Direct Solar Radiation Prediction using Multi-Model Evaluation with Trignometric Cyclic Process

Latif Bukari Rashid
Graduate Researcher
King Fahd University of Petroleum and Minerals (KFUPM)
Department of Mechanical Engineering
Saudi Arabia

Shahzada Zaman Shuja
Professor
King Fahd University of Petroleum & Minerals (KFUPM), Mechanical Engineering Department
Saudi Arabia

Shafiqur Rehman
Research Engineer
King Fahd University of Petroleum & Minerals (KFUPM), Interdisciplinary Research Center for Sustainable Energy Systems (IRC-SES) and Mechanical Engineering Department
Saudi Arabia

Accurate prediction of renewable sources in general and solar radiation is critical for optimal integration of solar energy systems. The studyexplores eight Machine Learning models namely Linear Regression Model (LRM),Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR),Gaussian Process Regression (GPR), , Artificial Neural Network (ANN), k-Nearest Neighbors (NN), Support Vector Regression (SVR), and Deep Learning (DL) for predicting thedirect solar radiation atclimatically distinctsix sites in Saudi Arabia. Models are evaluated using eight statistical metrics along with time series and absolute error analyses. The present work introduced the Trigonometric Cyclical Encoding (TCE), which significantly improved the temporal learning. Comparative SHAP-based analysis revealed that TCE enhanced the explanatory power of temporal features by 49.26% and 53.40% for monthly and daily cycles. Resultsshow that DL achieved the lowest Root Mean Square Error (RMSE) and highest coefficient of determination, while ANN consistently indicated high accuracy at all thesites. Error and time series analyses denoted stable predictions byANN and DL; whereas LR, RFR, and k-Nearest Neighbors (NN) showed largerfluctuations. The proposed TCE technique additionally improved the model outputby maintaining the overall fitness of the models between 81.79% and 94.36% in all scenarios. This studyreinforcethe effec–tive planning of solar energy integration in differentclimatic conditions.

Keywords: *Forecasting;Renewable Energy; Solar Energy;; Machine Learning; Deep Learning; Saudi Arabia.*

1. INTRODUCTION

The climate change and fast depletion conventional fuels have driventhe world to move at a faster pace to–wards global transition of energy production from clean and renewable sources instead of traditional fossil fuels [1–3]. Among technologically mature and commercially acceptable sources, solar energy has seen prominence as a clean source of energy asit places less stress on the environment and can be integrated into urban areas [4,5]. In the recent times, renewable energy has ac–counted for 86% of global power additions in 2023, lar–gely as a result of enormous growth in solar and wind power [6,7]. This growth is mostly associated with the technological development, decrease in cost, and proper government and institutional policies and frameworks have made the utilization of solar energy feasible and affordable [8,9].

However, solar power generation is variable/ inter–mittent and depends on weather, climate, and temporal patterns. The unpredictability of solar electricity gener–ation is a significant barrier to blending it into existing power infrastructure[10]. The fluctuating behaviour of solar resources is influenced by meteorological condi–tions, thereby creating uncertainty in estimating the power generation [10]. A study by [11] has shown that enhanced forecasting can improve the efficiency of the plant operations by up to 15% and the integration costs can also be cut down to a large extent. Solar forecasting could be applied to scheduling the use and storage of solar energy, evaluating and forecasting the perfor–mance of existing solar installations, sizing solar plants, and assessing the capacity and demands of electricity networks [12].

Two main techniques; the empirical and ML models, have received prominence in the literature, as efficient ways of estimating and forecasting the incoming solar energy by learning from historical time series of a parti–cular geographical setting [13]. Hissou et al. [14] reported a new technique by integrating several ML models with RFE and obtained a great success com–pared with the performance of LR model (RMSE= 0.003, R^2=0.999). Solano and Affonso [15]showed that combining multiple ML algorithms, through ensemble with weighted averaging gave commendable results: 6% decrease in Mean Absolute Error (MAE), 3% in Root Mean Square Error (RMSE) and 16% in Mean Absolute Percent Error (MAPE). Rehman et el.[16]and [17]used these machine learning techniques for wind speed pre–diction and wind farm layout optimization and reported good performance which further strengthen the role of these methodologies for forecasting the meteorological parameters for renewable energy applications. Mirkov et al. [18] proposed a preconditioner for Krylov method solution of linear systems for wind energy applications. When building ML model for a particular geographic

Received: January 2008, Accepted: March 2008
Correspondence to: Dr Shafiqur Rehman,
King Fahd University of Petroleum & Minerals
Dhahran, Saudi Arabia.
E-mail: srehman@kfupm.edu.sa
doi: 10.5937/fme2601128B

location, certain features contribute to its architecture [19]. Chaibiet al. [20] reported SVR-BO as the optimal performer for Moroccan climatic conditions (RMSE =0.4473). On the other hand, Bakir [21] found that best performing algorithms differ from site to site. There are also more advanced strategies that combine multiple methods, which are starting to yield better progress - for instance, Lu et al. [22] combined LibRadtran RTM and ML models and achieved an R^2 value of 0.98 but their work was limited to clear sky conditions. Likewise, Khafaga et al. [23] proposed an ADSSOA-LSTM which led to a low RMSE of 0.000388 however adding up computational burden.

The Kingdom of Saudi Arabia (KSA) stands out as one of the world's top locations for capturing solar energy. Forecasting solar radiation accurately is an im–portant step in the planning, design, and utilization of solar energy systems [24]. This study focuses on the performance evaluation of eight ML algorithms for six climatologically distinct sites in KSA, representing a geographical distribution within the coastal, inland, northern and western flanks of the kingdom with dif–ferent altitudes and topography. Meteorological data was sourced from NASA POWER on daily basis span–ning January 1, 2022to December 31, 2024for all sites. ML algorithms were evaluated using a wide range of metrics (MBE, RMSE, rRMSE, MABE, MAPE, R^2, t-stat, MSE, and MAE) to offer a reliable assessment of the prediction accuracy [25].

The present study seeks to address the gap in location-specific insights into ML model efficacy, cont–ributing to the identification and selection of optimal algorithms for accurate solar energy forecasting within varied geographical context across the middle eastern region particularly for the case of KSA [26]. Prior research has excessively used machine learning (ML) and deep learning (DL) models, but the time (months, days, and hours) are commonly treated as linear inputs, ignoring their inherent cyclic nature. The present study introduces TCE strategy to preserve seasonal periodicity and improve temporal learning in consistent manner. Unlike previous studies, the effectiveness of TCE app–roach is quantitatively validated using SHAP-based analysis. Furthermore, different ML/DL algorithms of linear, kernel-based, ensemble, and neural networks are evaluated across climatically distinct regions and offer a rare, large-scale, and geographically diversified assess–ment.Table 1 summarises research findings, most recent literature, methodologies, and limitations relevant to this study.

Table 1. Summary of Literature Review

Ref.	Methodology	Key Findings	Limitations
[15]	Multiple ML algorithms (RF, XGBoost, CatBoost)	Best performance with RF and CatBoost combination	Limited to Brazilian region
[20]	Comparison of 5 ML models with/without BO	SVR-BO performed best (RMSE=0.4473kWh/m²/day)	Single location study (Fez, Morocco) and Limited feature set
[21]	Multiple metaheuristic algorithms (GBO, HHO, BMO, SCA, HGSO) for distinct locations in Turkey	SCA best for Afyonkarahisar GBO best for Ağrı	Limited input variables
[22]	Comparison of six ML approaches	RTM-RF showed best performance (MAE 15.57 W/m²)	Limited to clear sky conditions
[23]	ADSSOA-LSTM hybrid comparison with GA, PSO, GWO	ADSSOA-LSTM achieved lowest RMSE (0.000388)	Limited feature exploration
[27]	Predicts daily global solar radiation data for 6 Pakistani cities	SVR achieves the best performance with R² values up to 0.99	No FE, No feature selection reported
[28]	Ensemble ML algorithms for solar power prediction in Saudi Arabia	RF outperformed other models (MAE=0.0141), (RMSE=0.0211)	Limited to Dhahran, Limited evaluation metrics
[29]	Multiple ML models (RF, GBM, LR, CART, DT)	LR and RF achieved lowest nMAE (-0.144, -0.151)	Limited feature selection methods
[30]	Compares RF with hyperparameter optimization with other ML models	95.98% accuracy with optimized RF	Limited to Queretaro, Mexico Focused on short-term predictions
[31]	Comparative analysis of BiLSTM-based LSTNet	RF-LSTNet performed best	Limited explanation of feature selection process
[32]	WRF Solar model	Superior performance compared to baseline models	Region-specific (Northwest China)
[33]	Radial Basis Function Neural Network (RBF-NN) for DSR and DNR	DSR; MAPE = 1.6%-9.3% DNR; MAPE= 0.49%-41%	Relatively old dataset (1998-2002)
[34]	Review of ML techniques	Decision trees, RF, XGBoost, and SVM are effective ML models	Inadequate use of FE, Limited context for KSA
[35]	Multiple ML algorithms comparison	XGBoost showed highest performance	Single location study
[36]	Comparison of next-gen ML algorithms	Random Forest outperformed other algorithms; MLP-ANN improved with feature selection	Limited to single application

2. METHODOLOGY

The study usesa structured ML pipeline for performance evaluation of eight algorithms for the prediction of DNI at six climatically distinctlocationsin KSA. The metho–dology includes site identification, data procurement, pre-processing, exploratory data analysis, and temporal feature engineering using Trigonometric Cyclic Enco–ding (TCE). Next, the model is trained and then exha–ustive hyperparameter optimization is performed tailored toeach site and ML algorithm. Finally, as et of

statistical metrics is used to assess the model perfor–mance. Figure 1 presents a step-by-step flow of the methodological approach use din the study.

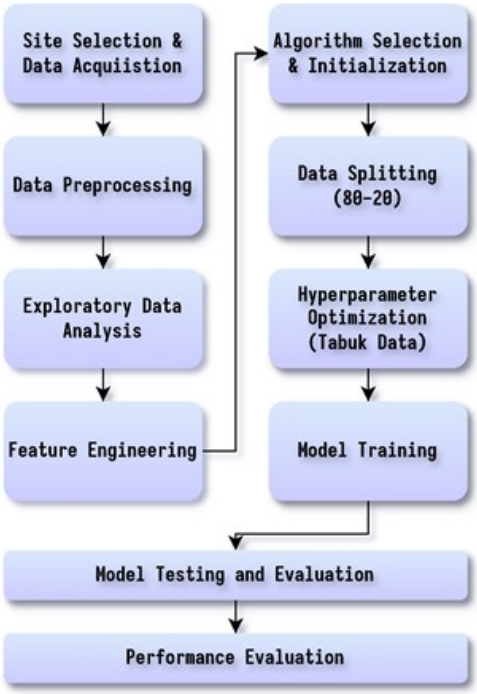


Figure 1. Methodological Flowchart

2.1 Case Study Area Description

The Kingdom of Saudi Arabia has become the hotspot for harnessing solar energy due to its unique geographic position, which allows it to enjoy high solar irradiance [28,37]. This study focuses on six key strategic sites representing diverse climatic and geographical regions in KSA. These sites arecarefully selected to ensure a wide coverage of different altitudes, terrains, and solar radiation profiles, to provide a robust foundation for evaluating and assessing the performance of the ML algorithms. Table 2 presents key geographic informa–tion about the sites considered in this study.Figure 2 presents the geographical locations of the six selected sites - Hagl, Tabuk, Timaa, Duba, AlWajh, and Umluj - spread across diverse climatic zones in KSA.Figure 3 provides an overview of DNI across KSA and showcases the yearly and daily averages of DNI.

2.2 Meteorological Data Overview

The data used was obtained from NASA POWER data–base. The present study included the essential input variables date, temperature, relative humidity, all sky clearness index, and wind speed on daily basis from January 1, 2022, to December 31, 2024 [38,39]. The dataset consists of 365X3 rows of daily values with 8 columns representing 7 input features and 1 target variable for all the sites. The target variable, labelled as All Sky Surface Shortwave Downward Direct Normal Irradiance, explicitly represents the direct component of solar radiation incident on a surface normal to the sun's rays under all-sky conditions (i.e., including the effects of clouds). Metadata and description of the meteoro–logical data is summarized in Table 3.

Table 2. Geographical Overview of Case Study Areas

Location	Region	Latitude (^{o}N)	Longitude (^{o}E)	Altitude (m)
Hagl	Northern	29.2899	34.9300	36
Tabuk	Northern	28.3829	36.4839	781
Timaa	Eastern	27.6173	38.5252	844
Duba	Western	27.3410	35.7229	45
AlWajh	Western	26.2561	36.4430	21
Umluj	Northern	25.0041	37.2738	10

Figure 2. Geographical Distribution of Case Study Sites Across Saudi Arabia (Source: Authors)

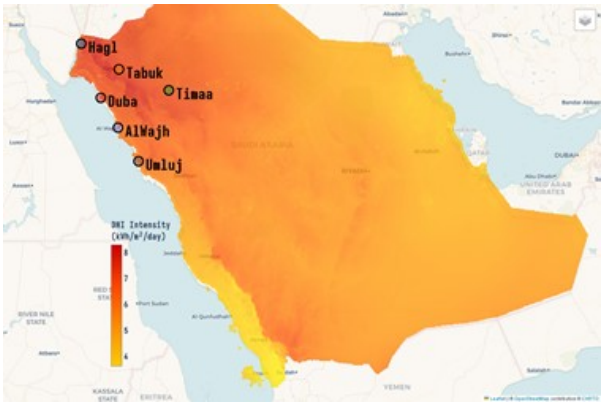


Figure 3. DNI Resource Map of KSA (Source: Solargis)

2.3 Machine Learning (ML) Algorithms

Eight ML algorithms are selected for the evaluation due to their known performance in non-linear regression and solar radiation forecasting tasks [41,42]. Table 4 sum–marizes the underlying mechanisms, strengths, limita–tions, and best use cases. Selected algorithms span from simple linear regressors (LRM) to ensemble-based learners (RFR, GBR) and kernel methods (SVR, GPR), as well as neural network models (ANN, DLM).

Table 3. Description of Meteorological Data [40]

Feature	Description	Unit
DT	Date	-
MO	Month	-
DY	Day	-
HR	Hour	hr
TMP	Temperature at 2 Meters	^{0}C
RH	Relative Humidity at 2 Meters	%
CI	All Sky Insolation Clearness Index	dimensionless
WS	Wind Speed at 10 Meters	m/s
DNI	All Sky Surface Shortwave Downward Irradiance	kWh/m^{2}/day

Table 4. Summary of ML Algorithms [43]

Algorithm	Strengths	Limitations	Use Case Fit	Ref
ANN	Captures complex non-linear patterns	Needs tuning, prone to overfitting	Good for moderately complex patterns and flexible modelling	[44]
DLM	Learns hierarchical features, handles time patterns	Requires large data, slow to train	Best for large datasets and capturing complex temporal/spatial patterns	[41, 43, 45]
GBR	High accuracy, customizable	Slow training, risk of overfitting	Ideal for maximizing accuracy on structured data	-
GPR	Probabilistic predictions, flexible	Computationally intensive	Useful when uncertainty estimates are important	-
KNN	Simple, no training phase	Sensitive to 'k' and scale of data	Useful for small datasets where local similarity matters	-
LRM	Simple, fast, interpretable	Fails to capture non-linear patterns	Best for simple, linear relationships	-
RFR	Robust to overfitting, handles non-linearity well	Slow for large forests, less interpretable	Great for noisy or non-linear tabular data	-
SVR	Strong performance on smaller datasets	Poor scalability to large datasets	Works well for small to medium datasets with clear margins	[46]

2.4 Hyperparameter Optimization

Table 5 summarizes the hyperparameter search space and the corresponding optimized values used for each ML algorithm in this study. The chosen ranges are designed to balance model flexibility and computational efficiency, drawn on established values from literature and prior experiences in regression tasks involving solar radiation forecasting. For instance, the range of parameters for RFR (estimators: 800 to 1200, depth: None to 20) are commonly used for high-dimensional, non-linear problems with similar data sizes as used in this study. The SVR was tuned using variations of C, epsilon, and kernel functions that are known to influence margin-based learning in noisy/non-linear data. The hyperparameter space for Gradient Boosting was selected to explore the trade-off between learning rate, tree depth, and model complexity. For ANN, the architecture and learning strategy waspre-defined rather than tuned via exhaustive search, following common literature practices in deep learning model building for tabular data [47].

Table 5. Hyperparameter Search Space and Selected Optimized Values for Classical ML Algorithms [33,34]

Model	Hyperparameter	Optimization Range	Optimized Hyperparameters
ANN	Hidden layer sizes	-	(128, 64, 32, 16)
	activation	-	relu
	solver	-	adam
	alpha	-	0.0001
	Learning rate	-	adaptive
GBR	estimators	100, 200, 300	1000
	Learning rate	0.01, 0.1, 0.2	0.03
	Max depth	3, 5, 7	6
	Sub sample	0.8, 1.0	0.9
	Min samples split	2, 5, 10	5
GPR	kernel	1.0 * RBF (length scale=1.0), 1.0 * Matern (length scale=1.0, nu=1.5)	1**2 * Matern (length scale=1, nu=1.5)
	alpha	1e-5, 1e-3, 1e-1	1e-1
	optimizer	fmin_l_bfgs_b	fmin_l_bfgs_b
	restarts	3, 5	5
KNN	neighbors	3, 5, 7, 10	10
	weights	uniform, distance	distance
	metric	euclidean, manhattan	manhattan
LRM	-	-	default
RFR	estimators	800, 1000, 1200,1800	1800
	Max depth	None, 10, 20	None
	Min samples split	2, 4, 6	5
	Min samples leaf	1, 2, 3	2
	Max features	0.3, 0.5, sqrt, log2	log2
SVR	C	1, 10, 50, 100	50
	epsilon	0.01, 0.1, 0.2,0.5	0.2
	kernel	linear, rbf	rbf
	gamma	scale, auto	scale

Deep Learning Model (DLM)

Given the practical constraints associated with hyper–parameter optimization for deep learning across mul–tiple geographical datasets, a manually configured arc–hitecture was implemented in this study. The design choices were guided by prior studies on similar fore–casting tasks [48]. Performance was monitored using MAE, and early termination was applied to minimize overfitting. Table 6 summarizes the configuration para–meters for DLM.

Table 6. Training Parameters for DLM [48]

Parameter	Value
Feature Selection	Top 3 features
Input Dimension	3 (based on FS output)
Hidden Layers	128, 64, 32, 16
Activation Function	relu

Dropout Rate	0.1
Optimizer	adam
Loss Function	MSE
Evaluation Metric	MAE
Learning Rate Strategy	adaptive
Max Iterations (Epochs)	1000
Batch Size	128, 64, 32, 16
Early Stopping	Yes

2.5 Trigonometric Cyclic Encoding (TCE)

Feature Engineering (FE) is an important method for transforming time dependent data into a more infor–mative and model-friendly format. This study focused on strategically creating and transforming time features to maximize the predictive power of the available information [14,30]. A key aspect of FE is to effectively handle temporal variables, acknowledge the cyclic nature of Month, and Day by using the TCE technique. Traditional time series numerical representations of days, and months often missed the periodicity of these measurements. For instance, the shift from 23:00 to 00:00 indicates closeness rather than a significant linear change, which is frequently overlooked by standard numerical encoding methods[30].

The present study applied TCE technique for cyc–lical features of time (days and months). Each temporal variable was broken into sine and cosine components, producing paired features that reflect the circular nature of time. This transformation helps machine learning models grasp the periodic relationships between time-based data [49]. This method remains inadequately un–explored [14]. The present study considers the cyclical encoding method to convert cyclic data into a format that is suitable to ML algorithms [14].

As shown in Figure 4, this study turns each time-related value (day or month) into a circular format, so that the smallest and the largest values sit next to each other. This is achieved byusing sine and cosine func–tions, which allow to represent time in a smooth and continuous way. In case of hours, the circle starts at midnight on the left and moves counterclockwise. This means that 11:59 PM is placed right next to 12:00 AM - just like it is in real time. The same kind of transfor–mation is applied to both, the month and the day values.

For days and months, the trigonometric cyclical transformation into sin and cosine components is exp–ressed mathematically as follows:

$$x_{\sin} = \sin\left(2\pi \frac{x}{T}\right) \tag{1}$$

$$x_{\cos} = \cos\left(2\pi \frac{x}{T}\right) \tag{2}$$

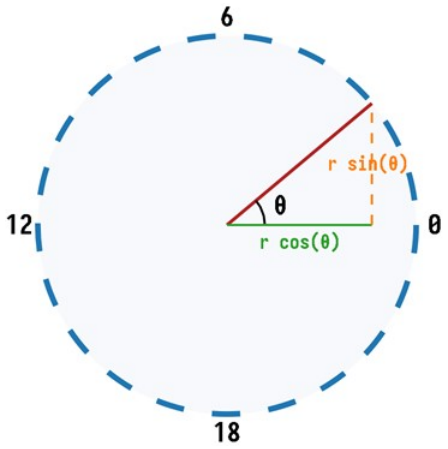


Figure 4. Cyclical Features (Source: Authors)

where x is the cyclical feature value (month or day), T is the period of the cycle (12 for months, 365 for days), x_{sin} and x_{cos} are the sin and cos transformed values.

To better capture the seasonal patterns of solar ra–diation, TCE was applied to convert the temporal vari–ables month and day into cos and sin features to pre–serve the periodic nature of time variables without int–roducing artificial breaks (for instance, between De–cember and January). Figure 5 presents the correlation between the TCE features and the core climatic vari–ables for each site. For instance, in Haql, cos_month has a strong negative correlation with TMP (-0.72) and a positive correlation with RH (0.63). Similar relati–onships are observed forTabuk, where cos_month cor–relates at --0.75 with TMP and at 0.76 with RH. In Timaa and Duba, which exhibit more extreme seasonal variability, sin-month and cos-month both maintain mo–derate correlation with DNI of up to -0.39. Additionally, the encoded features (sin_MO, cos_MO, sin_DY, cos_DY) remain nearly uncorrelated with each other, demonstrating no clear redundancy and multicollinearity.

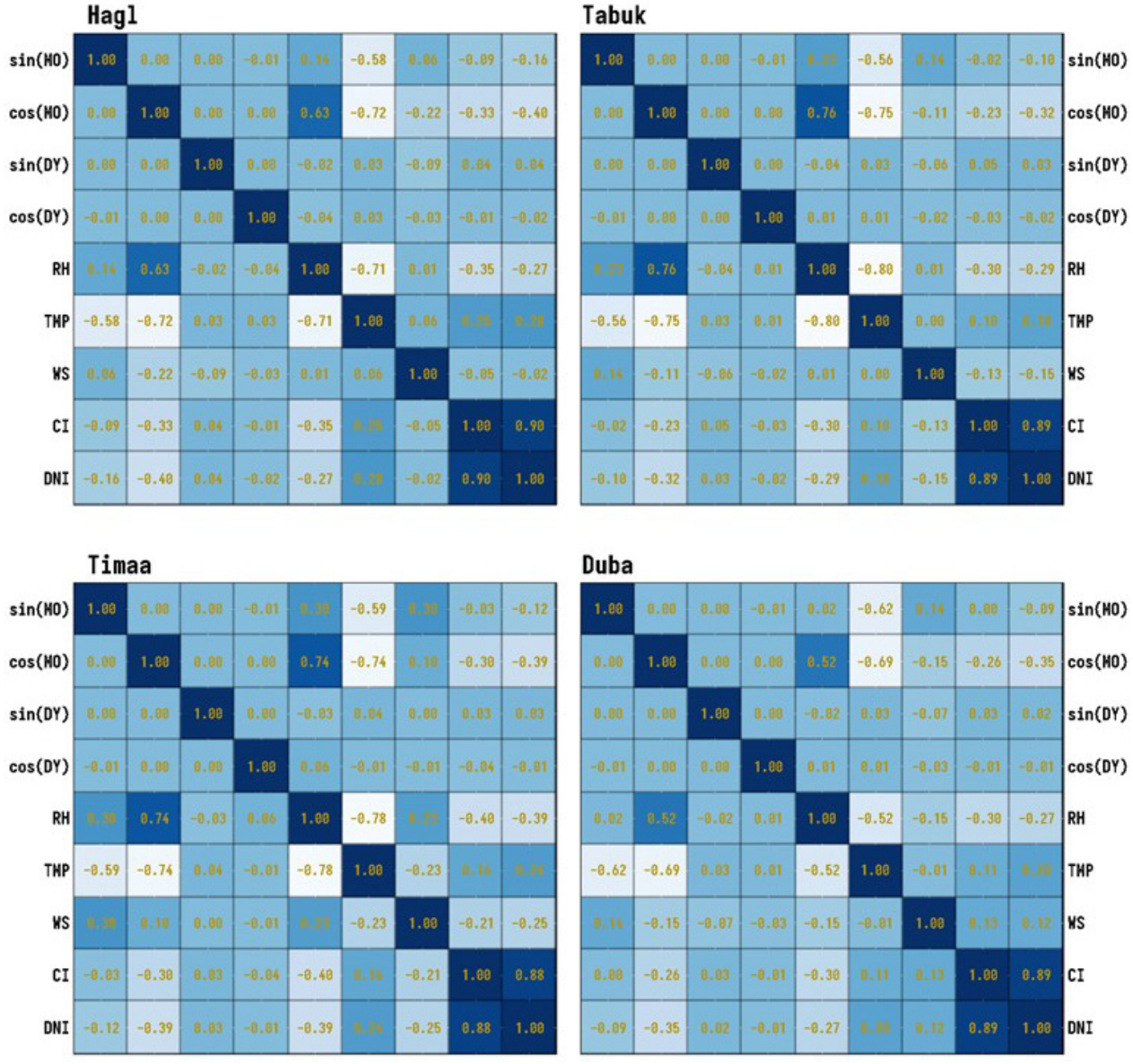
2.6 Performance Metrics

The performance metrics - MAE, MSE, RMSE, R^2, rRMSE, t-stat, MAPE, and MBE are used to assess the performance of the ML models. This metrics is care–fully chosen to conduct a thorough evaluation of the overall performance of estimating and predicting solar radiation in-terms of error magnitude. The mathematical expressions of these metrics is presented in Table 7. In Table 7, the term $\boldsymbol{y_i}$ is the observed value of the depen–dent variable, $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ is the predicted value from the ML models, $\boldsymbol{n}$ is the number of observations, and $\bar{\boldsymbol{y}}_{\boldsymbol{i}}$ is the mean of the observed values.

Table 7. Mathematical Model of Performance Metrics

Metrics	Mathematical Model	Description	Desired Output
MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	Measures the mean magnitude of errors between predicted and actual values without considering their direction[50,51]	closer to 0 is better
MSE	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	Measures the mean squared differences between predicted and actual values, and penalises larger errors more heavily[50]	closer to 0 is better

RMSE	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	Square root of MSE, providing error measure in the same units as the target variable[52]	closer to 0 is better
R^2	$R^2 = 1 - \left[\frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \overline{y}_i)^2}\right]$	Explains the variation in the target variable that's predictable from the input variable(s)[53]	closer to 1 is better
MAPE	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{y_i - \hat{y}_i}{y_i}\right\| \times 100$	Expresses accuracy as a percentage, showing the mean absolute percent difference between predicted and actual values[52,54]	closer to 0% is better
MBE	$MBE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)$	Used to evaluate bias of forecasting models[55]	closer to 0 is better
rRMSE	$rRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\overline{y}_i} \times 100$	Derived from RMSE[52]	closer to 0% is better



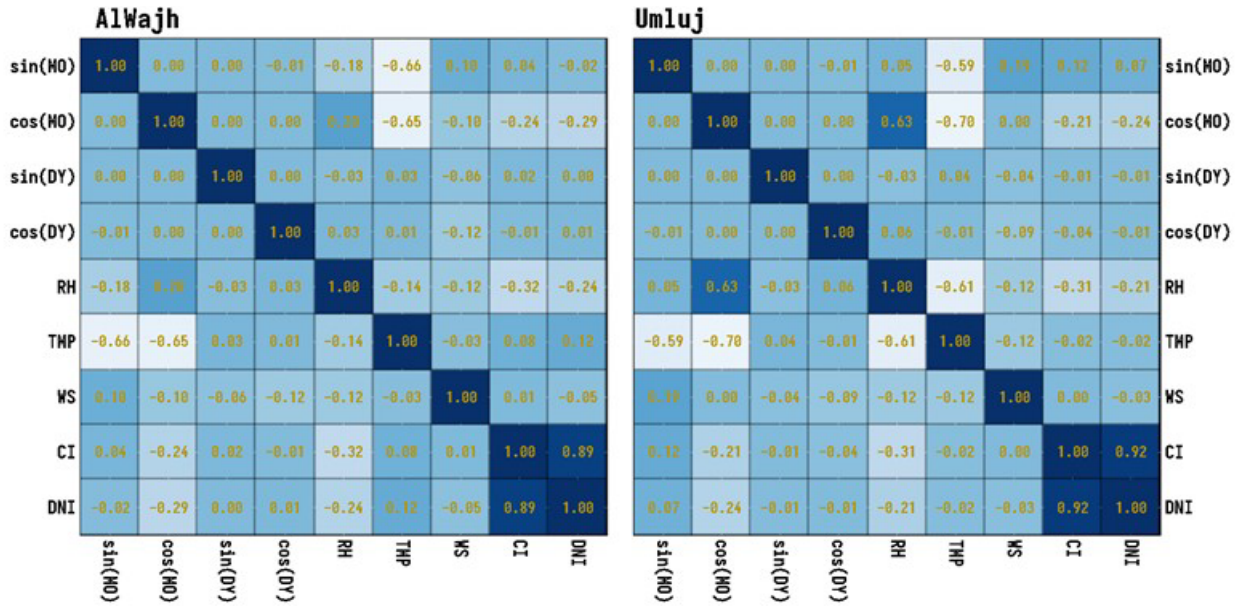


Figure 5. Correlation of TCE Features and Climatic Variables Across the Six Case Study Sites

3. RESULTS AND DISCUSSION

3.1 Temporal Patterns of DNI and Meteorological Parameters

The data used in this study covers a period of three years from January 1 2022 to December 31 2024. The daily mean values of meteorological data are used in this study, which are obtained from hourly mean values. Accurate forecasting of solar radiation relies on a clear under–standing of the local climatic dynamics influencing irradiance levels. This study identifies seasonal patterns and site-specific environmental behaviours of irradiance that affect the accuracy ofML model. Figures 6 to 10 present the trends of DNI, WS, CI, RH, and TMP, respectively;providing the forecasting challenges and regional variations present across the study locations.

DNI values in Tabuk are consistently high, during the summer months (June-September), where it frequently reaches around 8 to 10 kWh/m²/day, with noticeable lows during winter, Figure 6. Riyadh exhibits moderate variability, with values typically ranging from 3-7 kWh/m²/day and occasionally peaking above 9 kWh/m²/day during summer months. Haql shows the most moderate profile among all sites, generally maintaining values between 2.5 and 6.0 kWh/m²/day, with some peaks of7 kWh/m²/day during summer. Timaa demonstrates consi–derable fluctuation throughout the year, with values ranging from 2.5 to 8 kWh/m²/day. Duba presents a relatively stable profile with most values lying between 3 and 7 kWh/m²/day, though it experiences some spikes of up to 9.0 kWh/m²/day during certain periods. AlWajh shows a pattern somewhat similar to Tabuk, with high summer values often exceeding 8 kWh/m²/day, but exhibits more pronounced fluctuations during winter months, where values drop even below 3.0 kWh/m²/day.

AtHaql, the wind speed exhibits moderate patterns at 10 meters above ground level (AGL), with peaks reaching around 6.0 m/s showing stronger winds during mid-year, Figure 7. Tabuk's wind profile also exhibits moderate chnage, with speeds typically ranging between 3.0 and 7.0 m/s. Timaa demonstrates relatively larger fluctuations winds throughout the years, with occasional highs above 9.0 m/s. Duba, a coastal site, experiences steady winds with rarely surges above 6.0 m/s, but generally maintains moderate speeds around 3.0-5.5 m/s. AlWajh presents a somewhat erratic pattern with notable fluctuations between 3.0 and 7.0 m/s during three years. Umluj, a coastal site, experiences good winds with variations between 3.0 and 7.0 m/s but often reach 8.0 m/s and more.

Tabuk generally maintains higher and relatively stable CI values, often above 0.70, particularly during mid-year, while experiencing some dips during winter months, see Figure 8. Timaa shows moderate variability with CI values typically ranging between 0.45 and 0.75, with occasionally reaching below 0.50. Duba exhibits the same pattern as above with chnages between 0.5 and 0.7. At AlWajh, relatively consistent high values of CI , often above 0.65, with some lows near the end of the year are noticed.

Umlujshows moderately stable values varying bet–ween 0.60 and 0.70. Haql presents an interesting pattern with high values but more pronounced fluctuations du–ring the winter months.Haql, Duba, AlWajh, and Umluj maintain relatively stable and cyclic RH levels, typically ranging between 40-70%, with AlWajh showing slightly higher summer time values more consistent levels, as observed from Figure 9. Tabuk and Timaa demonstrate more pronounced variations, with higher RH in winter months (reaching 80%) and lower during summer (dropping to 15-20%).

With respect to ambient temperature variation over three years of data, all the sites, under consideration, show a cyclic variation with lows and highs during winter and summer times, as shown in Fig. 10. At Haql, Tabuk, and Timaa;overall lower values of temperature are observed during entire period compared to those at Duba, AlWajh, and Umluj. At first three sites, the temperature values reach as low as of 5°C in January and February and as high as 30°C during summer time (August and September). At Duba, Alwajh, and Umluj; all three coastal sites; the daily mean temperature values vary from 15°C to 37°C and has relatively lower variability in terms of annual magnitudes, Figure 10.

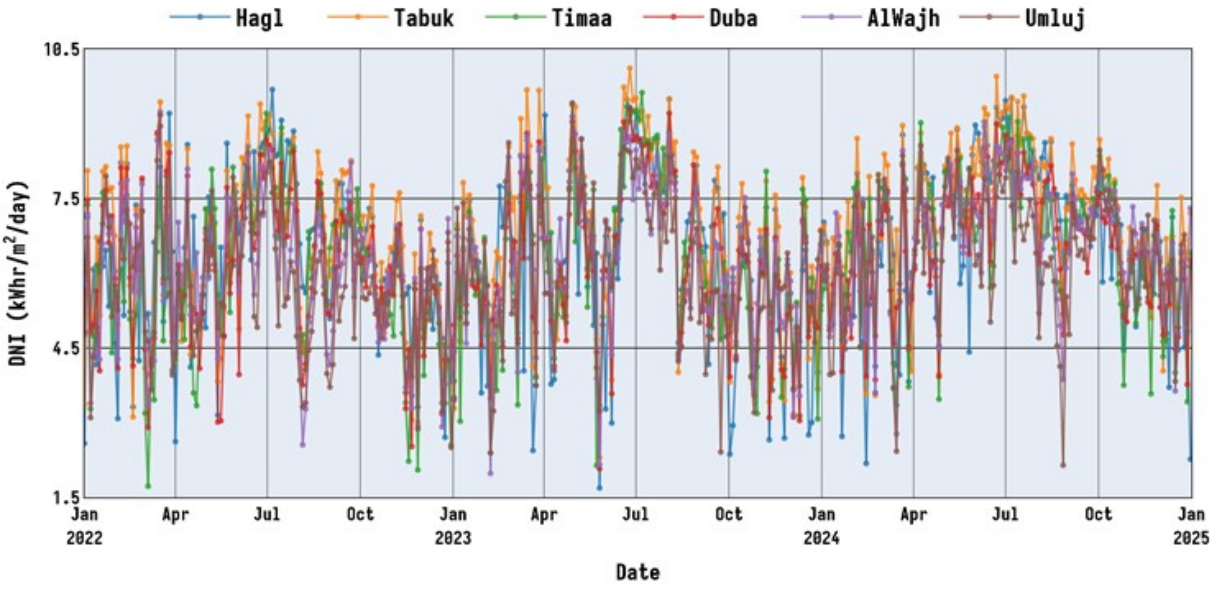


Figure 6: Direct Normal Irradiance trends across chosen sites

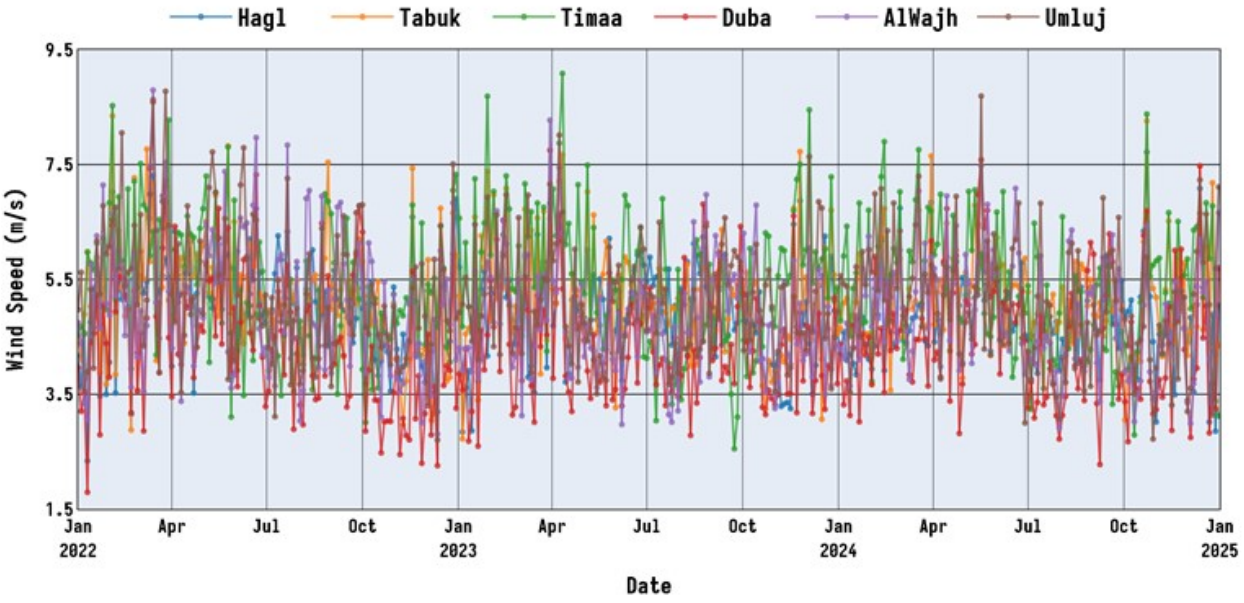


Figure 7: Wind Speed trends across chosen study sites

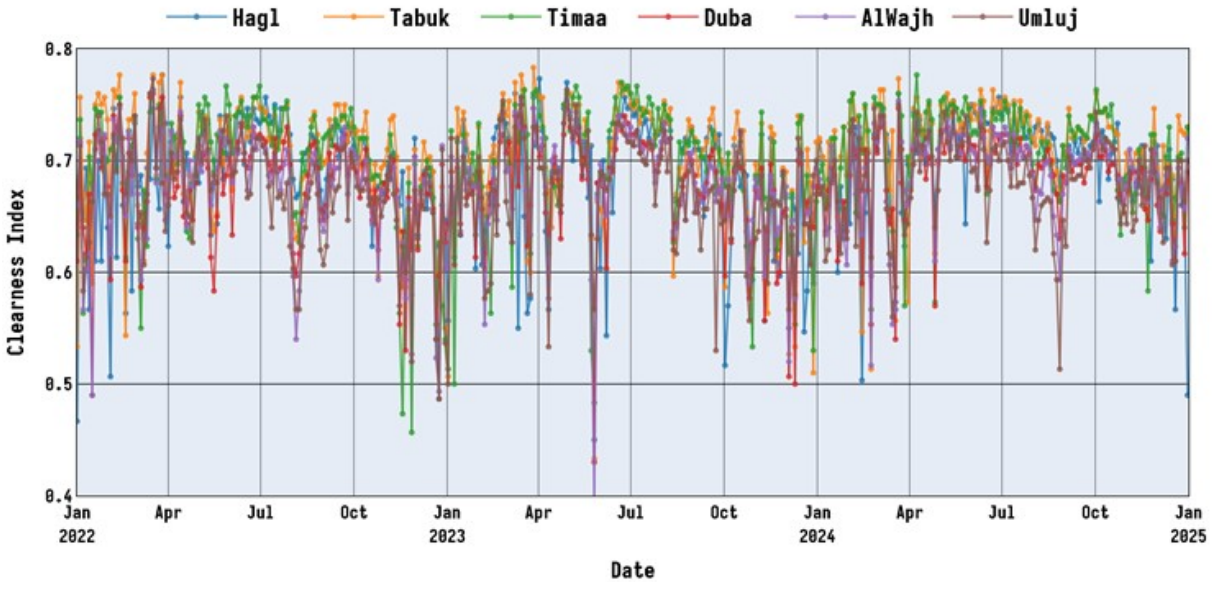


Figure 8: Clearness Index trends across chosen sites

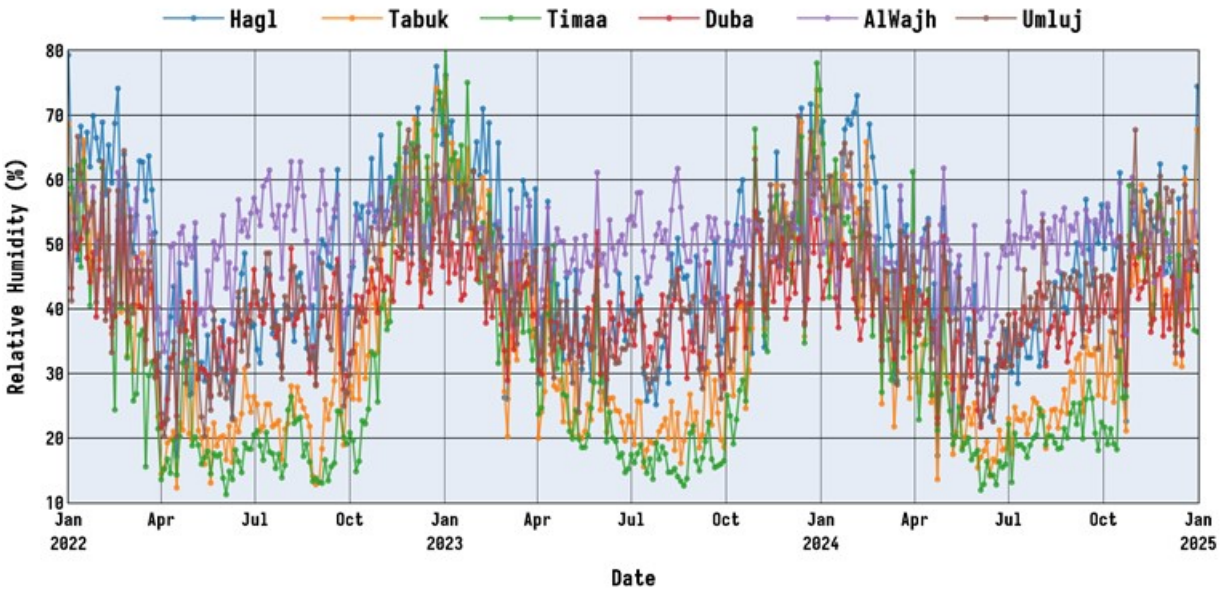


Figure 9: Relative Humidity trends across selected sites

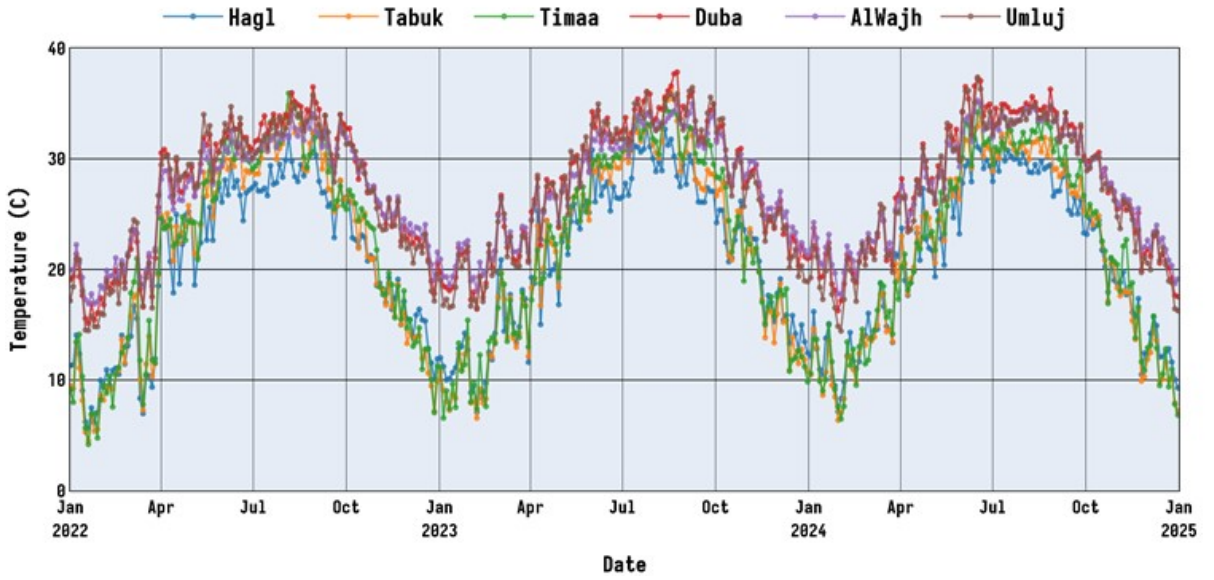


Figure 10: Temperature trends across selected sites

3.2 Model Performance Evaluation

The study utilizes daily mean values of DNI, tempe–rature, pressure, relative humidity, and wind speed over a period of three years, starting from 01 January 2022 and ending on 31 December 2024. The daily averages are obtained using the hourly mean values of each parameter. The data was obtained from NASA POWER. The performance comparison of eight (8) ML algorithms across six distinct climatic regions in KSA is carried out in terms of error metrics and the resulting values are depicted in Figure 11.The multi-metric and site-specific perspective, on how each algorithm performs across different regions, is presented.It captures relative trendsand central tendencies, trade-offs, and variability in model behaviour.

Multi-Metric Evaluation

In terms of MAE, models exhibit consistent variation across sites. The LRM notably records the highest MAE values atmost of the locations, peaking at Timaa with0.45 and Tabuk with0.4250. The ANN and DLM models show relatively lower MAE values of 0.32 and 0.31 for Umluj and Duba; respectively. A similar trend holds true for MSE and RMSE. However, RFR demon–strates relatively poor performance in Timaa (MSE: 0.335, RMSE: 0.575), while DLM and ANN maintain high accuracy at Dubaand Umlujwith MSE values of 0.150 and 0.155, respectively.

Regarding R^2, the DLM model maintains high performance across all sites, reaching 0.92 in Haql, Tabuk, and Duba.Similarly,ANN performs well, particularly in Haql(R^2=0.924) and Tabuk, Timaa, Duba, and AlWajh(R^2>0.91). However, models like KNN and GBR show relatively low values ofR^2, with KNN reaching a low of 0.865in AlWjhand the least is recorded forRFR in AlWajhat 0.851. SVR recorded the highest R^2 value of 0.893in Timaa.

Regarding MAPE, performance diverges more sharply. RFR and KNN report the highervalues between 5.11 and 7.89% corresponding to Duba for DLM and Timaa for RFR models. In general, the MAPE values are in the acceptable range and ANN and DLM models seem to be outperforming the others. MBE values

further differentiate model tendencies. ANN tend to show slightly negative biases of -0.013, -0.065, -0.013, and -0.007 at Haql, Tabuk, Alwajh, and Umluj; as can be observed from Radar diagram of Figure 11. DLM model showed slight positive biases at five sites in terms of MBE values varying from 0.025 to 0.214 at Umluj and Timaa with only negative bias of -0.00006 at AlWajh. However, GPR and SVR exhibit lower or slightly negative and positive biases atsomesites.

From the t-stat, values vary between -2.053 and 6.064 corresponding to models ANN and DLM at Tabuk and Timaa. It is worth to mention that t-stat values clustered around cluster around 1–2 and -0.1 to -2.0 for most models. rRMSEalso validatedDLM and ANN's robustness, with values consistently below 9% across most sites. DLM reports the lowest rRMSEof 5.99% at Duba, while LRM, RFR, and KNN frequently exceed 8%, particularly.

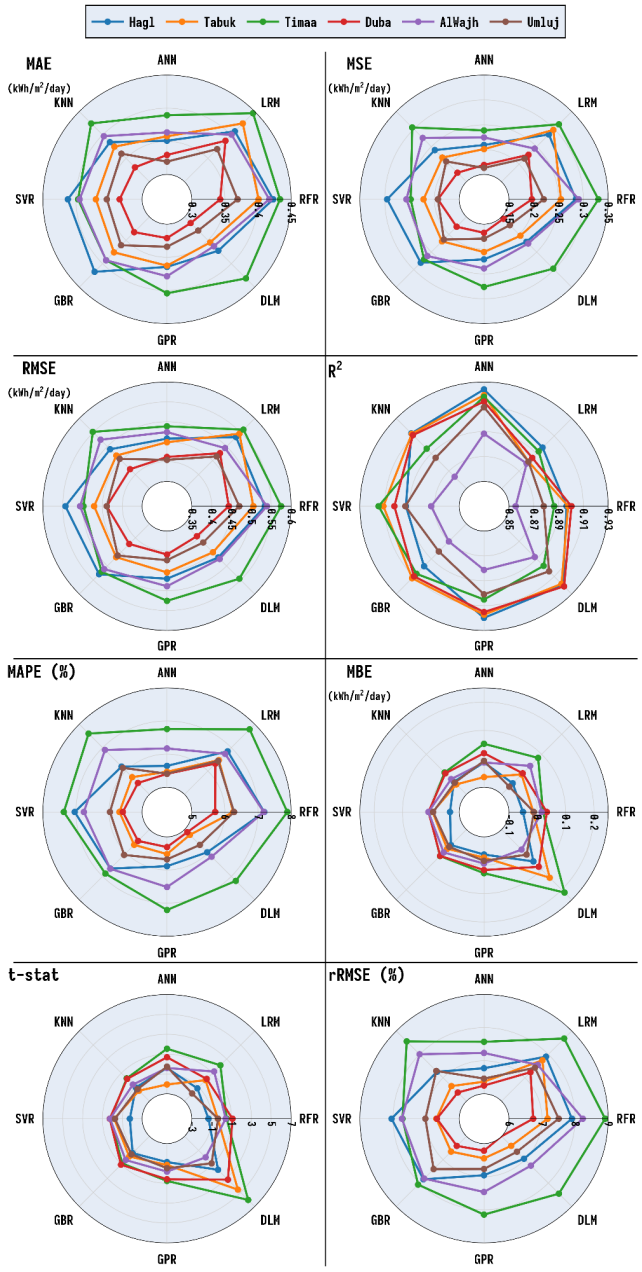


Figure 11: Multi-Metric Performance Across Sites

Multi-Site Evaluation

In Haql, ANNregistered the lowest RMSE of 0.3941kWh/m²/day and the highest R² of 0.9241. DLM also showed the lowest rRMSE (5.990%) and a positive bias in MBE (0.08564), indicating a slight over-estima–tion at Duba. ANN closely followed with an RMSE of 0.4113 and R² of 0.9143. LRM showed a reasonable R² (0.8851) butproduced a higher RMSE (0.4762) and MBE (0.0054). The highest MAE (0.4464) and RMSE (0.5599) in Dubawere associated with DLM and LRM, with an associated rRMSE of 8.68%. In Tabuk, ANN again produced the lowest RMSE at 0.4484, paired with the highest R² of 0.919. DLM closely matched this with an RMSE of 0.4513 and R² of 0.918. GPR also performed well (RMSE: 0.4541, R²: 0.917) and a MAPE of 5.52% against DLM (5.22%). LRMrecorded the highest error values atthis location, with RMSE of 0.546and rRMSE of 7.75%. KNN, SVR,and GBR also trailed behind, exhibiting RMSE values above 0.46and lower R² values under 0.92.

For Timaa, both ANN and GBR demonstrated competitive performance**,** with ANN achieving a lower RMSE (~0.469) and MAPE (~5.9%). The error spread among models was smaller at coastal sites, indicating relatively homogeneous radiative conditions. Never–theless, linear and simpler models showed inferior pre–dictive performance.At Duba, the coastal site, model performance showed increased atmospheric complexity due to humidity and aerosol effects. ANN provided the lowest RMSE (~0.486) and MAPE (~6.2%)**,** while GBR and RFR followed closely but with slightly higher bia–ses. The increased MAE and RMSE values compared to inland sites highlight the challenges of solar irradiance forecasting under marine-influenced conditions.

In Al Wajh, GBR marginally outperformed ANN in terms of RMSE (~0.498)**,** although ANN maintained a lower MAE and comparable MAPE. This suggests that ensemble models can better capture localized nonlinearities associated with coastal aerosol loading and cloud intermittency. Umluj presented a bit chal–lenging condition for accurate prediction, with all models exhibiting higher error levels compared to other sites. RMSE values exceeded 0.50 and MAPE appro–ached 7–8% for most models, particularly tree-based and kernel-based approaches. ANN still remained among the top performers but with reduced accuracy. These results highlight the need for site-specific tuning or hybrid modelling approaches for Umluj.

Across all six sites, ANN and GBR consistently ranked among the top-performing models, but the magnitude and ranking of errors varied by location. This site-wise variability in MAE, RMSE, MAPE, and MBE underscores the strong influence of local climatology and confirms thatno single machine learning model is universally optimalfor solar irradiance forecasting across the diverse geographic and atmospheric environ–ments of Saudi Arabia.

Figure 12 presents the model performance with respect to the variation in target variable (WS) expla–ined by the models for specific site-model combinations for two locations (Timaa and Duba) and for four models (RFR, ANN, SVR, and GBR) only for clarity.

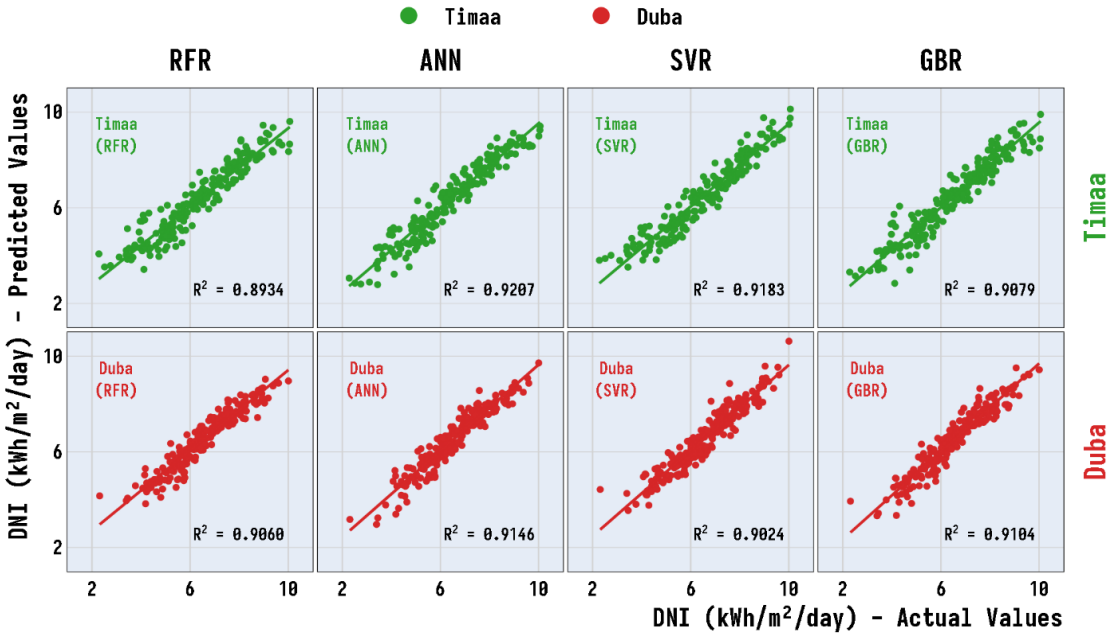


Figure 12: Actual vs. Predicted DNI across Sites and Models.

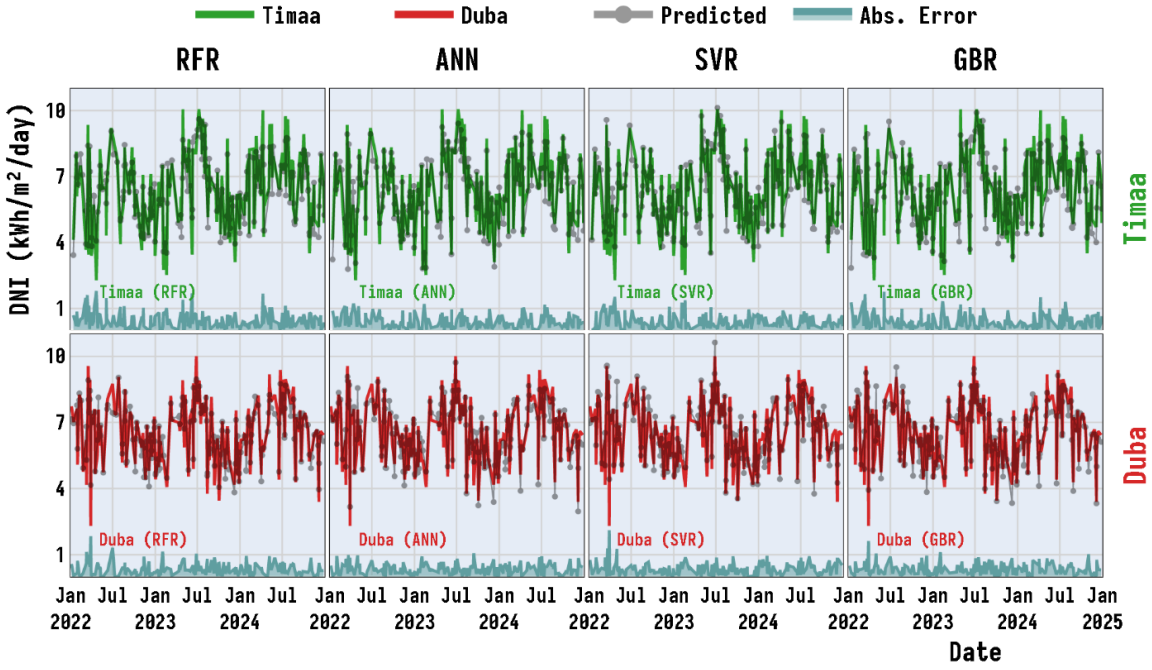


Figure 13. Prediction Patterns and Absolute Error Trends

The complete results on all the sites and for all the models are included in Figure A-1 in appendix-A. As shown in Figure A-1, ANN and DLM consistently demonstrate superior performance and near-perfect fits, with R^2 values exceeding 0.91atHaql, Tabuk, Timaa, Duba, and Umluj. On the other hand, LRM and RFR yielded the lowest R^2 values and showed more scattered prediction patterns, particularly in Timaa, AlWajh, and Umluj. Also, ANN demonstrated the consistently high R^2 scores across various sites, achieving an R^2 of 0.9241 in Haql. Similarly, in Tabuk, Timaa, Duba, and Umluj, ANN attained R^2 values of 0.9194, 0.9180, 0.9143, and 0.9097, respectively. In Alwajh, ANN recorded an R^2 of 0.8884, SVR recorded 0.8726 and LRM showing some promise with 0.8786, Figures 12 and A-1.

DLM performed remarkably well, recording the higher R^2 values of 0.9199, 0.9184, 0.9213, 0.9041 in Haql, Tabuk, Duba, and Umluj; respectively. However, the R^2 values at Timaa (0.8979) and Alwajh (0.8879)are a bit less than th0se discussed earlier.

SVR showed strong performance in Tabuk (0.9111), Duba (0.9023), and Timaa(0.9149), closely outperforming ANN and DLM in some instances, Figure 12. However, R^2 dropped in Haql(0.8929), AlWajh (0.8726) and Umluj(0.8934), see Figure A-1. GBR and GPR both produced moderately strong results, the former peaked in Tabuk (R^2 = 0.9120) and dropped to 0.8701 in AlWajh, while the latter maintained a relatively consis–tent performance across sites, scoring R^2 values around 0.905 and more. KNN demonstrated reasonably good fits in Haql(0.9129), Tabuk (0.9123), and Duba(0.9110), but performance compromise dslightly in Timaa (0.8954), AlWajh (0.8635), and Umluj(0.8851). RFR delivered R^2 of0.8555to 0.9005 at AlWajh and Duba.

LRM achieved highest R^2 value in Haql (0.8968) and lowest in AlWajh (0.8786), Figure A-1.

In addition to overall fit (see Figure A-1), the temporal performance of each model was assessed over the entire period. Figure 13 presents a comparative time series prediction trend of predicted versus actual values for Timaa and Duba covering RFR, ANN, SVR, and GBRmodels for clarity purpose only.

3.3 Temporal Performance Evaluation

However, the complete results for all the locations and the models are included as Figure A-2 in Appendix-A. In Figure 13, actual and predicted daily DNI values are plotted along with their corresponding absolute errors.

In Haql, predictions exhibit consistent seasonal trac–king across most models. ANN, DLM, and SVR show relatively smooth alignment with observed values, with visibly lower absolute errors across the mid-year high-radiation months. However, LRM, RFR and KNN disp–lay intermittent spikes in error, particularly around tran–sitional periods like April and October. For Tabuk, models ANN, SVR and DLM demonstrate minimal di–vergence from actual values, especially during peak summer, when the atmospheric conditions are relatively stable. The absolute error plots for SVR in Timaaremain lessaround the baseline throughout the year. On the other hand, LRM and KNN show pronounced error peaks between April and August.

In Duba, the predictive trajectories for DLM and ANN follow the actual DNI curve with considerable consistency, particularly from June through September. GBR and GPR show stable performance but RFR re–veals sharp deviations during the summermonths. AlWajh's results highlight increased volatility in pre–dictions across all models. KNN and LRM, showed larger deviations between predicted and actual values during the second and fourth quarters of the year. ANN maintains relative proximity to observed values during high solar periods, but frequent smaller oscillations in the absolute error trace indicate continuous minor pre–diction fluctuations.

In Umluj, both ANN and DLM show close agree–ment with actual DNI values for much of the year. The absolute error profiles remain suppressed throughout most of the year, notably from May to September. SVR and GBR also perform betterbut exhibit occasional surges in error, particularly during brief cloudy intervals typical in coastal regions.

It can be summarized that across all the sites, the non-linear ML models such as ANN, DLM, and GBR consistently provide excellent agreement between the predicted and observed DNI values, as observed from Figure 13 and Figure A-2. It is noticed that the site de–pendent climatic conditions strongly influence the predictability of the models with inland sites showing close agreement and coastal sites larger deviations.

3.4 Impact of Trigonometric Cyclical Encoding (TCE)

The study carried out a comparative analysis of feature importance to quantitatively evaluate the efficacy of TCE. It is analyzed that how TCE influences the explanatory power of temporal features. This was accomlished by training two separate RFR models under identical conditions. The first model utilised raw integer representations of temporal features (Month and Day), while the second employed cyclic encoded features (sin Month, cos Month, sin Day, and cos Day). The SHAP framework was then applied to both models to obtain a rigorous and consistent measure of each feature's marginal contribution to the ML algorithm[56]. For direct comparison, the sine and cosine components of each temporal concept were aggregated to represent the total impactof 'Cyclical Month' and 'Cyclical Day' features.

The Figure 14 presents SHAP summary plots for six sites (Hag1, Tabuk, Timaa, Duba, Alwajh, and Umluj), highlighting the relative importance and directional influence of input features on the model predictions. The results indicate that CI is the dominant predictor, exhibiting the widest spread of SHAP values, with higher CI values consistently contributing positively to the model prediction and lower values exerting a negative impact. The RH and TMP follow in importance, indication moderate but effective contributions, with mixed positive and negative SHAP distributions. The seasonal indicators (sin_MO and cos_MO) provide annual variability is captured by the model but cannot be the primary driver of predictions. In contrast, daily cycle variables (sin_DY and cos_DY) and WS exhibit SHAP values tightly clustered around zero. This simply means that these parameters have a marginal role in shaping the model output. Overall, the SHAP analysis confirms that the model behavior is mainly dominated by CI and other meteorological parameters.

The mean absolute SHAP importance of temporal features (month and day) using raw versus cyclical encoding across six sites (Haql, Tabuk, Timaa, Duba, Alwajh, and Umluj) is shown in Figure 15. For all locations, cyclical encoding markedly enhances the importance of the month feature, with substantial relative increases ranging from about 50% to over 130%, indicating a much-improved capture of seasonal periodicity. The day feature also shows consistent, though smaller, gains under cyclical encoding, reflecting better representation of intra-month continuity. Among the sites, Timaa and Duba exhibit the strongest sensitivity to cyclical month encoding, while Alwajh and Umluj show comparatively moderate but still notable improvements. Overall, the outcome demonstrates that cyclical encoding more effectively represents temporal periodicity than raw encoding, leading to stronger and more physically meaningful contributions of temporal features to the model across all sites.

4. CONCLUSION

This study comprehensively evaluated the performance of eight ML models for forecasting solar radiation across six climatically diverse sites in Saudi Arabia. The models evaluated include RFR, LRM, ANN, KNN, SVR, GBR, GPR, and DLM. Eight statistical metrics were used to assess predictive accuracy and genera–lizability across each site.

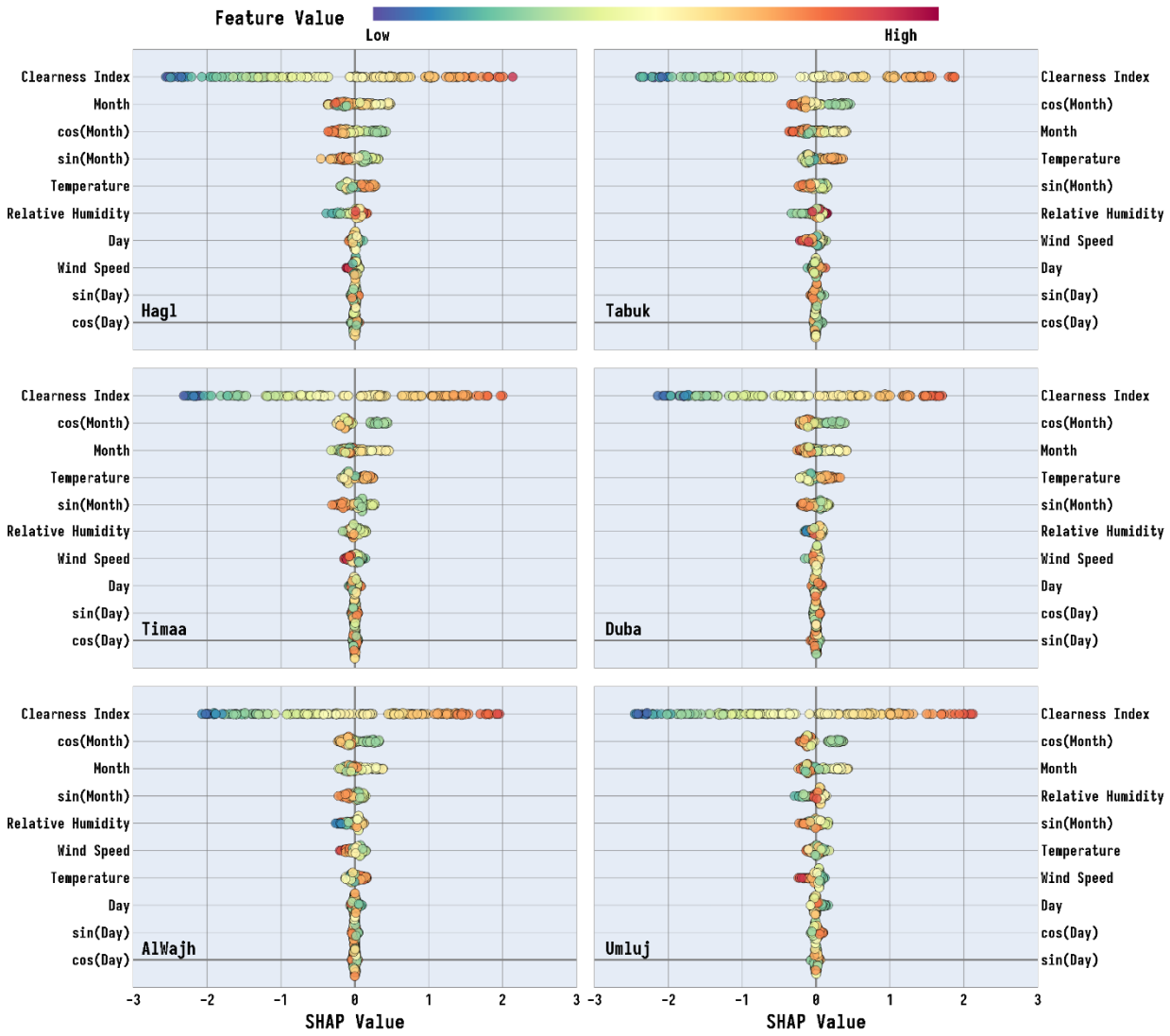


Figure 14. Comparison of SHAP summary of cyclically encoded features and raw features

The key findings, derived from an extensive multi metric-site evaluation, are summarized as follows:

- Deep Learning (DLM) and Artificial Neural Net–work (ANN) models demonstrated superior and consistent performance across most locations, with DLM achieving the lowest RMSE (as low as 0.3941 kWh/m²/day in Duba) and ANN showing remarkable stability and low errors (e.g., MAPE of 5.23% in Tabuk).
- Model effectiveness was significantly influenced by geographical and climatic conditions. Support Vec–tor Regression (SVR) excelled in specific arid in–land regions like Timaaand Tabuk, while other models such as RFR and KNN exhibited greater performance volatility.
- The implementation of Trigonometric Cyclical Encoding (TCE) for temporal features substantially enhanced model learning. A comparative analysis revealed that TCE increased the feature importance of temporal signals by over 50.1% to131.6% for monthly cycles and 13.4% to 58.4% for daily cycles, enabling models to more effectively capture fundamental periodic patterns in solar radiation.
- Time series and error analysis confirmed that ANN and DLM maintained the most stable prediction accuracy, particularly during high solar radiation seasons, whereas other models showed wider fluctuations.

Finally, based on the entire data analysis, the models are prioritised and ranked from 1 to 8, as given in Table 8.

Table 8. Ranking of Model Performance.

Model	Key Strengths	Key Limitations	Rank
ANN	Excellent MAPE and t-stat in variable climates, adapts to non-linearity	Requires careful tuning and larger datasets	1
DLM	Competitive in high-data scenarios	Exhibits systematic MBE	2
GPR	Low bias, reliable in moderate variability	Computationally intensive for large datasets	3
KNN	Good performance in stable inland climates	Poor in high-variability sites	4
SVR	Adequate in linear regimes	Poor in non-linear, climates; sensitive parameter choice	5
GBR	Lowest MAE/RMSE,	Slightly higher computational cost	6

	minimal bias, robust across all sites and irradiance levels		
LRM	Simple and fast	Highest errors and unable to capture non-linearity	7
RFR	Consistent and robust in high-irradiance	Slightly less accurate than GBR	8

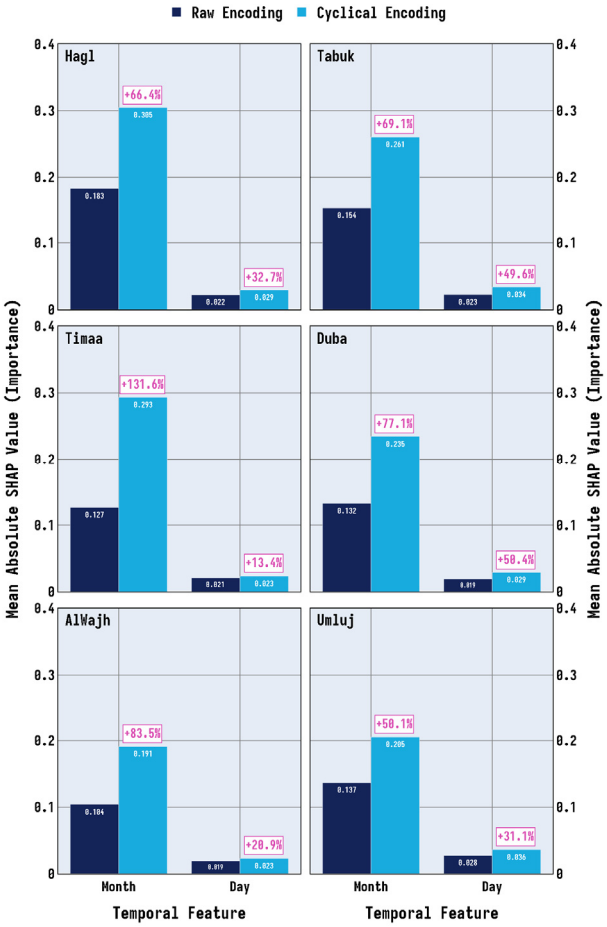


Figure 15. Comparison of aggregated SHAP importance for raw versus cyclically encoded temporal features

4.1 Limitations and future work

While this study provides a robust evaluation of ML models for intra-annual solar radiation forecasting, it is important to note its limitation regarding the temporal scope of the dataset. The use of data from a single calendar year (2023), while meticulously analyzed, may not encompass the full spectrum of inter-annual climatic variability and extreme weather anomalies specific to the regions of Saudi Arabia. Consequently, the absolute values of the reported error metrics should be inter–preted within this context. Future work will focus on expanding the dataset to include multiple years of historical data. This will allow for the development of more generalized models that are resilient to long-term climatic shifts and rare meteorological events, further enhancing their operational reliability for grid integration.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

LBR: Conceptualization, Methodology, Software, For–mal analysis, Investigation, Data curation, Writing – Original Draft; SZS: Visualization, Validation, Super–vision, Resources, Writing – Review & Editing; SR: Supervision, Validation, Project administration, Writing – Review & Editing. All authors had approved the final version.

ACKNOWLEDGMENT

Authors would like to acknowledge the technical sup–port provided by King Fahd University of Petroleum & Minerals to accomplish the research work reported in this manuscript.

DATA AVAILABILITY

The data that supports the findings of this study are ava–ilable from the corresponding author upon reaso–nable request.

REFERENCES

[1] M.D. Margaritou, E. Tzannatos, "A multi-criteria optimization approach for solar energy and wind power technologies in shipping," *FME Transactions*, vol. 46, pp. 374–380, 2018, doi: 10.5937/fmet1803374M.

[2] S. Bezari, S.M.E.A. Bekkouche, A. Benchatti, "Investigation and Improvement for a Solar Greenhouse Using Sensible Heat Storage Material," *FME Transactions*, vol. 49, pp. 154–162, 2020, doi: 10.5937/FME2101154B.

[3] B.P. Rašuo, A. Bengin, "Optimization of wind farm layout," *FME Transactions*, vol. 38, pp. 107–114, 2010,

[4] P. Dumka, D.R. Mishra, "Comparative experi–mental evaluation of conventional solar still (CSS) and CSS augmented with wax filled metallic finned-cups," *FME Transactions*, vol. 48, pp. 482–495, 2020, doi: 10.5937/FME 2002482D.

[5] M. Gojak, F. Ljubinac, M. Banjac, "Simulation of solar water heating system," *FME Tran–sactions*, vol. 47, pp. 1–6, 2019, doi: 10.5937 /fmet1901001G.

[6] B. Rašuo, M. Dinulović, A. Veg, A. Grbović, A. Bengin, "Harmonization of new wind turbine rotor blades development process: A review," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 874–882, 2014, doi: 10.1016/j.rser.2014.07.137.

[7] V. Parezanovic, B. Rasuo, M. Adzic, "Design of airfoils for wind turbine blades," 2005.

[8] M. Stojicevic, Z. Jeli, M. Obradovic, R. Obra–dovic, G.C. Marinescu, "Designs of solar con–centrators," *FME Transactions*, vol. 47, pp. 273–278, 2019, doi: 10.5937/fmet1902273S.

[9] F.M. Hussain, S. Rehman, F.A. Al-Sulaiman, "Per–formance Analysis of a Solar Chimney Power Plant for Different Geographical Loca–tions of Saudi Arabia," *FME Transactions*, vol. 49, pp. 64–71, 2020, doi: 10.5937/FME2101064H.

[10] G. Habtay, J. Buzas, I. Farkas, "Heat Transfer analysis in the chimney of the indirect solar dryer under natural convection mode," *FME Transactions*, vol. 48, pp. 701–706, 2020, doi: 10.5937/fme2003701H.

[11] A. Jakoplić, D. Franković, V. Kirinčić, T. Plavšić, "Benefits of short-term photovoltaic power production forecasting to the power system," *Optimization and Engineering*, vol. 22, pp. 9–27, 2021, doi: 10.1007/s11081-020-09583-y.

[12] X. Qing, Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," *Energy*, vol. 148, pp. 461–468, 2018, doi: 10.1016/j.energy.2018.01.177.

[13] H.C. Bayrakçı, C. Demircan, A. Keçebaş, "The development of empirical models for estimating global solar radiation on horizontal surface: A case study," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 2771–2782, 2018, doi: 10.1016/j.rser.2017.06.082.

[14] H. Hissou, S. Benkirane, A. Guezzaz, M. Azrour, A. Beni-Hssane, "A Novel Machine Learning Approach for Solar Radiation Estimation," *Sustainability (Switzerland)*, vol. 15, pp. 10609, 2023, doi: 10.3390/su151310609.

[15] E.S. Solano, C.M. Affonso, "Solar Irradiation Forecasting Using Ensemble Voting Based on Machine Learning Algorithms," *Sustainability (Switzerland)*, vol. 15, pp. 7943, 2023, doi: 10.3390/su15107943.

[16] S. Rehman, U.T. Salman, M.A. Mohandes, F.A. Al-Sulaiman, S. Adetona, L.M. Alhems, M.A. Baseer, "Wind Speed Prediction Based on Long-Short Term Memory using Nonlinear Autoregressive Neural Networks," *FME Transactions*, vol. 50, pp. 260–270, 2022, doi: 10.5937/fme2201260R.

[17] S. Rehman, S.A. Khan, L.M. Alhems, "The Effect of Acceleration Coefficients in Particle Swarm Optimization Algorithm with Application to Wind Farm Layout Design," *FME Transactions*, vol. 48, pp. 922–930, 2020, doi: 10.5937/fme2004922R.

[18] N. Mirkov, B. Rašuo, S. Kenjereš, "On the improved finite volume procedure for simulation of turbulent flows over real complex terrains," *Journal of Computational Physics*, vol. 287, pp. 18–45, 2015, doi: 10.1016/j.jcp.2015.02.001.

[19] M. Mohandes, H.H. Nuha, S.A. Mugitama, S. Rehman, A. Al-Shailkhi, "Global solar radiation prediction using machine learning approaches," *Sigma Journal of Engineering and Natural Sciences*, vol. 43, pp. 1725–1736, 2025, doi: 10.14744/sigma.2024.00128.

[20] M. Chaibi et al., "Machine Learning Models Based on Random Forest Feature Selection and Bayesian Optimization for Predicting Daily Global Solar Radiation," *International Journal of Renewable Energy Development*, vol. 11, pp. 309–323, 2022, doi: 10.14710/IJRED.2022.41451.

[21] H. Bakır, "Prediction of daily global solar radiation in different climatic conditions using metaheuristic search algorithms: a case study from Türkiye," *Environmental Science and Pollution Research*, vol. 31, pp. 43211–43237, 2024, doi: 10.1007/s11356-024-33785-x.

[22] Y. Lu, L. Wang, C. Zhu, L. Zou, M. Zhang, L. Feng, Q. Cao, "Predicting surface solar radiation using a hybrid radiative Transfer–Machine learning model," *Renewable and Sustainable Energy Reviews*, vol. 173, pp. 113105, 2023, doi: 10.1016/j.rser.2022.113105.

[23] D.S. Khafaga, A.A. Alhussan, M.M. Eid, E.S.M. El-kenawy, "Improving solar radiation source efficiency using adaptive dynamic squirrel search optimization algorithm and long short-term memory," *Frontiers in Energy Research*, vol. 11, 2023, doi: 10.3389/fenrg.2023.1164528.

[24] E. Zell, S. Gasim, S. Wilcox, S. Katamoura, T. Stoffel, H. Shibli, J. Engel-Cox, M. Al Subie, "Assessment of solar radiation resources in Saudi Arabia," *Solar Energy*, vol. 119, pp. 422–438, 2015, doi: 10.1016/j.solener.2015.06.031.

[25] M. Mohandes, S.A. Khan, S. Rehman, S. Arabia, A. Al-Shaikhi, B. Liu, K. Iqbal, "GARM: A Stochastic Evolution based Genetic Algorithm with Rewarding Mechanism for Wind Farm Layout Optimization," *FME Transactions*, vol. 51, pp. 575–584, 2023, doi: 10.5937/fme2304575M.

[26] G.S. Živković, N.S. Mirkov, D. V. Dakić, M.R. Mladenović, A.M. Erić, M.D. Erić, N.R. Rudonja, "Numerical simulation of thermo-fluid properties and optimisation of hot water storage tank in biomass heating systems," *FME Transactions*, vol. 38, pp. 63–70, 2010,.

[27] T. Bin Nadeem, S.U. Ali, M. Asif, H.K. Suberi, "Forecasting daily solar radiation: An evaluation and comparison of machine learning algorithms," *AIP Advances*, vol. 14, 2024, doi: 10.1063/5.0211723.

[28] M.K. Hossain, M. Arifuzzaman, M.E. Seliaman, A. Rahman, D. Sarker, H. Altammar, Ensemble Learning Algorithms for Solar Power Prediction in Saudi Arabia: A Data-Driven Approach, in: 2024 ASU Int. Conf. Emerg. Technol. Sustain. Intell. Syst. ICETSIS 2024, 2024,: pp. 1368–1372 doi: 10.1109/ICETSIS61505.2024.10459648.

[29] H. Hissou et al., Advanced Prediction of Solar Radiation Using Machine Learning and Principal Component Analysis, in: Lect. Notes Networks Syst., 2024,: pp. 201–207 doi: 10.1007/978-3-031-48573-2_29.

[30] C.G. Villegas-Mier, J. Rodriguez-Resendiz, J.M. Álvarez-Alvarado, H. Jiménez-Hernández, Á. Odry, "Optimized Random Forest for Solar Radiation Prediction Using Sunshine Hours," *Micromachines*, vol. 13, pp. 1406, 2022, doi: 10.3390/mi13091406.

[31] S. Wang, J. Ma, "A novel GBDT-BiLSTM hybrid model on improving day-ahead photovoltaic prediction," *Scientific Reports*, vol. 13, pp. 15113, 2023, doi: 10.1038/s41598-023-42153-7.

[32] J. Duan, H. Zuo, Y. Bai, M. Chang, X. Chen, W. Wang, L. Ma, B. Chen, "A multistep short-term

solar radiation forecasting model using fully con–volutional neural networks and chaotic aquila optimization combining WRF-Solar model results," *Energy*, vol. 271, pp. 126980, 2023, doi: 10.1016/j.energy.2023.126980.

[33] S. Rehman, M. Mohandes, "Splitting global solar radiation into diffuse and direct normal fractions using artificial neural networks," *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, vol. 34, pp. 1326–1336, 2012, doi: 10.1080/15567031003792403.

[34] W. Tercha, S.A. Tadjer, F. Chekired, L. Canale, "Machine Learning-Based Forecasting of Tempe–rature and Solar Irradiance for Photovoltaic Sys–tems," *Energies*, vol. 17, pp. 1124, 2024, doi: 10.3390/en17051124.

[35] O. Dikmen, "Predicting Solar Irradiance Using Machine Learning Approaches: The Case of Duzce, Turkey," *International Journal of Advanced Natu–ral Sciences and Engineering Researches*, vol. 8, pp. 133–145, 2024,. https://as-proceeding.com/index.php/ijanser.

[36] S. Soleymani, S. Mohammadzadeh, "Comparative Analysis of Machine Learning Algorithms for Solar Irradiance Forecasting in Smart Grids," 2023,. http://arxiv.org/abs/2310.13791.

[37] M. Mohandes, A. Balghonaim, M. Kassas, S. Rehman, T.O. Halawani, "Use of radial basis func–tions for estimating monthly mean daily solar radi–ation," *Solar Energy*, vol. 68, pp. 161–168, 2000, doi: 10.1016/S0038-092X(99)00071-7.

[38] B.P. Rasuo, A.D. Veg, Design, fabrication and verification testing of the wind turbine rotor blades from composite materials, in: ICCM Int. Conf. Compos. Mater., 2007, Kyoto, Japan: pp. 1–4.

[39] M.R. Dinulović, M.R. Trninić, B.P. Rašuo, D. V. Kožović, "Methodology for Aeroacoustic Noise Analysis of 3-Bladed H-Darrieus Wind Turbine," *Thermal Science*, vol. 27, pp. 61–69, 2023, doi: 10.2298/TSCI2301061D.

[40] B. Rašuo, A. Bengin, A. Veg, "On Aerodynamic Optimization of Wind Farm Layout," *Pamm*, vol. 10, pp. 539–540, 2010, doi: 10.1002/pamm.201010262.

[41] S.S. Mousavi, M. Schukat, E. Howley, Deep Rein–forcement Learning: An Overview, in: Lect. Notes Networks Syst., 2018,: pp. 426–440 doi: 10.1007/978-3-319-56991-8_32.

[42] O. Rohanian et al., "Using Bottleneck Adapters to Identify Cancer in Clinical Notes under Low-Re–source Constraints," *Proceedings of the Annual Meeting of the Association for Computational Lin–guistics*, pp. 62–78, 2023, doi: 10.18653/v1/2023.bionlp-1.5.

[43] A.A. Alabdulhadi, S. Rehman, A. Ali, M. Shafi–ullah, "Deep learning framework for wind speed prediction in Saudi Arabia," *Neural Computing and Applications*, vol. 37, pp. 3685–3701, 2025, doi: 10.1007/s00521-024-10766-2.

[44] S. Rehman, M. Mohandes, "Artificial neural net–work estimation of global solar radiation using air temperature and relative humidity," *Energy Policy*, vol. 36, pp. 571–576, 2008, doi: 10.1016/j.enpol.2007.09.033.

[45] V.K. Chauhan, J. Zhou, P. Lu, S. Molaei, D.A. Clifton, "A brief review of hypernetworks in deep learning," *Artificial Intelligence Review*, vol. 57, pp. 250, 2024, doi: 10.1007/s10462-024-10862-8.

[46] K. Uçak, G.Ö. Günel, "Adaptive stable back–stepping controller based on support vector regres–sion for nonlinear systems," *Engineering Applica–tions of Artificial Intelligence*, vol. 129, pp. 107533, 2024, doi: 10.1016/j.engappai.2023.107533.

[47] M.F. Tahir, M.Z. Yousaf, A. Tzes, M.S. El Moursi, T.H.M. El-Fouly, "Enhanced solar photovoltaic power prediction using diverse machine learning algorithms with hyperparameter optimization," *Renewable and Sustainable Energy Reviews*, vol. 200, pp. 114581, 2024, doi: 10.1016/j.rser.2024.114581.

[48] S.R. Young, D.C. Rose, T.P. Karnowski, S.H. Lim, R.M. Patton, Optimizing deep learning hyper-para–meters through an evolutionary algorithm, in: Proc. MLHPC 2015 Mach. Learn. High-Performance Comput. Environ. - Held Conjunction with SC 2015 Int. Conf. High Perform. Comput. Net–working, Storage Anal., 2015,: pp. 1–5 doi: 10.1145/2834892.2834896.

[49] V. V. Gurenko, B.I. Bychkov, V. V. Syuzev, An Approach to Simulation of Stationary and Non-sta–tionary Processes in the Harmonic Basis, in: Proc. 2021 IEEE Conf. Russ. Young Res. Electr. Elec–tron. Eng. ElConRus 2021, 2021,: pp. 2664–2667 doi: 10.1109/ElConRus51938.2021.9396627.

[50] M.V. Shcherbakov, A. Brebels, N.L. Shcherbakova, A.P. Tyukov, T.A. Janovsky, V.A. evich Kamaev, "A survey of forecast error measures," *World App–lied Sciences Journal*, vol. 24, pp. 171–176, 2013, doi: 10.5829/idosi.wasj.2013.24.itmies.80032.

[51] C. Chen, J. Twycross, J.M. Garibaldi, "A new ac–curacy measure based on bounded relative error for time series forecasting," *PLoS ONE*, vol. 12, pp. e0174202, 2017, doi: 10.1371/journal.pone. 0174202.

[52] H. Zang, L. Cheng, T. Ding, K.W. Cheung, M. Wang, Z. Wei, G. Sun, "Application of functional deep belief network for estimating daily global solar radiation: A case study in China," *Energy*, vol. 191, pp. 116502, 2020, doi: 10.1016/j.energy.2019.116502.

[53] L. Yang, Q. Cao, Y. Yu, Y. Liu, "Comparison of daily diffuse radiation models in regions of China without solar radiation measurement," *Energy*, vol. 191, pp. 116571, 2020, doi: 10.1016/j.energy.2019.116571.

[54] S.G. Gouda, Z. Hussein, S. Luo, Q. Yuan, "Model selection for accurate daily global solar radiation prediction in China," *Journal of Cleaner Pro–duction*, vol. 221, pp. 132–144, 2019, doi: 10.1016/j.jclepro.2019.02.211.

[55] J. Fan, X. Wang, L. Wu, F. Zhang, H. Bai, X. Lu, Y. Xiang, "New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: A case study in South China," *Energy Conversion and Management*, vol. 156, pp. 618–625, 2018, doi: 10.1016/j.enconman.2017.11.085.

[56] O. Petrosian, Y. Zhang, "Solar Power Generation Forecasting in Smart Cities and Explanation Based on Explainable AI," *Smart Cities*, vol. 7, pp. 3388–3411, 2024, doi: 10.3390/smartcities7060132.

NOMENCLATURE

ADSSOA	Adaptive Dynamic Squirrel Search Optimization Algorithm
AI	Artificial Intelligence
CART	Classification And Regression Tree
DNI	Direct Normal Irradiance
DNN	Deep Neural Networks
DNR	Direct Normal Radiation
DSR	Direct Solar Radiation
DT	Decision Tree
FE	Feature Engineering
FS	Feature Selection
GA	Genetic Algorithm
GBM	Gradient Boosting Machine
GHI	Global Horizontal Irradiance
GPR	Gaussian Process Regression
GWO	Grey Wolf Optimizer
HHO	Harris Hawks Optimization
IEA	International Energy Agency
IRENA	International Renewable Energy Agency
KSA	Kingdom of Saudi Arabia
LR	Linear Regression
LSTNet	Learning Spectral Transformer Network
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLP	Multi-Layer Perceptron
NASA	National Aeronautics and Space Administration
NLP	Natural Language Processing
POWER	Prediction of Worldwide Energy Resources
PSO	Particle Swarm Optimization
PV	Photovoltaic
RF	Random Forest
RFE	Recursive Feature Elimination
RFR	Random Forest Regressor
RMSE	Root Mean Squared Error
R2	R-squared
RTM	Referential Translation Machine
SDG	Sustainable Development Goal
SGDR	Stochastic Gradient Descent Regressor
SVR-BO	Support Vector Regression- Bayesian Optimization
TCE	Trigonometric Cyclic Encoding
UN	United Nations
XGBoost	Extreme Gradient Boosting

APPENDIX – A

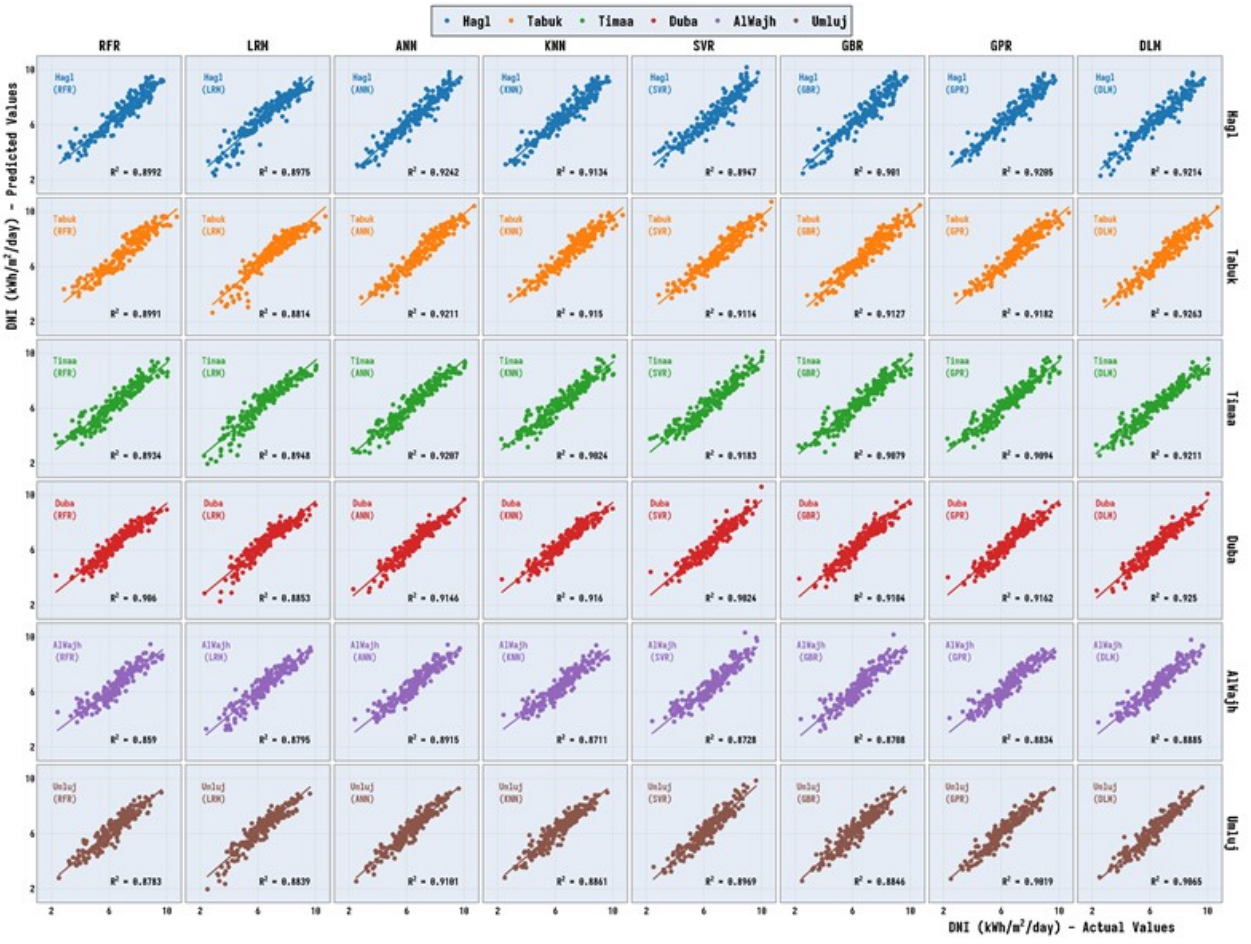


Figure A-1: Actual vs. Predicted DNI across Sites and Models.

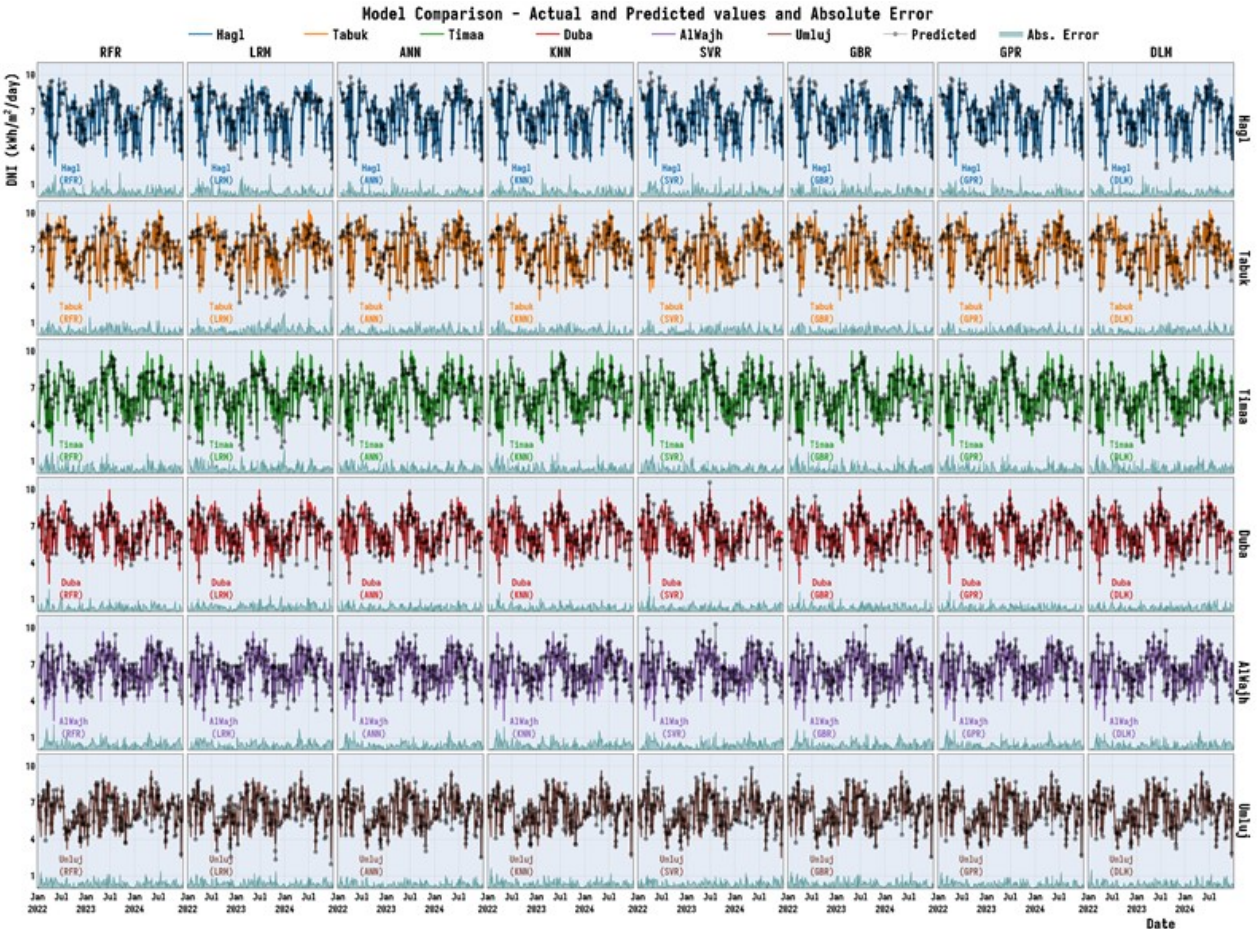


Figure A-2. Prediction Patterns and Absolute Error Trends

ПРЕДВИЂАЊЕ ДИРЕКТНОГ СУНЧЕВОГ ЗРАЧЕЊА КОРИШЋЕЊЕМ ВИШЕМОДЕЛСКЕ ЕВАЛУАЦИЈЕ СА ТРИГОНОМЕТРИЈСКИМ ЦИКЛИЧНИМ ПРОЦЕСОМ

Л.Б. Рашид, Ш.Ц. Шуџа, Ш. Рехман

Тачно предвиђање обновљивих извора уопште и сунчевог зрачења је кључно за оптималну интеграцију система соларне енергије. Студија истражује осам модела машинског учења, наиме модел линеарне регресије (LRM), регресор случајне шуме (RFR), регресор градијентног појачавања (GBR), Гаусову процесну регресију (GPR), вештачку неуронску мрежу (ANN), k-најближе суседе (NN), регресију вектора подршке (SVR) и дубинско учење (DL) за предвиђање директног сунчевог зрачења на шест климатски различитих локација у Саудијској Арабији. Модели се евалуирају коришћењем осам статистичких метрика, заједно са временским серијама и анализама апсолутних грешака. У овом раду је уведено тригонометријско циклично кодирање (TCE), које је значајно побољшало темпорално учење. Компаративна анализа заснована на SHAP-у открила је да је TCE побољшао објашњавајућу моћ временских карактеристика за 49,26% и 53,40% за месечне и дневне циклусе. Резултати показују да је DL постигао најнижу средњоквадратну грешку (RMSE) и највиши коефицијент детерминације, док је ANN константно показивао високу тачност на свим локацијама. Анализе грешака и временских серија указују на стабилна предвиђања ANN и DL; док су LR, RFR и k-најближи суседи (NN) показали веће флуктуације. Предложена TCE техника је додатно побољшала резултат модела одржавајући укупну подобност модела између 81,79% и 94,36% у свим сценаријима. Ова студија појачава ефикасно планирање интег–рације соларне енергије у различитим климатским условима.