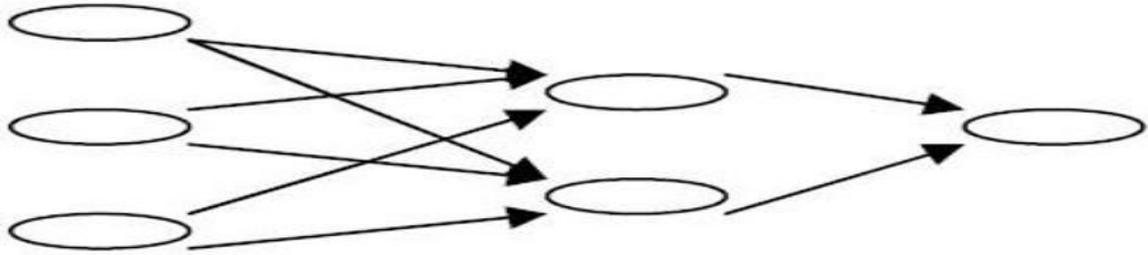


Figure 2.1. Feedforward ANN

2.2.1.2. Feedback ANN

In a feedback ANN, the output of at least one neuron is fed back as input to itself or to other neurons. This feedback typically occurs through a delay element and may take place either between neurons within the same layer or across different layers.

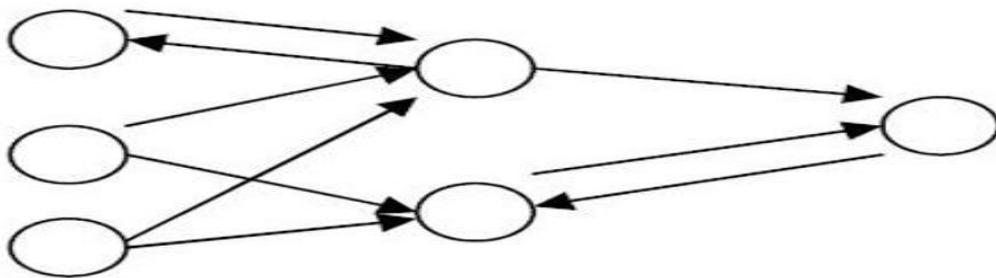


Figure 2.2. Feedback ANN

As shown in Figure 2.2, the output of neurons in a feedback ANN model can serve either as input to the next layer or as input to preceding neurons. Although this type of ANN possesses dynamic memory capabilities, the networks can be structurally complex and powerful. Therefore, such networks are capable of producing realistic and accurate results (Doğan, 2019).

2.2.2 Levenberg–Marquardt (LM) Algorithm

The Levenberg–Marquardt algorithm is notable for providing rapid convergence in the training of feedforward neural networks. Recent studies have confirmed that the LM algorithm offers high accuracy, particularly in regression tasks involving small datasets. Its use in combination with hybrid models has also become increasingly common (Ahmad et al., 2022).

2.2.3 Bayesian Regularization (BR) Algorithm

The Bayesian Regularization method prevents overfitting by optimizing the distribution of weights and error functions. Studies have shown that the BR method produces more stable results and improves generalization performance, particularly in models working with limited data (Zhao and Yang, 2023).

2.2.4 Scaled Conjugate Gradient (SCG) Algorithm

The Scaled Conjugate Gradient (SCG) algorithm is a method that enables fast convergence without requiring second-order derivative information. Studies have indicated that this algorithm is particularly suitable for large datasets and is notable for its low memory usage. However, in certain scenarios, its accuracy has been reported to be lower compared to the Levenberg–Marquardt algorithm (Li and He, 2021).

3. TEST STUDIES AND ANALYSIS

3.1 Test Studies

The datasets used in this study consist of actual solar energy production data collected from Facility A for the period 2019–2022 and from Facility B for the period 2020–2022. These data were collected on a daily basis. A total of 8,112 records were obtained from Facility A, and 1,062 records from Facility B. The training/testing split was set at 70% and 30%, respectively. Dividing the dataset into 70% for training and 30% for testing is a widely adopted approach in ANN-based forecasting studies, as it enables an objective assessment of the model's generalization capability (Aktaç, 2016; Biçer, 2018). This ratio ensures sufficient data for training while allowing the model's independent accuracy to be evaluated on the test set. A feedforward multilayer artificial neural network was used as the network architecture. A single hidden layer with 10 neurons was employed. This structure is commonly preferred in energy forecasting studies in the literature, as it prevents the model from becoming overly complex while maintaining adequate learning capacity. Moreover, different numbers of layers and neurons were tested, and the lowest error and highest accuracy were achieved with 10 hidden neurons. This approach enhances the generalization capability of the ANN and helps to prevent overfitting (Ahmad et al., 2022). All datasets were scaled to the range of 0–1 using min-max normalization before being fed into the ANN model. This preprocessing step facilitates faster and more stable learning and is a standard procedure recommended in the literature (Biçer, 2018; Orman and Köse, 2022). In order to maximize model performance, a systematic optimization was carried out on hyperparameters such as the number of hidden layers, the number of neurons, and the learning rate. As suggested in the literature, different combinations were tested, and the configuration yielding the best results was selected (Ahmad et al., 2022; Zhao and Yang, 2023). A sigmoid activation function was used in the hidden layer, while a linear activation function was employed in the output layer. The algorithm parameters were set according to the default settings in MATLAB (Version 9.13.0.2193358, MathWorks, USA), and three different algorithms—Levenberg–Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient—were compared.

Table 3.1. Actual Data Values from Facility A

Season–Month		Start Date	End Date	Total Number of Records
Winter	DECEMBER	01.01.2019	31.12.2022	1632
	JANUARY	01.01.2019	31.12.2022	
	FEBRUARY	01.01.2019	31.12.2022	
Spring	MARCH	01.01.2019	31.12.2022	2160
	APRİL	01.01.2019	31.12.2022	
	MAY	01.01.2019	31.12.2022	
Summer	JUNE	01.01.2019	31.12.2022	2160
	JULY	01.01.2019	31.12.2022	
	AUGUST	01.01.2019	31.12.2022	
Autumn	SEPTEMBER	01.01.2019	31.12.2022	2160
	OCTOBER	01.01.2019	31.12.2022	
	NOVEMBER	01.01.2019	31.12.2022	
TOTAL				8112

The actual energy production data for Facilities A and B are presented in Table 3.1 and Table 3.2, respectively. In the analyses conducted, the data were divided as 70% training and 30% testing, and the number of hidden layer neurons was set to 10. Among the algorithms evaluated—Levenberg–Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient—the Levenberg–Marquardt algorithm demonstrated the best performance and was therefore selected. The performance comparison results of these three algorithms are provided in Table 3.3.

Table 3.2. Actual Data Values from Facility B

Season–Month		Start Date	End Date	Total Number of Records
Winter	DECEMBER	01.01.2020	31.12.2022	252
	JANUARY	01.01.2020	31.12.2022	
	FEBRUARY	01.01.2020	31.12.2022	
Spring	MARCH	01.01.2020	31.12.2022	270
	APRIL	01.01.2020	31.12.2022	
	MAY	01.01.2020	31.12.2022	
Summer	JUNE	01.01.2020	31.12.2022	270
	JULY	01.01.2020	31.12.2022	
	AUGUST	01.01.2020	31.12.2022	
Autumn	SEPTEMBER	01.01.2020	31.12.2022	270
	OCTOBER	01.01.2020	31.12.2022	
	NOVEMBER	01.01.2020	31.12.2022	
TOTAL				1062

Table 3.3. Algorithm Performance Comparison Results

ALGORİTM		TRANİNG		TEST	
		MSE	R	MSE	R
Scaled Conjugate Gradient		3.9133e+05	-0.0224	2.0145e+05	0.7430
Levenberg-Marquardt		4.7906e+05	0.5894	4.0227e+04	0.9541
Bayesian Regularization		5.5566e+05	0.2710	2.6135e+04	0.9719

As shown in the tables, the number of data records obtained from Facility A is greater than that of Facility B. This is because the data collected from Facility A were sourced from multiple solar energy production plants.

Metrics

Mean Squared Error (MSE)

Mean Squared Error is the average of the squared differences between the actual and predicted values. The formula is presented in Equation (1);

Here, \hat{y}_i represents the predicted value of y_i ,

e_t denotes the error term.

$$e_t = y_i - \hat{y}_i$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (e_t)^2 \tag{1}$$

Root Mean Square Error (RMSE)

Root Mean Square Error indicates the standard deviation of the predicted error magnitude. It reflects how much the predicted values deviate from the actual ones. The formula is presented in Equation (2);

Here, \hat{y}_i represents the predicted value of y_i ,

e_t denotes the error term.

$$e_t = y_i - \hat{y}_i$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_t^2} \tag{2}$$

R-Squared (R²) Value

The R-squared value represents the percentage of variance in the outcome variable that is explained by a linear model. It always ranges between 0 and 1. A value of 0 indicates that the model explains none of the variability in the data, whereas a value of 1 indicates that the model explains all the variability in the data (Erbudak, 2022). The formula is presented in Equation (3).

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{3}$$

Table 3.4. Table display of metrics and formulas

Metrics	Formulas
Root Mean Square Error- RMSE	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_t^2}$
Mean Square Error- MSE	$MSE = \frac{1}{n} \sum_{i=1}^n (e_t)^2$
R-Squared (R ²) Value	$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$

3.2. Analyses

The data analyzed on a seasonal basis for a 4-year period belongs to Facility A. The input values consist of data from January and February for the years 2019–2022, and from December for the years 2019–2021. The December 2022 value was designated as the test data.

Similarly, seasonal analysis was conducted on a total of 3 years of data for Facility B. The input values included data from January and February for the years 2020–2022, and from December for the years 2020 and 2021. The December 2022 value was entered as the test data. All seasonal analysis results for Facility A and Facility B are presented respectively in Figures 3.1 and 3.2 below.

REGRESSION ANALYSIS GRAPHS FOR ALL SEASONS – FACILITY A

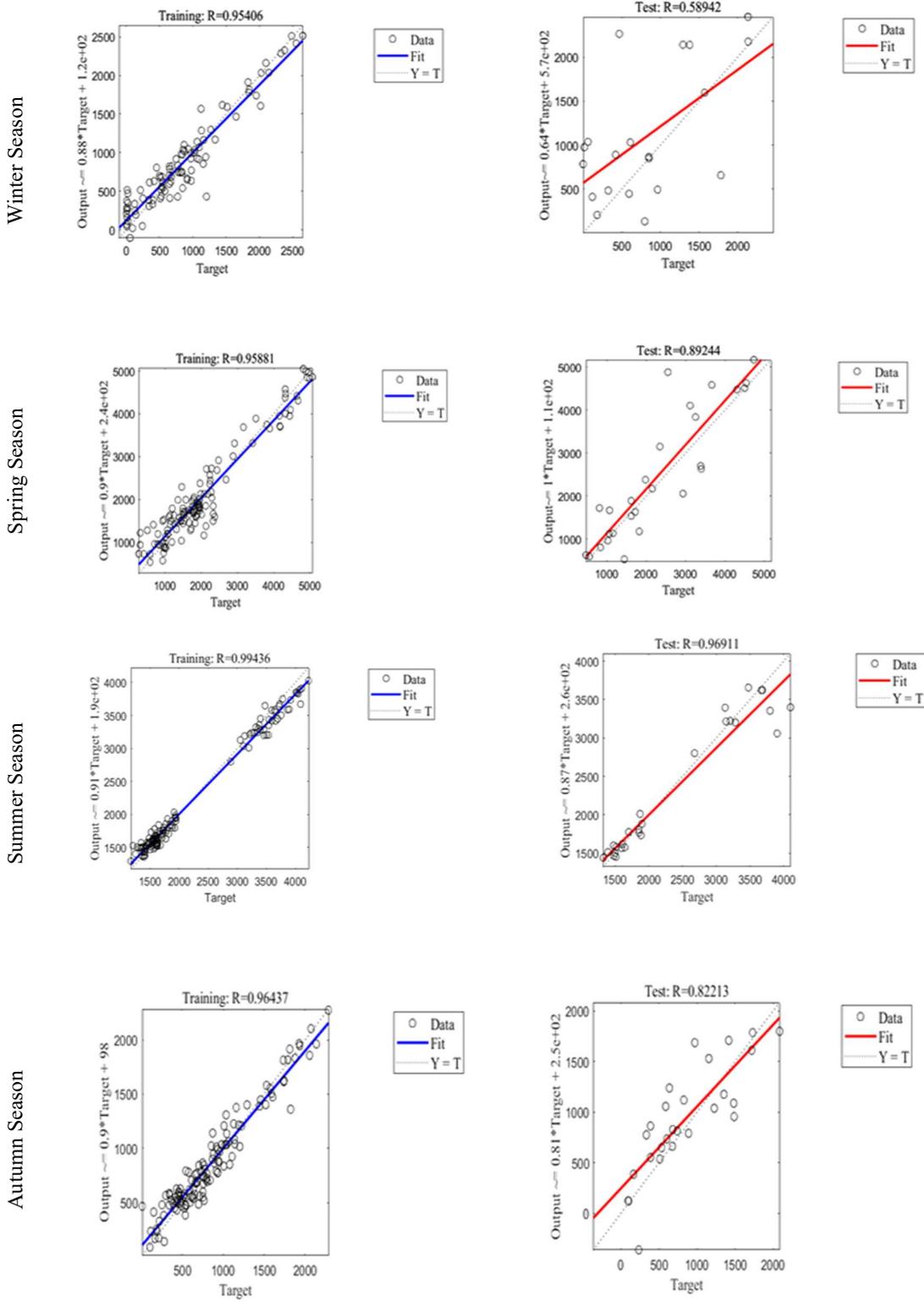


Figure 3.1. Analysis Results for All Seasons – Facility A

REGRESSION ANALYSIS GRAPHS FOR ALL SEASONS – FACILITY B

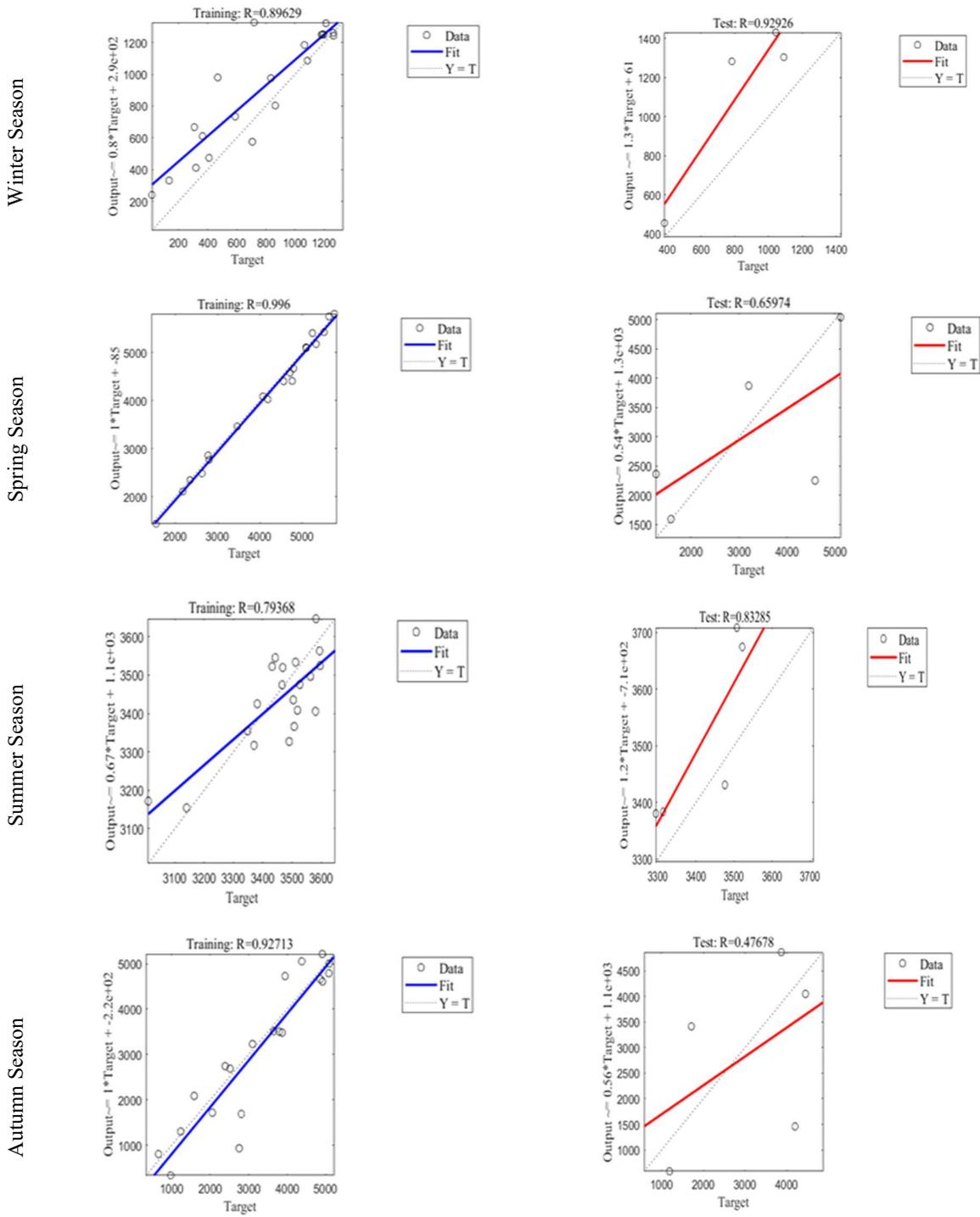


Figure 3.2. Analysis Results for All Seasons – Facility B

For Facility A and Facility B, the seasonal analysis and testing studies conducted using the ANN method are summarized separately in Tables 3.5 and 3.6.

Table 3.5. Training Results for All Seasons – Facility A

SEASON	VALUE NAME	OBSERVATION	MSE	R
AUTUMN	Training	1512	1.7514e+04	0.9644
	Test	648	1.0857e+05	0.8221
SUMMER	Training	1512	1.6727e+04	0.9944
	Test	648	6.1014e+04	0.9691
SPRING	Training	1512	1.1729e+05	0.9588
	Test	648	4.8258e+05	0.8924
WINTER	Training	1142	4.0227e+04	0.9541
	Test	490	4.7906e+05	0.5894

Table 3.6. Training Results for All Seasons – Facility B

SEASON	VALUE NAME	OBSERVATION	MSE	R
AUTUMN	Training	189	3.5931e+05	0.9271
	Test	81	2.3854e+06	0.4768
SUMMER	Training	189	8.2041e+03	0.7937
	Test	81	1.5609e+04	0.8329
SPRING	Training	189	1.6476e+04	0.9960
	Test	81	1.4159e+06	0.6597
WINTER	Training	176	5.0956e+04	0.8963
	Test	76	1.1284e+05	0.9293

4. CONCLUSION

In this study, solar energy production data were predicted using the Artificial Neural Network (ANN) method. The prediction results based on data obtained from two different publicly owned power plants demonstrated that the Levenberg-Marquardt algorithm provides high accuracy. The findings indicate that the ANN-based model can serve as an effective tool in solar energy production planning and contribute to strategic planning efforts for public institutions.

ANALYSES

The data presented in Tables 3.3 and 3.5 highlight that the number of data values for the winter season is noticeably lower compared to other seasons. Two primary factors account for this situation. The first is that the length of February differs by year, containing either 28 or 29 days. The second reason is that the number of data collected from the facilities during the winter months is generally lower due to breakdowns, maintenance, and other environmental factors. However, it has been observed that with rising temperatures and more favorable weather conditions, such issues decrease, and the data are collected more reliably. The limited and insufficient actual energy production data in the winter months do not yield favorable results. In general, the prediction results based on data obtained during March, April, May, June, July, and August were found to be more successful compared to those from September, October, November, December, January, and February. Therefore, it is suggested that future studies not only increase the number of historical data points used but also ensure higher data quality. Incorporating meteorological variables such as cloud cover, temperature, pressure, and precipitation is expected to enhance the reliability of the results. Moreover, the use of different methods and machine learning algorithms is recommended, as they may yield improved analysis outcomes. According to the obtained results, it was observed that the training performed with the Levenberg-Marquardt algorithm yielded low average error values and R (correlation coefficient) values above 0.98. This indicates that the model is capable of making highly accurate predictions. The MSE (Mean Squared Error) and R (Linear Correlation Coefficient) values presented in Tables 3.5 and 3.6 demonstrate both the prediction accuracy of the model and that the estimations were performed using actual data. In particular, the low MSE and high R values observed during the summer months indicate that the model successfully captures the linear relationship (Es et al., 2014). Furthermore, the low MSE and high R values, especially during the summer months, indicate that the model performs better on a seasonal basis. In the regression analysis graphs shown in Figures 3.1 and 3.2, the horizontal axis represents the actual production values, while the vertical axis shows the values predicted by the ANN model. The

clustering of data points along a line close to 45 degrees in the graphs indicates that the model provides highly accurate predictions (Orman and Köse, 2022). To illustrate the uncertainty of the prediction results, a 95% confidence interval was calculated for the test set predictions, and it was observed that the predictions fell within the ± 1.96 *standard error range. This approach supports the statistical significance and reliability of the model's predictions (Zhao and Yang, 2023). The presented MSE and R values reveal that the Levenberg-Marquardt algorithm outperforms the other algorithms. The Bayesian Regularization algorithm provided high generalizability, albeit with a longer training time. Although the Scaled Conjugate Gradient algorithm had a shorter training duration, its accuracy level was relatively lower. Based on these findings, the Levenberg-Marquardt algorithm yielded the most suitable results for the dataset and problem type used in this study.

ACKNOWLEDGMENT

I would like to thank Prof. Dr. Esma Uzunhisarcıklı and Dr. Ramazan Aldemir for their contributions to this study.

AUTHOR'S CONTRIBUTIONS

The authors contributed equally.

CONFLICTS OF INTEREST

There is no conflict of interest.

RESEARCH AND PUBLICATION ETHICS

The author declares that this study complies with Research and Publication Ethics.

REFERENCES

- Adalı, E.Y., (2015). Gün Öncesi Piyasasında Saatlik ve Günlük Elektrik Fiyatları Tahmini, Master Thesis, İstanbul Technical University Energy Institute, İstanbul, Turkey.
- Ahmad, F., Khalid, S., & Khan, M. A., (2022). Enhanced solar radiation forecasting using hybrid Levenberg–Marquardt trained neural networks. *Neural Computing and Applications*, 75(1), 1783-1800, DOI:10.32604/nca.2023.035736
- Aktaş, M.Y., (2016). Rüzgâr Güç Yoğunluğunun Yapay Sinir Ağı (YSA) ile Tahmini, Master Thesis, Siirt Üniversitesi, Institute of Science, Siirt, Turkey.
- Aydemir, M., (2020). Yapay Sinir Ağları ile Bütçe Gelirlerinin Tahmini, Master Thesis, İnönü University, Institute of Social Sciences, Malatya, Turkey.
- Bıçer, A., (2018). Enerji Talep Tahminine Yönelik Program Geliştirme ve Bir Bölge İçin Uygulaması, Master Thesis, Afyon Kocatepe Üniversitesi Institute of Science, Afyon, Turkey
- Cihat, A., Korkmaz, O., (2009). Türkiye Kısa Süreli Elektrik Talebinin Saatlik Olarak Tahmin Edilmesi. *Journal of Yasar University, Enerji*, İstanbul, Turkey
- Doğan, Z., (2019). Nötral Tipteki Zaman Gecikmeli Dinamik Yapay Sinir Ağlarının Kararlılık Özelliklerinin İncelenmesi, Master Thesis, İstanbul University-Cerrahpaşa Graduate Education Institute, İstanbul, Turkey.
- Erbudak, A. E., (2022). Veri Madenciliği ve Makine Öğrenimi ile Döviz Kuru Tahmini Uygulaması, Master Thesis, Altınbaş University, Institute of Graduate Education, İstanbul, Turkey.
- Es, H. A., Kalender, F. Y., Hamzaçebi, C., (2014). Yapay sinir ağları ile Türkiye net enerji talep tahmini, *Journal of the Faculty of Engineering and Architecture of Gazi University*, Vol 29, No 3, 495-504, Doi:10.17341/gummfd.41725

Harika, Ü. L. K. Ü., Yalçır, Ş., (2021). Enerji talep tahmini için metodoloji geliştirme: 2030 yılı Türkiye örneği. Niğde Ömer Halisdemir University Journal of Engineering Sciences,10(1), 188-201., Doi:10.28948/ngumuh.814134.

İnalı, K., (2018). Keban ve Karakaya HES’de Enerji Veriminin İklim Parametrelerine Bağlı Olarak Akıllı Sistemlerle Analizi, Master Thesis, Fırat Üniversitesi Institute of Science, Elazığ, Turkey

Li, Z., & He, Y. A Comparative Study of Scaled Conjugate Gradient and Adam for Deep Learning Optimization. Journal of Intelligent & Fuzzy Systems, 40(2), 2781–2792, 2021

Orman, E., Köse, B., (2022). İzmir Bakırçay üniversitesi güneş enerji santralinin elektrik üretiminin saatlik tahmini. (ICADA’22) 2nd International Conference on Artificial Intelligence and Data Science, (pp1), Balıkesir, Turkey

TEİAŞ, “İstatistikler” <https://www.teias.gov.tr/turkiye-elektrik-uretim-iletim-istatistikleri> [Access Date:13-March-2023].

Zhao, H., & Yang, L. Bayesian Regularization Neural Networks for Robust Indoor Air Quality Prediction. Environmental Modelling & Software, 158, 105617, 2023