Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

GU J Sci, Part C, 14(1): 318-327 (2026)

Machine Learning – Based Prediction of Daily Global Solar Radiation Using Local Meteorological Observations

İbrahim Çağrı BARUTÇU[1*]

[1]*Hakkari University, Çölemerik Vocational School, Department of Electricity and Energy, Hakkari, Turkey*

Article Info

Research article
Received: 26/11/2025
Revision: 08/01/2026
Accepted: 15/01/2026

Keywords

Solar Radiation
Machine Learning
Meteorological Data
Prediction

Makale Bilgisi

Araştırma makalesi
Başvuru: 26/11/2025
Düzeltme: 08/01/2026
Kabul: 15/01/2026

Anahtar Kelimeler

Güneş Radyasyonu
Makine Öğrenme
Meteorolojik Veri
Tahmin

Graphical/Tabular Abstract (Grafik Özet)

The study predicts daily global horizontal irradiance (GHI) using 2015–2022 daily meteorological measurements from Hakkari (RH, wind speed, temperature, atmospheric pressure, sunshine duration) and compares SVR, RF, XGBoost, and LR with R^2, MAE, RMSE, and R metrics. / Bu çalışma, Hakkâri'de 2015–2022 arasında günlük ölçülen bağıl nem, rüzgâr hızı, sıcaklık, atmosfer basıncı ve güneşlenme süresi gibi verilerle günlük toplam küresel güneş radyasyonunu (GHI) tahmin etmektedir. SVR, RF, XGBoost ve LR modellerini R^2, MAE, RMSE ve R ölçütleriyle karşılaştırmaktadır.

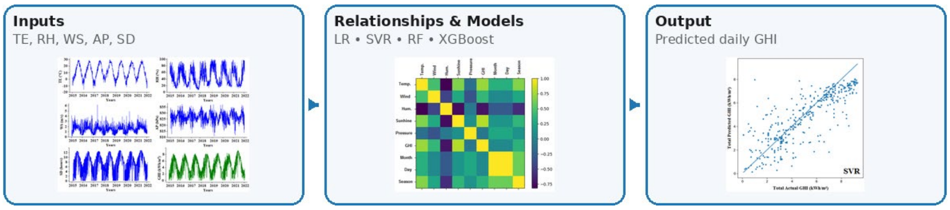

***Figure A**: Graphical abstract of the study / **Şekil A**:.Çalışmanın grafik özeti*

Highlights (Önemli noktalar)

- *GHI was predicted using 2015–2022 local meteorological observations in Hakkari. / Hakkâri'de 2015–2022 dönemine ait yerel meteorolojik gözlemler kullanılarak GHI tahmin edilmiştir.*
- *SVR, RF, XGBoost and LR models were compared using R^2, RMSE, MAE and R metrics. / SVR, RF, XGBoost ve LR modelleri, R^2, RMSE, MAE ve R ölçütleriyle karşılaştırılmıştır.*
- *Sunshine duration, temperature and atmospheric pressure were the most effective determinants of daily solar radiation. / Güneşlenme süresi, sıcaklık ve atmosfer basıncı, günlük güneş radyasyonunun en etkili belirleyicileri olmuştur.*
- *Findings supported reliable, cost-effective prediction for PV planning in high-altitude, variable-climate regions. / Bulgular, yüksek rakımlı ve değişken iklimli bölgelerde PV planlaması için güvenilir ve düşük maliyetli tahmini desteklemiştir.*

Aim (Amaç): *This study aims to predict GHI in Hakkari using 2015–2022 daily meteorological variables (RH, WS, TE, AP, SD) and to compare the performance of SVR, RF, XGBoost, and LR models. / Bu çalışma, Hakkâri'de 2015–2022 arasında günlük ölçülen meteorolojik değişkenleri (RH, WS, TE, AP, SD) kullanarak günlük toplam küresel güneş radyasyonunu tahmin etmeyi ve SVR, RF, XGBoost ve LR modellerinin performansını karşılaştırmayı amaçlamaktadır.*

Originality (Özgünlük): *The paper's originality lies in filling a literature gap by developing and benchmarking multiple machine-learning models for daily GHI prediction in the high-altitude, topographically complex Hakkari region using daily meteorological data and analyzing key driving variables. / Makalenin özgünlüğü, literatürde sınırlı çalışılan Hakkâri gibi yüksek rakımlı ve karmaşık topografyalı bir bölgede, günlük meteorolojik verilerle günlük GHI tahmini için birden fazla makine öğrenmesi modelini karşılaştırmasıdır.*

Results (Bulgular): *Results indicate sunshine duration, temperature, and atmospheric pressure are the most influential variables, and SVR performs best (R^2=0.703), showing reliable and cost-effective prediction is feasible even under Hakkari's high-altitude, variable climate. / Sonuçlar, en etkili değişkenlerin güneşlenme süresi, sıcaklık ve atmosfer basıncı olduğunu; ayrıca SVR'nin (R^2=0.703) en iyi performansı vererek yüksek rakım ve değişken iklime rağmen yerel verilerle güvenilir ve düşük maliyetli tahmin yapılabileceğini gösterdiğini ortaya koymaktadır.*

Conclusion (Sonuç): *SVR best predicted daily GHI, mainly driven by sunshine duration, temperature, and pressure. / SVR, özellikle güneşlenme süresine, sıcaklığa ve basınca bağlı olan GHI'yı en iyi şekilde tahmin etmiştir.*

*Corresponding author, e-mail: ibrahimcagribarutcu@hakkari.edu.tr
DOI: 10.29109/gujsc.1830524

GU J Sci, Part C, 14(1): 318-327 (2026)

Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

http://dergipark.gov.tr/gujsc

Machine Learning – Based Prediction of Daily Global Solar Radiation Using Local Meteorological Observations

İbrahim Çağrı BARUTÇU[1*]

[1]*Hakkari University, Çölemerik Vocational School, Department of Electricity and Energy, Hakkari, Turkey*

Article Info

Research article
Received: 26/11/2025
Revision: 08/01/2026
Accepted: 15/01/2026

Keywords

Solar Radiation
Machine Learning
Meteorological Data
Prediction

Abstract

This study aims to predict the daily total global solar irradiance (GHI) utilizing machine learning methods using meteorological indicators such as relative humidity (RH), wind speed (WS), temperature (TE), atmospheric pressure (AP), sunshine duration (SD) measured daily between 2015 and 2022 at the station within the Hakkari Provincial Directorate of Meteorology. Solar irradiance prediction is of great importance, especially in terms of determination of the solar energy potential, photovoltaic (PV) system performance analysis, and energy planning. In this context, Extreme Gradient Boosting (XGBoost), random forest (RF), support vector regression (SVR), linear regression (LR) algorithms were applied, and the prediction successes of the models were compared using coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (R) performance criteria. The findings denoted that SD, TE and AP were the most effective determinants of daily solar radiation. Furthermore, among the four different algorithms used in the GHI estimation study, the SVR model was found to exhibit superior performance compared to the other models, with R^2=0.703, RMSE=1.381, MAE=0.983, and R=0.838. In the LR model, R^2 was calculated as 0.685, RMSE as 1.422, MAE as 1.082, and R as 0.827. The RF and XGBoost algorithms, respectively, showed similar performance with R^2=0.654, RMSE=1.491, MAE=1.061, R=0.808 and R^2=0.646, RMSE=1.509, MAE=1.094, R=0.803. The findings demonstrate that, despite Hakkari's high altitude and variable climate conditions, reliable and cost-effective solar radiation prediction models can be developed using local meteorological data, making significant contributions to renewable energy applications.

Yerel Meteorolojik Gözlemler Kullanılarak Günlük Küresel Güneş Radyasyonunun Makine Öğrenmesine Dayalı Tahmini

Makale Bilgisi

Araştırma makalesi
Başvuru: 26/11/2025
Düzeltme: 08/01/2026
Kabul: 15/01/2026

Anahtar Kelimeler

Güneş Radyasyonu
Makine Öğrenme
Meteorolojik Veri
Tahmin

Öz

Bu çalışmada, Hakkâri ili meteoroloji il müdürlüğü bünyesinde bulunan istasyondan 2015–2022 yılları arasında günlük olarak ölçülen bağıl nem (RH), rüzgâr hızı (WS), sıcaklık (TE), atmosfer basıncı (AP), güneşlenme süresi (SD) gibi meteorolojik göstergeler kullanılarak, günlük toplam küresel güneş radyasyonunun (GHI) makine öğrenmesi yöntemleri ile tahmin edilmesi amaçlanmıştır. Güneş radyasyonu tahmini, özellikle güneş enerjisi potansiyelinin belirlenmesi, fotovoltaik (PV) sistem performans analizleri ve enerji planlaması açısından büyük bir öneme sahiptir. Bu kapsamda, Extreme Gradient Boosting (XGBoost), rastgele orman (RF), destek vektör regresyonu (SVR), doğrusal regresyon (LR) algoritmaları uygulanmış; modellerin tahmin başarıları coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE) ve correlation coefficient (R) performans kriterleri ile karşılaştırılmıştır. Sonuçlar, SD, TE ve AP'nin günlük güneş radyasyonunun en etkili belirleyicileri olduğunu göstermiştir. Ayrıca GHI tahmin çalışmasında kullanılan dört farklı algoritma arasından SVR modelinin R^2=0.703, RMSE=1.381, MAE=0.983 ve R=0.838 ile diğer modellere kıyasla daha üstün performans sergilediği belirlenmiştir. LR modelinde R^2=0.685, RMSE=1.422, MAE=1.082 ve R=0.827 olarak hesaplanmıştır. RF ve XGBoost algoritmaları ile sırasıyla hesaplanan R^2=0.654, RMSE=1.491, MAE=1.061, R=0.808 ve R^2=0.646, RMSE=1.509, MAE=1.094, R=0.803 değerleri birbirine yakın performans göstermiştir. Elde edilen bulgular, Hakkâri'nin yüksek rakımlı ve değişken iklim koşullarına rağmen yerel meteorolojik verilerin kullanılmasıyla güvenilir ve düşük maliyetli güneş radyasyonu tahmin modellerinin geliştirilebileceğini ortaya koymakta ve yenilenebilir enerji uygulamalarına önemli katkılar sunmaktadır.

*Corresponding author, e-mail: ibrahimcagribarutcu@hakkari.edu.tr
DOI: 10.29109/gujsc.1830524

1. INTRODUCTION (GİRİŞ)

Solar radiation prediction has become a critical research area for accurately determining solar energy potential, modeling PV energy production, and designing efficient energy systems. Increasing renewable energy demand and efforts to combat climate change have significantly increased scientific interest in data-driven approaches to solar radiation prediction [1]. The proliferation of machine learning has significantly improved prediction accuracy, indicating that machine learning-based models will be becoming increasingly dominant in the field of solar irradiance prediction [2].

Machine learning methods offer advantages such as capturing nonlinear relationships in solar radiation predicting, modeling interactions between different meteorological variables, and delivering higher performance on high-dimensional datasets. These models are particularly effective in uncovering relationships among parameters such as TE, RH, WS, SD [3, 4]. Furthermore, multi-model approaches and hybrid predicting structures are also recognized in the literature as important tools for improving prediction performance [5].

Hyperparameter optimization is a crucial component that directly affects model performance, particularly in tree-based (RF) algorithms and boosting (XGBoost) methods. The XGBoost model is frequently preferred for solar radiation prediction, and various studies have demonstrated that prediction accuracy can be increased with correct hyperparameter settings [6–8]. Furthermore, appropriate input variable selection is also an important factor affecting the success of prediction models. It has been shown in the literature that more optimized prediction models can be developed by eliminating unnecessary or low-contribution variables using methods such as entropy theory and principal component analysis (PCA) [9, 10].

Studies conducted in different geographical conditions indicate that the impact of meteorological parameters on solar radiation can vary depending on regional dynamics. Studies conducted in regions with high insolation potential, such as Saudi Arabia and India, have shown that temperature and sunshine duration are the most decisive variables [11, 12]. Studies conducted specifically for Turkey have indicated that the effects of humidity and wind vary seasonally and vary according to regional geographical conditions [13]. Furthermore, the higher accuracy of ensemble models compared to single regression approaches is clearly evident in various comparative studies [14, 15].

A general trend in the literature suggests that machine learning methods offer an effective approach to solar radiation prediction; however, model success varies depending on regional climate characteristics, the quality of the dataset, and the hyperparameter settings of the algorithm used [16, 17]. Studies conducted in high-altitude, topographically complex regions in Turkey are limited, and new data-driven modeling approaches in such regions fill a significant gap in the literature. In that manner, solar irradiance prediction studies conducted in regions with high altitude, harsh climate, and geographical variability, such as the Hakkari Meteorological Station, will make valuable contributions to both Turkish and international research. These contributions can be listed as follows:

- For strengthening scientific validity for data used within the study, an original study based on verified daily meteorological data obtained from the official institution was presented.
- Data-driven solar radiation prediction models were investigated for a region like Hakkari, which has a high altitude and complex topography. This high altitude causes rapid temperature fluctuations and significant atmospheric pressure changes throughout the day. These characteristics make solar radiation modeling, particularly in Hakkari, a complex but scientifically valuable problem.
- By comparing multiple machine learning methods (XGBoost, RF, SVR, LR), algorithm that gives the best performance was determined.
- Scientific comparison was made by testing LR, SVR, RF and XGBoost models on the same data set.
- The effects of environmental variables including SD, TE, AP on solar radiation were analyzed in detail.

The present work is organized as in the following: Section 2 presents study area from which the dataset was obtained, as well as the processing and organization of the recorded dataset. This section gives the basis for methods utilized to predict total global GHI and details on the metrics used to evaluate their performance. Section 3 presents the general findings and discussions obtained from each prediction model. Section 4 summarizes the main insights for work, and offers suggestions of future studies in field of solar irradiance prediction.

2. MATERIALS AND METHODS (MATERYAL VE METOD)

2.1. Organizing the Workspace and Dataset
(Çalışma Alanı ve Veri Setinin Düzenlenmesi)

The meteorological data for this work was investigated through official correspondence from station affiliated with General Directorate of Meteorology (GDM) in Hakkari province, located in the southeast of Turkey (37°34'28.2"N and 43°44'19.8"E) and includes daily measurements between 2015 and 2022 [18]. The Hakkari region, with its high altitude, harsh continental climate, and variable atmospheric conditions, constitutes an important research area for solar irradiance prediction. In that manner, it attracts attention of many energy companies in terms of PV system investments, and many investments are being made for field of PV systems.

The dataset contains approximately 2,570 records, and meteorological variables including mean TE (°C), RH (%), WS (m/s), SD (hours), AP (hPa), and daily total global solar radiation (kWh/m²) were included in the study. These variables are fundamental meteorological parameters that have direct or indirect relationships with solar radiation and are widely used in solar radiation prediction models. They have sufficient density for data quality and continuity. The temporal changes of the parameters in the dataset across all years are shown in Figure 1.

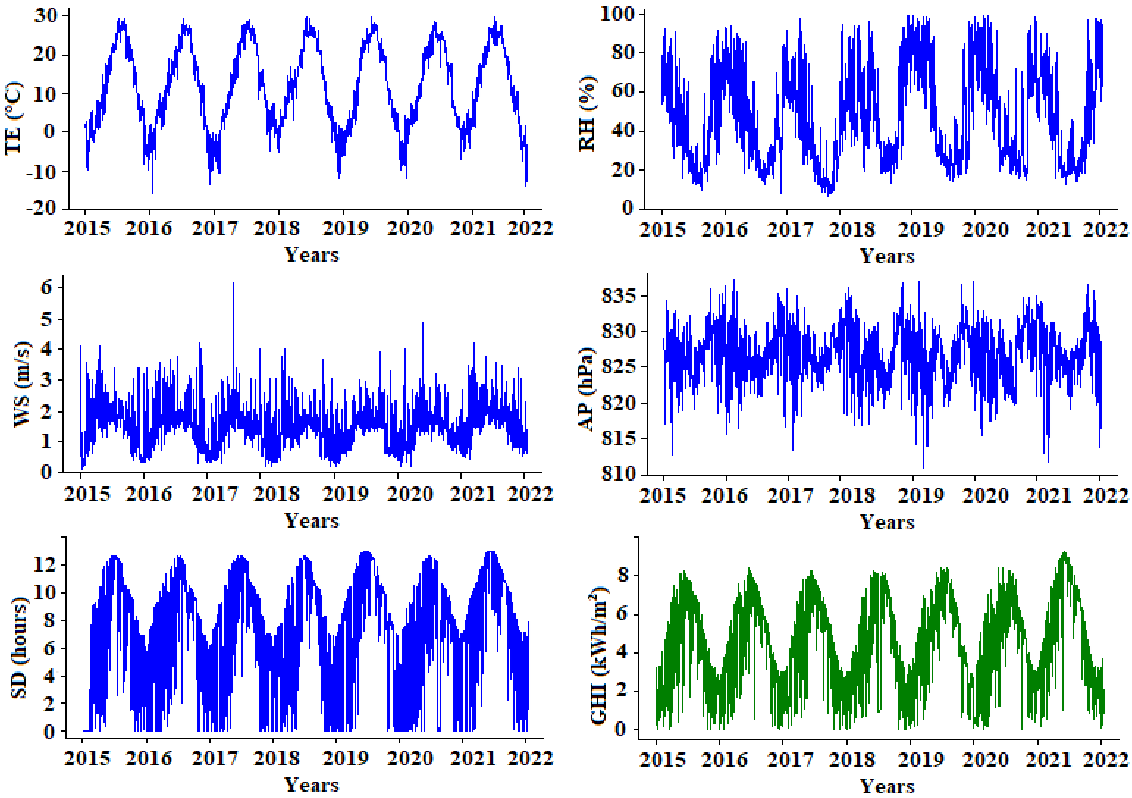

Figure 1. Time series change graphs for all parameters (Tüm parametrelere ait zaman serisi değişim grafikleri)

Examining the graphs in Figure 1 reveals that while the data set is generally consistent, the "0" value for sunshine duration or radiation measurements on some days is generally due to cloudy days. Furthermore, amount of missing days for data is quite limited, which will increase the accuracy of the model. Missing or outlier values were analyzed and removed before model training. However, days with no GHI values were not excluded from the data; they were corrected using appropriate padding methods after statistical control.

Since meteorological time series contain serial dependence, a time-based discrimination approach was preferred instead of the classical random splitting method. In this context, the periods covering 2015–2020 were used as training and validation data (approximately 2192 data points), and the periods covering 2021–2022 were used as testing data (approximately 378 data points). These values indicate that approximately 85% of the dataset was used for training and validation, while the remaining 15% was used for testing. This approach is compatible with the assumption that models should learn from past data and predict future values. Most of the machine learning methods are vulnerable to variable scaling. Therefore: Although models such as RF and XGBoost do not require scaling, in order for SVR and LR models to provide more stable results, the Min–Max Scaling method was applied and the data were normalized. In this context, the data were scaled within [0,1] with Min–Max normalization for the SVR and LR models; no scaling was applied for the RF and XGBoost models. The mathematical equation for Min–Max normalization is shown in Equation 1 [19–21].

$$P' = \frac{P - P_{min}}{P_{max} - P_{min}} \quad (1)$$

where P_{min} and P_{max} show lower and upper values in data set, respectively. P shows raw data to be processed, P' denotes normalized data.

2.2. Theoretical Foundations of the Model Used (Kullanılan Modelin Teorik Temelleri)

2.2.1. LR (Doğrusal Regresyon)

It was used as the most basic comparative model. It expresses the effect of independent variables on solar radiation using a linear combination. Its advantage is its simplicity and speed, but its capacity to capture nonlinear relationships is limited. The mathematical expression of this model is shown in Equation 2.

$$z = \alpha_0 + \sum_{k=1}^{m} \alpha_k u_k + \delta \quad (2)$$

where z is variable that is dependent, α shows parameter vector, u denotes independent variable, δ presents error of the model [22, 23].

2.2.2. SVR (Destek Vektör Regresyonu)

SVR has capacity to capture nonlinear relationships by transforming into high-dimensional spaces. SVR can provide strong predictive performance, especially when the relationships between input variables are nonlinear. The RBF kernel was chosen in this study. The model mathematical representation can be shown as follows:

$$K(u, u_j) = exp\left(-\frac{1}{\sigma^2} u - u_j^2\right) \quad (3)$$

where u, u_j show vectors in input space. In other words, they are feature vectors. σ is function parameter of the RBF kernel [24, 25].

2.2.3. RF (Rastgele Orman)

RF is the method including huge amount for decision trees. It is constructed utilizing samples of data with random selection. It performs particularly well on environmental and meteorological data [26–28].

2.2.4. XGBoost

XGBoost presents the widely used decision-based robust algorithm known for its efficiency and flexibility [29]. The basic formulation of the model is based on optimizing sequential trees to reduce errors [30, 31].

Using the four different machine learning models described above, comprehensive analyses were conducted to improve the accuracy of daily total GHI predictions. These analyses were implemented for training both machine learning methods utilized for evaluating results. Hyperparameter adjustments were made to maximize model performance, and the optimal hyperparameter set was determined for each method. The selected hyperparameters represent the values that provide the best prediction performance for the respective models. Training and testing were then conducted for verifying the effectiveness for domain-specific test dataset. Table 1 presents the baseline hyperparameters of the four different prediction models used in this study.

Table 1. Hyperparameters of the models (Modellerin hiperparametreleri)

Hyperparameter	LR	SVR	RF	XGBoost
Normalize	No	-	-	-
Fit_intercept	Yes	-	-	-
Solver	auto	-	-	-
Kernel	-	RBF	-	-
C	-	10	-	-
Gamma	-	scale	-	-
Epsilon	-	0.1	-	-
Shrinking	-	True	-	-

n_estimators	-	-	300	300
max_depth	-	-	6	6
learning_rate	-	-	0.05	0.05
colsample_bytree	-	-	-	0.8
booster	-	-	-	gbtree
objective	-	-	-	squared_error

2.3. Model Performance Measures (Model Performans Ölçütleri)

This section presents the metrics and their optimal value ranges for evaluating the prediction models. These metrics were used to determine the accuracy of the data obtained by analyzing four methods utilized for predicting daily total GHI. The equations and optimal value ranges for the performance metric criteria are shown in Table 2 [32–35].

Table 2. Details of performance evaluation criteria (Performans değerlendirme kriterlerinin ayrıntıları)

Model Name	Mathematical Equation	Ideal Value Range
RMSE	$RMSE = \frac{1}{M} \times \sqrt{\left(\sum_{k=1}^{M}\left(f_t - f_g\right)^2\right)}$	$0 \leq RMSE$
MAE	$MAE = \frac{1}{M} \times \left(\sum_{k=1}^{M}\left|f_t - f_g\right|\right)$	$0 \leq MAE$
R^2	$R^2 = 1 - \left(\frac{\sum_{k=1}^{M}\left(f_t - f_g\right)^2}{\sum_{k=1}^{M}\left(f_g - \bar{f}_g\right)^2}\right)$	$R^2 \leq \%100$

where M shows total data amount, f_t represents computed or predicted solar irradiance, f_g denotes measured irradiance.

3. RESULTS AND DISCUSSION (BULGULAR VE TARTIŞMA)

The purpose of this work is to obtain the effective approach by comparing the performance of algorithms predicting the daily total GHI using four different machine learning methods, based on daily meteorological measurements officially obtained from the Hakkari Meteorological Station between 2015 and 2022. The comparison is presented in Table 3.

Table 3. Comparison results of prediction models according to performance metric criteria (Tahmin modellerinin performans metriği kriterlerine göre karşılaştırma sonuçları)

Model	RMSE	MAE	R^2	R
LR	1.422	1.082	0.685	0.827
SVR	1.381	0.983	0.703	0.838
RF	1.491	1.061	0.654	0.808
XGBoost	1.509	1.094	0.646	0.803

When Table 3 is examined, the SVR model produced the lowest error values with RMSE=1.381 and MAE=0.983; it also explained the largest portion of the variance in solar radiation with R^2=0.703 and R=0.838 values. This result is an indication of SVR's ability to effectively model nonlinear relationships. Although the LR model showed high performance with R^2=0.685 and R=0.827, it fell behind SVR due to the limitations of linear assumptions. The RF and XGBoost models exhibited lower results with R^2=0.654 and R^2=0.646. The R values for these models were calculated as 0.808 and 0.803. This results from the fact that the GHI variable exhibits more continuous and smoother change patterns. The direction and magnitude of relations for meteorological parameters used in study and daily total GHI are shown in Figure 2.

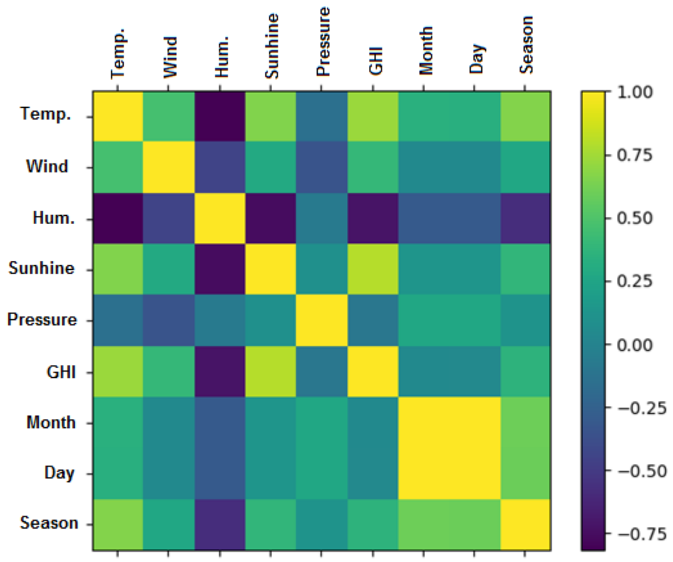

Figure 2. Correlation matrix for meteorological variables (Meteorolojik değişkenler için korelasyon matrisi)

The color intensities in the correlation matrix shown in Figure 2 represent correlation coefficients between -1 and +1, and the relationships between GHI and meteorological inputs are particularly critical for solar radiation prediction. The strongest positive correlation with GHI is observed for the sunshine duration variable. As sunshine duration increases, the amount of radiation reaching ground level naturally increases. The temperature variable was found to exhibit a moderate positive correlation with GHI. This is due to the fact that high temperatures are generally associated with clear, sunny weather conditions, and are associated with increased solar radiation.

The relative humidity variable is negatively correlated with the GHI. High humidity means more water vapor in the atmosphere, leading to scattered solar radiation and reduced radiation reaching the surface. This finding suggests that the humidity variable plays an indirect role in radiation prediction. Wind speed appears to have a weak correlation with the GHI. While wind does not have a strong direct effect on radiation, a limited correlation is observed due to its indirect effects on atmospheric mixing and cloud movement. There is a weak but positive correlation between atmospheric pressure and the GHI. Since high pressure generally indicates clear and stable weather conditions, it is expected to be associated with days with increased radiation values.

An overall evaluation of the correlation matrix reveals that sunshine duration, temperature, and time-based variables are the most effective inputs for the prediction models. This also supports the higher accuracy in study, as SVR can successfully model nonlinear relationships and seasonal patterns. Figure 3 depicts comparative distribution for actual daily total GHI values and predicted values for four different methods across time series.

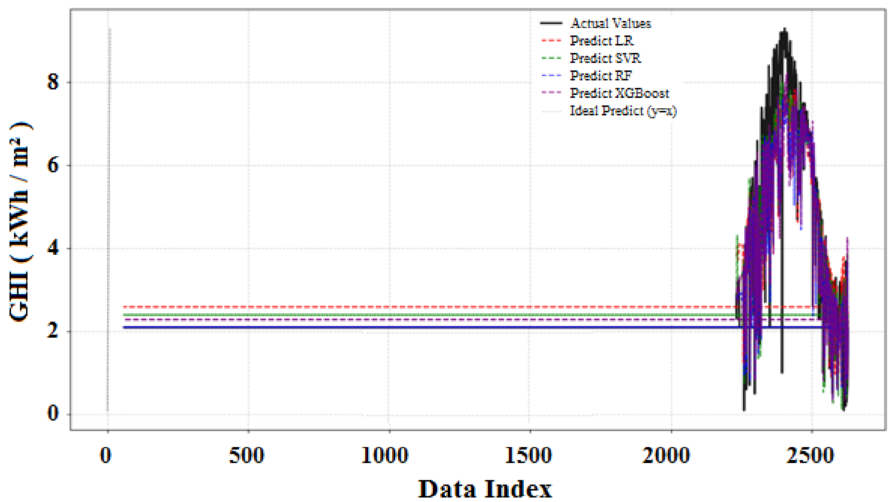

Figure 3. Comparison of actual and predicted daily total GHI for all models (Tüm modeller için gerçek ve tahmin edilen günlük toplam GHI'nın karşılaştırılması)

In the graph shown in Figure 3, the "ideal predict (y=x)" line represents the situation where model predictions are equal to the true values. The closeness of the model curves to this ideal line is an indicator of prediction accuracy, and SVR produced the curve closest to this line. While RF and XGBoost produced a prediction curve farther from the linear region, the LR model produced the prediction distribution farthest from the ideal line. This finding is consistent with the R, R^2, MAE, RMSE values presented in Table 3, supporting SVR model as the most suitable method for this study.

When the models' behavior over time series is evaluated, it is observed that the SVR model draws a prediction curve closer to the actual values than the other models. The SVR curve better captures the actual trend, particularly in peak regions where high radiation values are observed. This finding is consistent with SVR delivering the highest performance in terms of RMSE, R^2, and consequently, the R value. LR model's prediction curve is close to a horizontal line on the graph, failing to capture the complex structure of the actual values. This is an indication of the LR model's limited linear representation ability. Therefore, the LR model performed poorly for a parameter with high variability such as GHI. While the RF and XGBoost models followed the actual curve to a certain extent during periods of volatility, they did not respond as quickly to sudden changes as the SVR. The prediction curves of both models exhibit a flatter structure than the actual values, particularly during peak periods when radiation values increase. This can also be seen in the regression graphs of the actual and predicted GHI values demonstrated in Figure 4.

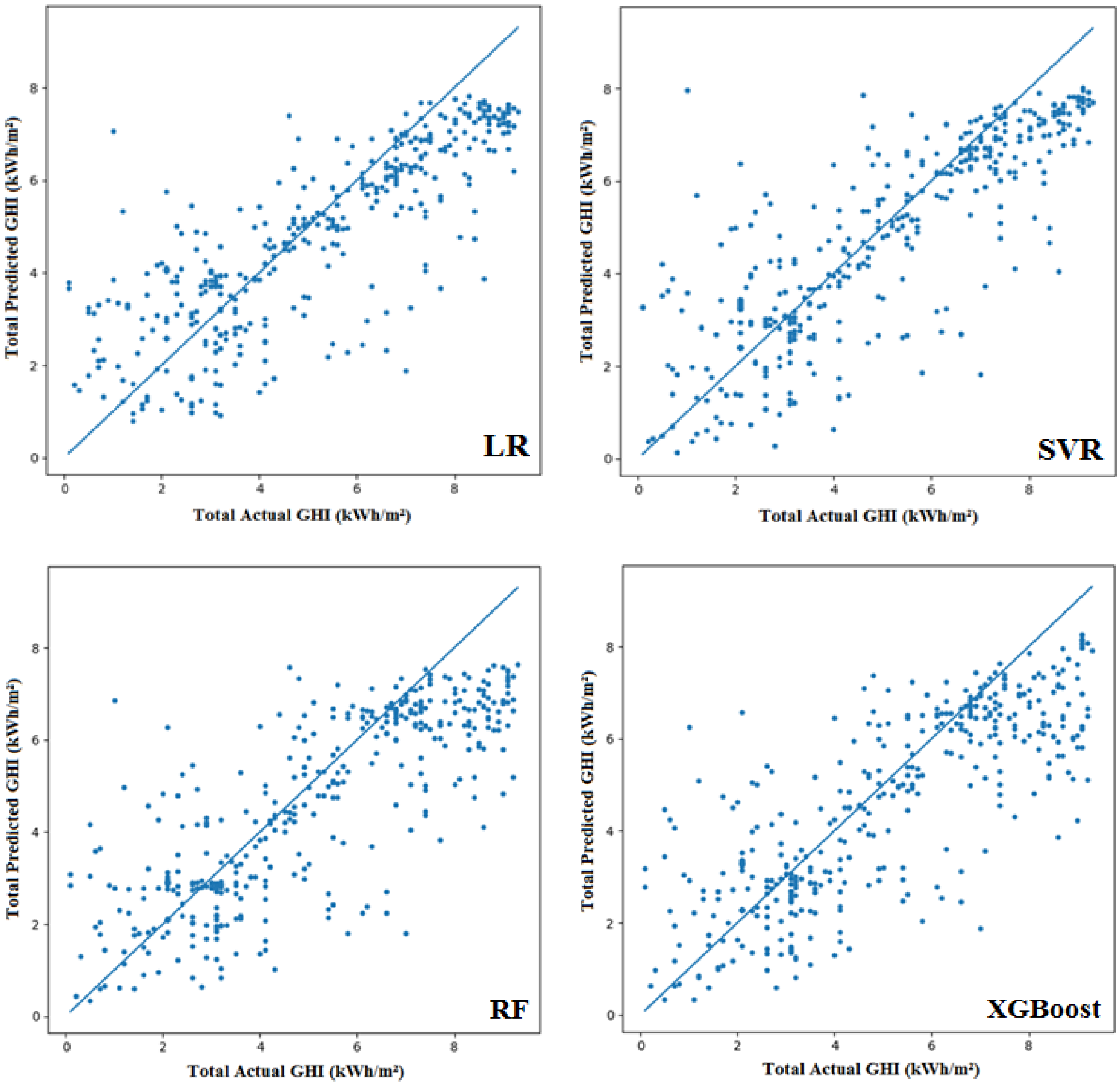

Figure 4. Regression graphs of actual and predicted GHI values (Gerçek ve tahmini GHI değerlerine ait regresyon grafikleri)

In the regression graphs for each model in Figure 4, the vertical axis shows the predicted daily total GHI values, and the horizontal axis shows the actual daily total GHI values measured by the meteorological station. These graphs are based on R values determined with the results of methods and visually compare daily GHI prediction capabilities for methods. An R value close to 1 is desirable, indicating a clear linear relationship. Model performance is directly proportional to density for data around the ideal line (y=x).

In Figure 4, when the regression graph of the LR model is examined, it is seen that this model exhibits a more pronounced distribution compared to the other models. Large deviations are observed, especially in the low and medium-level daily total GHI values. This indicates that the meteorological data used as input parameters for LR are insufficient to predict GHI. In the regression graph of the SVR model, the distribution of data points is more densely observed compared to LR. The SVR model exhibits a more pronounced distribution, especially at higher GHI values. SVR is sensitive to extreme values and exhibits better fit than other models in the medium GHI range.

The RF and XGBoost models present regression plots where data points are concentrated at the highest density along the line. However, the RF model still exhibits an irregular distribution of data at low GHI values, indicating that the model produces higher uncertainty at low radiation levels. In the XGBoost model, data points are observed to be more tightly clustered around the line across the entire GHI range. Due to its gradient boosting-based structure, XGBoost is understood to have a high capacity to capture complex patterns. Therefore, XGBoost exhibits the highest accuracy performance according to graphical evaluation. However, when evaluated together with the results obtained with performance metrics, it is observed that the RF and XGBoost models achieve lower GHI prediction success.

4. CONCLUSIONS (SONUÇLAR)

In this study, a daily total GHI prediction was performed using daily meteorological data from the Hakkari Meteorological Station between 2015 and 2022. For this purpose, four different machine learning models: LR, SVR, RF, and XGBoost were comparatively evaluated. The performance of the models was measured using RMSE, MAE, R^2, and R metrics, and the results revealed the main findings of the study. The findings show that the SVR model performed superiorly compared to the other models. SVR successfully modeled the nonlinear nature of solar radiation, particularly with low error rates (RMSE=1.381, MAE=0.983) and average coefficient of determination and correlation (R^2=0.703, R=0.838). Potential nonlinearity, outliers in the dataset, and the flexibility provided by SVR's kernel-based structure have contributed to its superior performance in solar radiation prediction compared to other models. Although the LR model performed better than expected (R^2=0.685, R=0.827), it was limited by the nonlinear nature of the radiation data. The RF and XGBoost models exhibited poor performance because of limited sample amount in dataset, which was measured in a region with high altitude and variable meteorological conditions. Therefore, the coefficient of determination and correlation coefficient values of these models were calculated as R^2=0.654, R=0.808 and R^2=0.646, R=0.803, respectively.

Another important result of the study is that sunshine duration, temperature, and atmospheric pressure are the most significant meteorological factors affecting the GHI. This finding is consistent with studies conducted in different climatic regions in the literature and further confirms the critical importance of these variables in solar radiation modeling. In this context, findings of methods are an effective tool for solar radiation prediction, even in topographically and climatically challenging regions like Hakkari. The present work presents novelties in terms of assessment for solar energy potential on Turkey's high-altitude regions and serves as an important reference for regional energy planning.

DECLARATION OF ETHICAL STANDARDS
(ETİK STANDARTLARIN BEYANI)

The author of this article declares that the materials and methods they use in their work do not require ethical committee approval and/or legal-specific permission.

Bu makalenin yazarı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

AUTHORS’ CONTRIBUTIONS (YAZARLARIN KATKILARI)

***İbrahim Çağrı BARUTÇU*:** He collected the data, conducted the simulations, analyzed the results and performed the writing process.

Verileri toplamış, simulasyonları yapmış, sonuçları analiz etmiş ve makalenin yazım işlemini gerçekleştirmiştir.

CONFLICT OF INTEREST (ÇIKAR ÇATIŞMASI)

There is no conflict of interest in this study.

Bu çalışmada herhangi bir çıkar çatışması yoktur.

REFERENCES (KAYNAKLAR)

[1] Sri Revathi B. A survey on advanced machine learning and deep learning techniques assisting in renewable energy generation. Environmental Science and Pollution Research. 2023; 30: 93407–93421.

[2] Gürel AE, Ağbulut Ü, Bakır H, Ergün A, Yıldız G. A state of art review on estimation of solar radiation with various models. Heliyon. 2023; 9(2): e13167.

[3] Dobruna L, Berisha A. A Data-Driven Approach for Predicting Solar Energy Production Using Machine Learning Techniques. In 2025 MIPRO 48th ICT and Electronics Convention. 2025; 180–185.

[4] Vanlalchhuanawmi C, Deb S, Islam MM, Ustun TS. Solar radiation prediction: A multi-model machine learning and deep learning approach. AIP Advances. 2025; 15(5): 055201.

[5] Sehrawat N, Vashisht S, Singh A. Solar irradiance forecasting models using machine learning techniques and digital twin: A case study with comparison. International Journal of Intelligent Networks. 2023; 4: 90–102.

[6] Obiora CN, Ali A, Hasan AN. Implementing extreme gradient boosting (XGBoost) algorithm in predicting solar irradiance. In 2021 IEEE PES/IAS PowerAfrica. 2021; 1–5.

[7] Qiu R, Liu C, Cui N, Gao Y, Li L, Wu Z, Jiang S, Hu M. Generalized Extreme Gradient Boosting model for predicting daily global solar radiation for locations without historical data. Energy Conversion and Management. 2022; 258: 115488.

[8] Kumar M, Namrata K, Kumar N. Data-driven hyperparameter optimized extreme gradient boosting machine learning model for solar radiation forecasting. Advances in Electrical & Electronic Engineering, 2022; 20(4): 549–559.

[9] Bouzgou H, Gueymard CA. Fast short-term global solar irradiance forecasting with wrapper mutual information. Renewable Energy. 2019; 133: 1055–1065.

[10] Soltani-Gerdefaramarzi S, Askarizadeh M. Application of Entropy Theory and Principal Component Analysis to Determine Input Variables for Estimating Solar Radiation using Machine Learning Algorithms. Physical Geography Research. 2024; 56(4): 73–87.

[11] Irfan M, Shaf A, Ali T, Zafar M, AlThobiani F, Almas MA, Almawgani AH. Global horizontal irradiance prediction for renewable energy system in Najran and Riyadh. AIP Advances. 2024; 14(3): 035137.

[12] Husain S, Khan UA. Development of machine learning models based on air temperature for estimation of global solar radiation in India. Environmental Progress & Sustainable Energy. 2022; 41(4): e13782.

[13] Güzel B, Sevli O, Okatan E. Forecasting solar radiation based on meteorological data using machine learning techniques: A case study of Isparta. International Journal of Engineering Research and Development. 2023; 15(2): 704–713.

[14] Gupta R, Yadav AK, Jha SK, Pathak PK. A robust regressor model for estimating solar radiation using an ensemble stacking approach based on machine learning. International Journal of Green Energy. 2024; 21(8): 1853–1873.

[15] Mugware FW, Ravele T, Sigauke C. Short-Term Predictions of Global Horizontal Irradiance Using Recurrent Neural Networks, Support Vector Regression, Gradient Boosting Random Forest and Advanced Stacking Ensemble Approaches. Computation. 2025; 13(3): 72.

[16] Hissou H, Benkirane S, Guezzaz A, Azrour M, Beni-Hssane A. A novel machine learning approach for solar radiation estimation. Sustainability. 2023; 15(13): 10609.

[17] Gaboitaolelwe J, Zungeru AM, Yahya A, Lebekwe CK, Vinod DN, Salau AO. Machine learning based solar photovoltaic power forecasting: A review and comparison. IEEE Access. 2023; 11: 40820–40845.

[18] TSMS. Turkish State Meteorological Service. https://www.mgm.gov.tr

[19] Song Z, Cao S, Yang H. An interpretable framework for modeling global solar radiation using tree-based ensemble machine learning and Shapley additive explanations methods. Applied Energy. 2024; 364: 123238.

[20] Wang H, Cai R, Zhou B, Aziz S, Qin B, Voropai N, Gan L, Barakhtenko E. Solar irradiance forecasting based on direct explainable neural network. Energy Conversion and Management. 2020; 226: 113487.

[21] Chiranjeevi M, Karlamangal S, Moger T, Jena D. Solar irradiation prediction hybrid framework using regularized convolutional BiLSTM-based autoencoder approach. IEEE Access. 2023; 11: 131362–131375.

[22] Li Q, Bessafi M, Li P. Mapping prediction of surface solar radiation with linear regression models: case study over reunion island. Atmosphere. 2023; 14(9): 1331.

[23] Rabehi A, Guermoui M, Lalmi D. Hybrid models for global solar radiation prediction: a case study. International Journal of Ambient Energy. 2020; 41(1): 31–40.

[24] Ramedani Z, Omid M, Keyhani A, Shamshirband S, Khoshnevisan B. Potential of radial basis function based support vector regression for global solar radiation prediction. Renewable and Sustainable Energy Reviews. 2014; 39: 1005–1011.

[25] Mohammadi K, Shamshirband S, Anisi MH, Alam KA, Petković D. Support vector regression based prediction of global solar radiation on a horizontal surface. Energy Conversion and Management. 2015; 91: 433–441.

[26] Karasu S, Altan A. Recognition model for solar radiation time series based on random forest with feature selection approach. In 2019 11th international conference on electrical and electronics engineering (ELECO). 2019; 8–11.

[27] Ibrahim IA, Khatib T. A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. Energy Conversion and Management. 2017; 138: 413–425.

[28] Benali L, Notton G, Fouilloy A, Voyant C, Dizene R. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. Renewable Energy. 2019; 132: 871–884.

[29] Yan H, Yan K, Ji G. Optimization and prediction in the early design stage of office buildings using genetic and XGBoost algorithms. Building and Environment. 2022; 218: 109081.

[30] Zhang C, Zhang Y, Pu J, Liu Z, Wang Z, Wang L. An hourly solar radiation prediction model using eXtreme gradient boosting algorithm with the effect of fog-haze. Energy and Built Environment. 2025; 6(1): 18–26.

[31] Fan J, Wang X, Wu L, Zhou H, Zhang F, Yu X, Lu X, Xiang Y. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. Energy Conversion and Management. 2018; 164: 102–111.

[32] Yuzer EO. Comparison of calculation methods and artificial neural network results in regional solar irradiation prediction. Electrical Engineering. 2025; 107(6): 7115–7135.

[33] Arslan G, Bayhan B, Yaman K. Mersin/Türkiye için Ölçülen Global Güneş Işınımının Yapay Sinir Ağları ile Tahmin Edilmesi ve Yaygın Işınım Modelleri ile Karşılaştırılması. Gazi University Journal of Science Part C: Design and Technology. 2019; 7(1): 80–96.

[34] Ertürk S, Kara H, Akkuş C, Genç G. Türkiye'de Farklı İklim Kuşakları İçin Yapay Sinir Ağları Kullanılarak Güneş Işınımının Tahmini. Gazi University Journal of Science Part C: Design and Technology. 2023; 11(4): 885–892.

[35] Uzun S, Arslantaş H. Determination of Radiation Value by Month Using Artificial Neural Network Model; Ankara, Sivas, Erzurum example. Gazi University Journal of Science Part C: Design and Technology. 2024; 12(1): 315–323.