Tree- Enhanced Deep Neural Network Framework for Forecasting Solar Power Generation

Shereen A. El-aal, Yomna M. Elbarawy, Asmaa Mohamed, Sahar A. Gomaa and Nahed El Desouky

Abstract—**Photovoltaic (PV) systems have gained recent attention as a crucial component of the advancement of renewable energy. Smooth solar power generation requires precise and trustworthy forecasts to maximize the performance of the power grid and to prevent issues with solar generation due to instability. This research presents a study that employes regression models with feature engineering combined with Deep Neural Networks (DNN) to forecast solar power generation. The proposed model used a Tree-Enhanced Deep Neural Network (TEDNN) architecture. It is divided into three phases: phase one, optimizing the Hyperparameters (HPs) of the regression models. The HPs were optimized using four algorithms, Grid Search (GS), and Bayesian Optimization (BO), and Particle Swarm (PSO). In phase two, two feature engineering techniques (Squared Irradiation and Temperature Differential) were applied over extracted data from phase one. Phase three combines different tree-based regression algorithms with DNN to forecast solar power generation. Multiple regression techniques were compared, including Decision Tree Regression (DTR), Random Forest Regression (RFR), Gradient Boosting Regression (GBR) and Xgboost Regression (XgbR) techniques. The proposed model was applied over a real-world data from a PV solar power plant in India and evaluated according to performance metrics. The results show that the proposed model can accurately forecast solar power with GBR+DNN (With Feature Engineering), achieves the highest closeness coefficient (0.781) due to strong performance across all metrics. Furthermore, it was found that the PSO outperforms BO and GS in terms of accuracy and execution time. In addition, feature engineering improves performance, as models with feature engineering dominate the top 3 ranks and models combined with DNN show significantly better results than their non-DNN counterparts.**

Index Terms—**Photovoltaic Solar Power, Bayesian Optimization, Grid Search, Particle Swarm Optimization, Decision Tree, Random Forest, Gradient Boosting, Xgboost**

I. INTRODUCTION

AN essential part of producing renewable energy is photovoltaic (PV) energy generation. Because of its abundance, cleanliness, and environmental friendliness, PV energy has gradually increased in popularity in recent years [1]. Furthermore, the generated power from PV plant energy can differ according to time, location, and the panel size. In addition, it is also influenced by the atmosphere's conditions such as irradiation, temperature, wind speed, and humidity [2]. Forecasting PV solar power is an important task for the efficient management of renewable energy resources [3], [4]. Accurate forecasting of solar power generation can help to reduce the risk of over-supply or under-supply of electricity, and to optimize the use of available resources. Furthermore, power forecasting keeps the balance between power production and consumption, which contributes to the financial benefits of electrical utilities. Moreover, forecasting reduces the additional costs related to weather dependence and raises the quality of the energy delivered to the grid [5]. Statistical analysis and machine learning are two significant strategies for solar power forecasting that compete with one another. Supervised machine learning algorithms are frequently applied for classification and forecasting problems [6], [7]. Regression-based supervised machine learning attempts to effectively predict the continuous outputs based on the input data. Moreover, regression models that use statistical methods such as decision tree regression to discover the link between a criterion and a predictor variable can be used to forecast the generated power [8]. In addition, the process by which an output is produced from an input and a series of decisions is also parametrically represented as each decision in the decision tree algorithm [9]. Random forest regression is an ensemble learning algorithm that combines multiple decision tree to improve performance, which is known as a bagging algorithm, while gradient boosting regression combines multiple weak models to create a strong model. XgbR is a more regular model that uses advanced configuration to improve model capabilities [10].

To adapt a machine learning model to a dataset, certain HPs must be set. Nominating the optimal HP combination is challenging to achieve optimal performance of a prediction model that minimizes a predetermined loss function on given independent data. Optimization techniques are frequently used to improve the predictive ability of Machine Learning (ML) algorithms because the performances of these models are highly sensitive to their HPs [11]. The authors in [12] provide a practical solution for improving the performance of machine learning models by optimizing their HPs based on BO and comparing the results with grid search and random search methods. Choosing an optimization method is not entirely easy, as some of these methods are suitable for small HP configuration spaces such as BO and others for large configuration spaces such as PSO [13]. Furthermore, PSO has the benefits of quick convergence, minimal processing time, and excellent precision [14]. GS is one of the optimization algorithms that evaluates the model for a given HP vector using cross-validation. The independence of the HP settings is one benefit of the grid search; due

Manuscript received May 8, 2025; revised July 16, 2025.

Shereen A. El-aal is Lecturer at Department of Mathematics, Faculty of Science, Al-Azhar University, Cairo, Egypt (E-mail: shereen.a.elaal@azhar.edu.eg).

Yomna M. Elbarawy is Lecturer at Department of Mathematics, Faculty of Science, Al-Azhar University, Cairo, Egypt (E-mail: y.elbarawy@azhar.edu.eg).

Asmaa Mohamed is Lecturer at Department of Mathematics, Faculty of Science, Al-Azhar University, Cairo, Egypt (E-mail: asmaamohamed89@azhar.edu.eg).

Sahar A. Gomaa is Lecturer at Department of Mathematics, Faculty of Science, Al-Azhar University, Cairo, Egypt (E-mail: sahargomaa@azhar.edu.eg).

Nahed El Desouky is Assistant Professor at Department of Mathematics, Faculty of Science, Al-Azhar University, Cairo, Egypt (E-mail: nahedeldesouky5922@azhar.edu.eg).

to this, parallel computation can be done [15], [16], while BO is a sequential optimization technique that builds a probability model of the objective function to choose the most favorable HPs to test against the objective function [17].

The literature on solar PV forecast estimation presents different types of models, some of which are illustrated in Table I which summarizes some research work on solar forecasting using the most popular machine learning algorithms. As this brief review indicates, irradiance and temperature are the most frequently utilized features in solar power forecasting, making them essential features to estimate a PV system's effectiveness. In addition, there are numerous machine learning models that have been utilized for forecasting solar power; the variations among these models are related to their formulation, forecasting range, data processing, dependency, and model sophistication.

In this paper, a novel approach for forecasting solar power generation is presented using feature engineering and DNN employed to enhance models' output. Additionally, this paper will provide an overview of the HPs behind each model and how they can be tuned for optimal performance. Feature engineering preprocessing and DNN employed to enhance models' output. The remaining paper is organized as follows: Section II has some literature about how solar energy generating systems work and the performance metrics that the proposed model uses. Section III demonstrates the proposed model. In Section IV, the dataset description, dataset visualization, and experimental findings are presented. The conclusion is presented in Section V.

II. Literature Review

A. Solar Energy Generation

A solar power generation system uses solar cells (PV panels) to convert sunlight into electricity. PV panels consist of semiconductor materials like silicon, which absorb photons from sunlight and generate direct current (DC) electricity through the photovoltaic effect. This electricity is then converted into alternating current (AC) using inverters for use in homes, businesses, or the grid [26]. Solar power generating systems provide clean, renewable energy, contributing to sustainability goals by reducing greenhouse gas emissions and dependence on fossil fuels. Fig. 1 illustrates the flow of solar power systems that generate electricity.

B. Performance Metrics

The objective of these metrics is to measure how well the model can predict unknown values based on known input. The most commonly used performance metrics for regression methods are Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination R^2 [27] [28]. Hence, in this study, the performance of the suggested model strategy is examined using the following indices:

$$MAE = 1/N \sum_{j=1}^{N} |y_j - \hat{y_j}| \tag{1}$$

$$MSE = 1/N \sum_{j=1}^{N} (|y_j - \hat{y_j})^2 \tag{2}$$

$$RMSE = \sqrt{1/N \sum_{j=1}^{N} (|y_j - \hat{y_j})^2} \tag{3}$$

$$R^2 = 1 - \frac{\sum(y_j - \hat{y_j})}{\sum(y_j - \bar{y_j})} \tag{4}$$

where, R^2 represents the proportion of the variance in the dependent variable that is predictable from the independent variables in the model, N is the total number of samples, y_j and $\hat{y_j}$ are state of charge of truth values and the predicted values from the regression model, respectively, and $\bar{y_j}$ is the mean value of y_j. To compare the extracted results, TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is used. It is a multi-criteria decision-making method that ranks alternatives by their distance from the ideal solution (best performance) and negative-ideal solution (worst performance) [29]. Since this study applied over one dataset, formal statistical tests (e.g., Friedman test with post-hoc analysis) would not have a significant output. The average rank method is a non-parametric, simplified approach often used in machine learning and statistical comparisons to compare algorithms across multiple metrics or datasets. Algorithm 1 has the detailed steps of TOPSIS.

Algorithm 1: TOPSIS

1) Construct the Decision Matrix
2) Normalize the Decision Matrix - Convert all metrics to a common scale using vector normalization:

   $r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}}$

   x_{ij}: Original value of metric j for model i. r_{ij}: Normalized value.
3) Weight the Normalized Matrix
4) Identify Ideal (A^+) and Negative-Ideal Solutions (A^-)
   - Ideal Solution: Best value for each criterion (min for errors, max for R^2).
   - Negative-Ideal Solution: Worst value for each criterion (max for errors, min for R^2).

5) Calculate Distances Compute the Euclidean distance of each model from A^+ and A^-:

   $$D_i^+ = \sqrt{\sum_{j=1}^{m} (r_{ij} - A_j^+)^2},$$

   $$D_i^- = \sqrt{\sum_{j=1}^{m} (r_{ij} - A_j^-)^2}$$
6) Computes Closeness Coefficient
   $C_i = (D_i^-)/(D_i^+ + D_i^-)$
   Higher C_i = better overall performance.
7) Rank Models
   - Sort models by C_i in descending order.

III. Methodology

This section provides the proposed model methodology, including feature engineering, and proposed model architecture.

TABLE I: Literature on solar PV forcast estimation models

Reference	Year	Regression Method implemented	Features used	Metric	Optimization method
[17]	2018	Gradient Boost Decision Tree	irradiance, ambient temperature	RMSE, MAPE	- -
[18]	2019	Linear regression, ANN, AdaBoost, SVM ,RFR K-NN	radiation, vapor pressure, surface temperature, atmospheric pressure	MSE , R^2, RMSE	- -
[19]	2021	ANN	air (ambient) temperature, cell temperature, wind speed, humidity, cloud opacity, snow depth	MAE, MSE , R2, RMSE	- -
[20]	2021	Decision tree regression	irradiance, grid availability, equipment availability	MAPE, MAE, MSE	- -
[21]	2022	Multiple Linear Regression, RFR, XgbR	module temperature, irradiance, ambient temperature, wind speed	MAE, R^2,RMSE	- -
[22]	2022	Random forest	irradiance, humidity, temperature, pressure	MAE, MSE , R^2 , RMSE,MAPE	RF optimizer
[23]	2022	Support Vector Machine (SVM)	temperature , humidity , irradiance	MAE, RMSE	Bayesian Optimization (BO)
[24]	2022	Regression Trees, Random Forests, Bagged Trees Regression, Support Vector Regression, Multi-Layer Perceptron	irradiance, sunshine duration, temperature, pressure, humidity, precipitations, and wind speed	MAE, R^2, RMSE	BO
[25]	2023	Gradient Boosting Machine (GBM), RFR	solar insolation (irradiance), ambient temperature, humidity, wind speed	MAE, R^2, RMSE	- -

Fig. 1: Solar Power Generation System

A. Feature Engineering

In this study, two engineered features were incorporated **a.** *Squared Irradiation (IRRADIATION_SQ)* that captures non-linear relationships between solar irradiance and power output. It is a physically feature that improves model flexibility for low/high irradiance regimes without violating monotonicity [30]. It is calculated by Eq. 5.

$$IRRADIATION_{SQ} = IRRADIATION^2 \quad (5)$$

b. *Temperature Differential (TEMP_DIFF)* which quantifies the thermal gradient between module and ambient temperatures. Additionally, it quantifying the thermal gradient affecting PV panel efficiency [31]. It is given by Eq. 6.

$$TEMP\_DIFF = MODULE\_TEMPERATURE - AMBIENT\_TEMPERATURE \quad (6)$$

B. Proposed Model

Forecasting of solar power means analysing data to predict solar power generation over various time spans to minimize the impact of solar fluctuation. Regression algorithms depend on a set of HPs whose values need to be optimized to get the optimal model prediction form. Therefore, this research is divided into three phases: phase 1: regression HPs optimization based on optimization algorithms, phase 2: data processing via feature engineering, and phase 3: forecasting solar power via TEDNN. The proposed hybrid forecasting

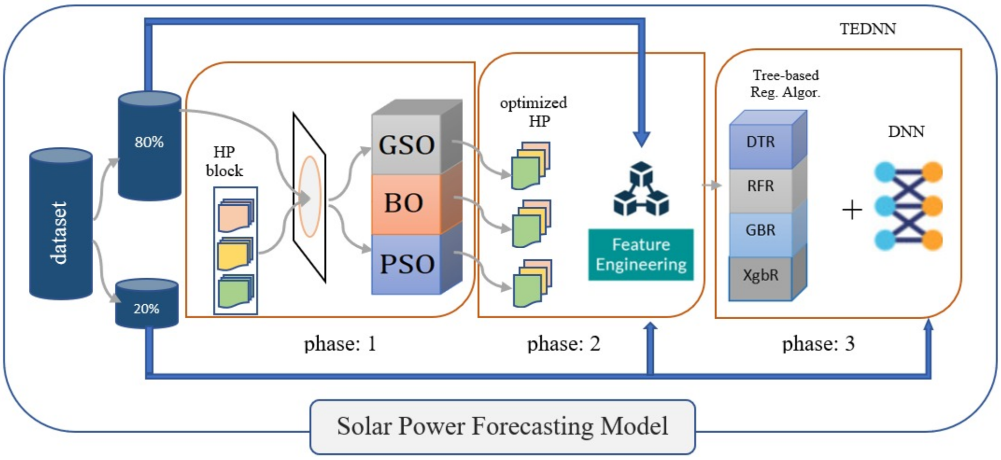


Fig. 2: Proposed Model for Solar Power Forecasting

model aims to enhance solar power (AC power) prediction accuracy by integrating tree-based regression algorithms with a deep neural network (DNN). This architecture operates through two parallel pathways. In the tree-based regression pathway, each model independently generates predictions and produces leaf embeddings (terminal node indices) that encode decision paths, serving as structured feature representations. Concurrently, in the deep learning pathway, a neural network processes both the original engineered features and the tree embeddings in separate sub-networks. These two pathways are then merged via concatenation, enabling the DNN to learn complex interactions between the structured decision rules from the tree models and the raw feature representations. Table II summarizes the architecture of TEDNN.

Fig. 2 depicts a general overview of the steps forming the development process. As demonstrated, it starts by feeding the data to the model, where it is divided into 80% and 20%, which represent training and test sets, respectively. The training set is utilized through the three phases. For the optimization phase, three optimization methods were used to tune the HPs of regression algorithms including GSO, BO and PSO. The HPs for regression methods and their ranges are shown in Table III. In addition, the objective function of the optimization methods is based on minimizing MAE. Then, the raw data undergoes systematic feature engineering, which includes outlier handling and derived features generation. Finally, for forecasting, the optimized HPs from phase 1 and engineered features from Phase 2 are utilized in two parallel workflows. First, tree-based regression algorithms including DTR, RFR, GBR and XgbR [32], using optimized HPs from Phase 1 for interpretable predictions. Second, tree-based regression algorithms + DNN to capture non-linear patterns.

IV. Simulation and Results

The proposed learning model is tested using the PV solar power dataset. The structure of the used dataset is described in subsection 4.1. The dataset analysis and visualization are in subsection 4.2. Finally, the outcomes of the proposed model are displayed according to performance metrics in subsection 4.3. A study of the influence of parameters on output was also presented.

A. Dataset

A solar power system uses photovoltaic (PV) panels to transform sunlight into electrical energy. The PV solar power dataset that was used in this study came from two separate solar plants in India [33]. Each plant generates two files of data: one for weather data and the other for solar power generation. The weather data file includes ambient temperature, module temperature, and irradiation, while the generation data file includes AC power, DC power, daily yield, and total yield. Over the course of 34 days, data was gathered at a 15-minute interval. Results from the experiment in this study were based on information from plant-2. The plant-2 data file shows there are 22 inverters (S) with only one weather monitoring unit. The DC power generation for all inverters is shown on Fig. 3. Two data files are merged to train and test the proposed model. The experimental results are based on three features which are irradiation, ambient temperature, and module temperature, to predict AC power.

B. Data analysis and visualization

This section provides a graphical representation of the dataset to extract information and identify the dataset patterns. Fig. 3 illustrates the DC power generation from all inverters over the course of the day. The PV cells generate DC power, which is transmitted to a total of 22 inverters. The highest three DC power values were achieved by inverters S5, S8, and S21, with respective values of 27,709.2, 27,430.7, and 27,240.1. Conversely, the figure also indicates the three lowest DC power values obtained by inverters S6, S4, and S12, with values of 18,859.7, 18,071.2, and 16,640.6, respectively. The lower DC power outputs from these inverters could be attributed to various factors, such as potential faults in the solar cells connected to the inverters or the presence of shading obstructing the solar cells' exposure to sunlight. Table IV presents the inverters' source key and their alternative symbols.

Fig. 4 displays the time-dependent profiles of the AC and DC power generated over a span of 34 days. As depicted in the figure, the power generation remains constant at zero from 18:30 pm to 6:00 am, representing a duration of approximately 12.30 hours every day. Additionally, the AC power exhibits variability within the range of [0, 1358.4] Watt, while the DC power fluctuates within the range of

TABLE II: Tree-Enhanced Deep Neural Network Architecture

Layer	Input	Output Size	Activation	Regularization
Tree Embedding Input	Leaf indices (n_estimators)	–	–	
Dense 1	Leaf embeddings	64	ReLU	L2 (0.01)
Dense 2	Output from Dense 1	32	ReLU	L2 (0.01)
Raw Feature Input	Engineered features (scaled)	–	–	
Dense 3	Raw features	32	ReLU	
Concatenation	Outputs from Dense 2 & 3	64	–	
Final Dense Layer	Concatenated features	1	Linear	

TABLE III: HPs ranges for regression algorithms

Regression Algorithm	HP	Values
DTR	max_depth	[2, 10]
	min_impurity_decrease	[0.0 , 0.5]
	Criterion	{squared_error, friedman_mse, absolute_error}
	min_samples_leaf	[1:70]
RFR	n_estimators	[10, 1000] (log)
	max_depth	[2, 10]
	max_features	{0.25, 1.0, sqrt, log2}
	min_weight_fraction_leaf	[0.0, 0.001 ,0.01, 0.1]
GBR	n_estimators	[10, 1000] (log)
	max_depth	[2, 10]
	max_features	{0.25, 1.0, sqrt, log2}
	learning_rate	[0.01, 0.3]
	subsample	[0.1, 1]
XgbR	n_estimators	[10, 1000] (log)
	max_depth	[2, 10]
	colsample_bylevel	{0.25, 1.0, sqrt, log2}
	learning_rate	[0.01, 0.3]
	reg_lambda	{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4}
	gamma	{0, 0.1, 0.2, 0.3, 0.4, 1.0, 1.5, 2.0}

TABLE IV: Inverters' source key and their alternative symbols

Source Key	Symbol
4UPUqMRk7TRMgml	S1
81aHJ1q11NBPMrL	S2
9kRcWv60rDACzjR	S3
Et9kgGMDl729KT4	S4
IQ2d7wF4YD8zU1Q	S5
LYwnQax7tkwH5Cb	S6
LlT2YUhhzqhg5Sw	S7
Mx2yZCDsyf6DPfv	S8
NgDl19wMapZy17u	S9
PeE6FRyGXUgsRhN	S10
Qf4GUc1pJu5T6c6	S11
Quc1TzYxW2pYoWX	S12
V94E5Ben1TlhnDV	S13
WcxssY2VbP4hApt	S14
mqwcsP2rE7J0TFp	S15
oZ35aAeoifZaQzV	S16
oZZkBaNadn6DNKz	S17
q49J1IKaHRwDQnt	S18
rrq4fwE8jgrTyWY	S19
vOuJvMaM2sgwLmb	S20
xMbIugepa2P7lBB	S21
xoJJ8DcxJEcupym	S22

[0, 1420.93] Watt per day, and the maximum values of total power generated with time of the day for AC and DC are 474924.4, 486466.9, respectively. Notably, the maximum power output is attained around 13:00 pm.

C. Results and discussion

This section holds the optimization and model evaluation results. In addition, parametric analysis was performed to show the influence of different parameters on the output model.

1) Optimization results: Table V presents the optimized HPs for each regression algorithm across three optimization methods, with varying values due to MSE minimization. These HPs are tested to evaluate performance. Results show differing metrics for each algorithm. The initialization parameters for PSO dictate the algorithm's operation in searching for optimal solutions. Table VI presents these parameters, including a population size (N) of 10 particles and 3 dimensions (D) for optimization. The maximum weight (Wmax) is set at 0.9 and the minimum weight (Wmin) at 0.5, balancing exploration and exploitation. The acceleration coefficients influence particle behavior; the first coefficient is zero, meaning particles focus solely on their best positions, while the second is 0.3, allowing them to consider neighbors' best positions as well.

Table VII displays the performance metrics for DTR, RFR, GBR and XgbR based on the optimal HP results achieved by GSO, BO and PSO. As demonstrated in the table, for DTR, the maximum value of MAE is achieved by BO, which equals 1503.6, followed by 1347.7 for GSO, while PSO achieves the minimum value, which equals 1213.3. Additionally, for RFR, PSO attained the least MAE value, which equals 1074.01, while GSO obtained the highest value, which is equal to 1507.6. Furthermore, for GBR, the PSO method yields the lowest MAE of 917.7 and highest R^2 of 0.901, indicating optimal predictive accuracy compared to BO and GSO.

2) Model Performance: For XgbR, PSO method demonstrates the optimal performance with a MAE of 1174.9 and an R^2 of 0.913. In overall, PSO emerges as the most effective optimization method based on the analysed metrics.

Table VIII holds the results of applying TOPSIS method to rank regression models and optimized HPs results achieved by GSO, BO and BSO optimization methods based on their performance metrics. The results are sorted by the closeness coefficient, where higher values indicate better overall performance. This ranking depends on equal weights

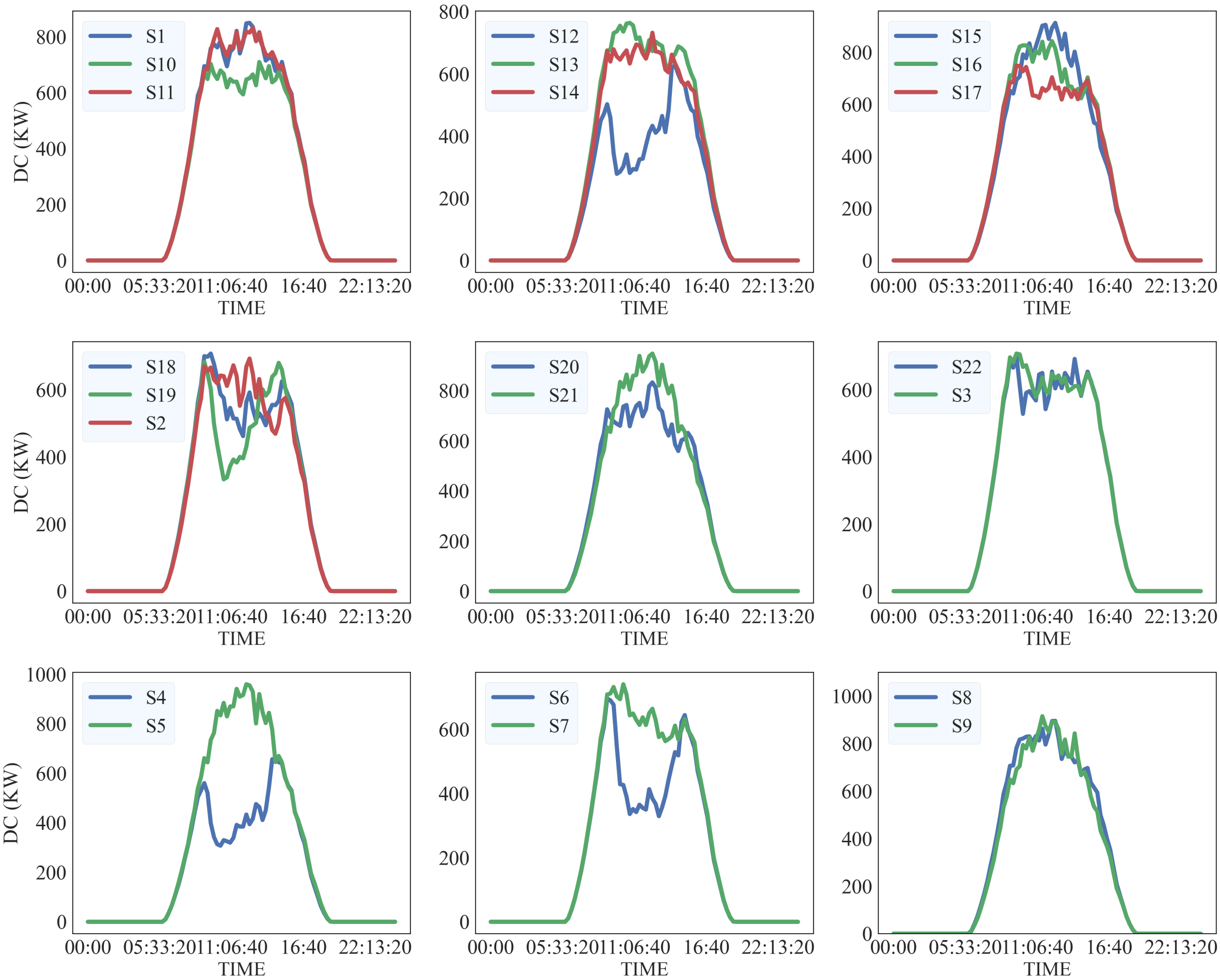


Fig. 3: DC power generation in KW for all inverters

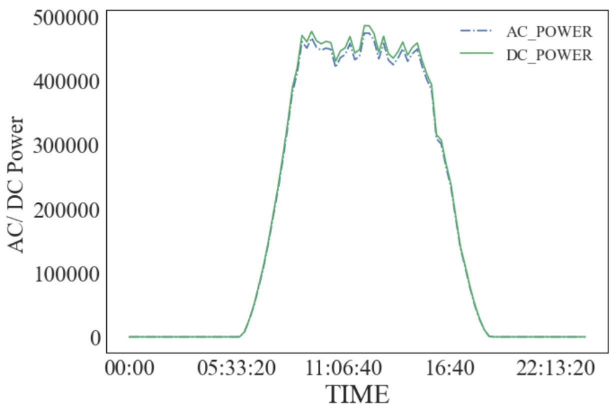


Fig. 4: Total power generated with time of the day

assumption. RFR_PSO and DTR_PSO top the list despite higher errors, primarily due to their trade-off across all metrics with relatively good R^2. Table IX holds a comparative evaluation of regression models DTR, RFR, GBR, and XgbR with and without feature engineering across metrics MAE, MSE, RMSE, and R^2. The GBR model combined with deep learning (GBR+DNN) shows better predictive power with feature engineering, reaching the greatest R^2 score (0.941), lowest MAE (708.6), and least RMSE (11560.9), which highlights its efficiency in reducing error and clarifying variance. However, the standalone GBR model also shows great predictive performance (MAE: 798.2, R^2: 0.931), implying that the algorithm is naturally strong. Interestingly, deep neural network inclusion invariably enhances base models' efficiency: XgbR +DNN results in R^2 upgrade to 0.933 from 0.915. These results show that DNN assists existing models. Also, through feature engineering, the performance of the models is further improving. After feature engineering, GBR+DNN has the best metrics (MAE of 708.6, RMSE of 11560.9, R^2 of 0.941). Table X holds the results of applying TOPSIS method to rank regression models performance with and without feature engineering based on their performance metrics. The results are sorted by the closeness coefficient, where higher values indicate better overall performance. Best model is GBR+DNN (With Feature Engineering), achieves the highest closeness coefficient (0.781) due to strong performance across all metrics (MAE = 708.6, R^2 = 0.941). Feature engineering improves performance, as models with feature engineering dominate the top 3 ranks. Example: GBR+DNN (With) outperforms GBR+DNN (Without) by 16 in closeness. Models with DNN (e.g., XgbR+DNN) show significantly better results than their non-DNN counterparts. Gradient boosting methods (GBR, XGBR) occupy the top ranks, while decision trees (DTR) and random forests (RFR) trail behind.

Fig. 5 illustrates the comparison between actual AC and

TABLE V: Optimized HP values

Regression Algorithm	HP	GS	BO	PSO
	max_depth	3	2	5
DTR	min_impurity_decrease	0.1	0.4	0.48
	Criterion	squared_error	squared_error	absolute_error
	min_samples_leaf	10	10	10
	n_estimators	100	123	134
RFR	max_depth	3	5	5
	max_features	sqrt	Log2	sqrt
	min_weight_fraction_leaf	0.0	0.01	0.001
	n_estimators	100	778	550
	max_depth	3	3	5
GBR	max_features	sqrt	sqrt	sqrt
	learning_rate	0.1	0.25	0.2
	subsample	0.2	0.4	0.4
	n_estimators	150	100	50
	max_depth	4	3	3
XgbR	colsample_bylevel	0.1	log2	log2
	learning_rate	0.015	0.1	0.2
	reg_lambda	2	3	3
	gamma	0	0	0

TABLE VI: PSO parameters values

Parameters	Values
Population Size N	10
No. of Dimensions (D)	3
No. of Iterations (T)	100
Maximum Weight (W_{max})	0.9
Minimum Weight (W_{min})	0.5
Acceleration Coefficient (c_1, c_2)	0.0, 0.3

TABLE VII: Performance metrics results for regression algorithms based on the optimal HPs results achieved by GSO, BO and PSO optimization methods

Optimization method	MAE	MSE	RMSE	R^2
		DTR		
GSO	1347.7	4.032047e+06	2007.99	0.882
BO	1503.6	4.790283e+06	2188.6	0.883
PSO	**1213.3**	4.855904e+06	2203.6	**0.898**
		RFR		
GSO	1507.6	4.597589e+06	2144.19	0.883
BO	1105.2	4.477948e+06	2116.11	0.886
PSO	**1074.01**	4.795537e+06	2189.87	0.88
		GBR		
GSO	917.7	3.976826e+06	1994.198	0.899
BO	836.93	4.142630e+06	2035.34	0.895
PSO	**823.92**	3.884271e+06	1970.8	**0.901**
		XgbR		
GSO	1213.3	4.032047e+06	2007.99	0.898
BO	1190.8	3.886385e+06	1971.39	0.901
PSO	**1174.9**	3.592529e+06	1895.3	**0.913**

TABLE VIII: TOPSIS method to rank regression models performance based on the optimal HPs results achieved by GSO, BO and PSO optimization methods

Rank	Opt. _Model	Closeness Coefficient
1	RFR_PSO	0.9497
2	DTR_PSO	0.9463
3	DTR_BO	0.8693
4	DTR_GSO	0.6284
4	XGBR_PSO	0.6284
6	RFR_BO	0.6040
7	RFR_GSO	0.5834
8	XGBR_BO	0.5632
9	XGBR_GSO	0.5128
10	GBR_PSO	0.2426
11	GBR_BO	0.2231
12	GBR_GSO	0.1653

predicted AC values for four engineered regression models + DNN. Within the figure, a single band is displayed, representing the 95% confidence range. This band serves as an indicator of the accuracy of the estimation of the models. Additionally, a prediction range is represented by another band, with its width corresponding to the 95% prediction interval. This interval provides a comprehensive understanding of the predictive capability of the models. The figure compares the performance accuracy of four learning models of TEDNN, based on the values of HPs selected by PSO. The red dashed diagonal line that started at the origin (0,0) and has a 45-degree angle represents the line of equality, which means, the ideal scenario where the actual values are exactly equal to the predicted values. Additionally, the regression equation for each model is represented by (y = slope x + intercept) as illustrated in table XII. As shown in table XII, all four models+DNN demonstrated strong predictive accuracy, with correlation coefficients (r) exceeding 0.966, confirming a robust linear relationship between predicted and actual values. The GBR outperformed others, achieving the highest correlation (r= 0.9682) and a near-ideal slope (0.9044), which indicates both precise predictions and minimal systematic bias. While RFR exhibited the steepest slope (0.9276) and lowest intercept (327.10), its marginally lower correlation (r =0.9662) suggests slightly less consistent predictions compared to GBR. XgbR and DTR delivered intermediate performance; XgbR's elevated intercept (395.25) implies systematic overestimation for lower power values, whereas DTR's balanced slope (0.9223) and correlation (r= 0.9665) underscore its reliability despite its simpler structure.

3) Parametric Study: Parametric analysis study the influence of different parameters on the output model. To study the impact of the parameter on the model, a set of data was generated, where the values of each parameter ranged between the minimum and maximum of its values on the original dataset. In addition, the values of the remaining parameters were determined by their mean value. Minimum, maximum, and average values for the three parameters are demonstrated in Table XI. Subsequently, the generated data is passed through each model (engineered regression algorithm + DNN) to get the prediction output. Figure 6 shows the parametric study of each parameter with the prediction output

TABLE IX: Regression Models Performance with and without Feature Engineering

	Model	MAE	MSE	RMSE	R^2
Without feature engineering	DTR	1347.7	4.855904e+06	2203.6	0.882
	DTR+ DNN	1482.4	4.387759e+06	2094.6	0.893
	RFR	1507.6	4.795537e+06	2189.9	0.88
	RFR + DNN	936.2	3.871631e+06	1967.64	0.902
	GBR	917.7	3.884271e+06	1970.8	0.901
	GBR + DNN	932.8	3.840170e+06	1959.6	0.902
	XGBR	1190.8	3.592529e+06	1895.3	0.913
	XgbR + DNN	844.1	2.842114e+06	1685.8	0.931
With feature engineering	DTR	836.9	4.142630e+06	2035.3	0.895
	DTR+ DNN	942.1	3.613706e+06	1900.9	0.912
	RFR	1190.8	3.592529e+06	1895.3	0.913
	RFR + DNN	823.9	3.976827e+06	1994.1	0.899
	GBR	798.2	2.843335e+06	1686.2	0.931
	GBR +DNN	**708.6**	2.436520e+06	11560.9	**0.941**
	XgbR	1155.6	3.512624e+06	1874.2	0.915
	XgbR +DNN	807.7	2.820365e+06	1679.3	0.933

TABLE X: TOPSIS method to rank regression Models Performance with and without Feature Engineering based on their performance metrics.

Rank	Model	Closeness Coefficient
1	GBR+DNN (With)	0.781
2	XgbR+DNN (With)	0.733
3	GBR (With)	0.697
4	XGBR+DNN (Without)	0.662
5	GBR+DNN (Without)	0.621
6	XGBR (Without)	0.589
7	RFR+DNN (With)	0.554
8	DTR (With)	0.522
9	RFR+DNN (Without)	0.498
10	DTR+DNN (With)	0.475
11	XgbR (With)	0.452
12	RFR (With)	0.433
13	DTR+DNN (Without)	0.417
14	GBR (Without)	0.398
15	RFR (Without)	0.382
16	DTR (Without)	0.361

for every model. As demonstrated in the Figure 6 (a) , for all models, the predicted AC power increases with increasing irradiation values, this indicates that a rise in solar radiation causes an increase in output power, which improves a solar panel's efficiency [34].

For temperature, Figure 6(b) shows that the output power of PV is not affected much by an increase in the ambient temperature. Additionaly, Figure 6 : (c) show that as module temperature increases, the output power increases slightly with DTR and RFR, while the output power of GBR + DNN and XgbR + DNN decreases with increasing module temperature. Consequently, increasing module temperature results in a decrease in the output power of the PV system, while the output of PV modules improves significantly with increasing irradiance levels. This indicates that GBR and XgbR demonstrate the correct relationship between PV output power, temperature, and irradiance.

TABLE XI: Feature Measure

Feature	Max	Min	Avg
Irradiation	1.099	0	0.23
Ambient temp.	39.18	20.94	28.07
Module Temp.	66.64	20.27	32.77

4) Comparative study: Table XIII presents the accuracy of the proposed model compared with a related one. The authors in [35] employed time series models such as ARIMA

TABLE XII: Comparative Performance Metrics of Engineered Regression Methods + DNN

Engineered model	Slope	Intercept	Correlation coefficient (r)
DTR+DNN	0.9223	338.6497	0.9665
RFR+DNN	0.9276	327.1008	0.9662
XgbR+DNN	0.9012	395.2470	0.9669
GBR+DNN	0.9044	349.0124	0.9682

TABLE XIII: Comparison of the proposed model with related

Reference	Year	Technique	Accuracy
[35]	2021	Prophet Model	89.4%
[36]	2023	Random Forest	84.3%
[37]	2023	Polynomial Regression	93.7%
Proposed		GBR based PSO	94.1%

and Prophet to obtain meaningful results for forecasting solar power. The objectives of the study include understanding and managing the output variability of solar power generation using machine learning algorithms for power generation that achieved 89.4% accuracy based on the Prophet model. In addition, authors in [36] utilized machine learning techniques for solar forecasting and the highest accuracy was achieved by RF with an accuracy of 84.3%. Moreover, authors in [37] used polynomial regression based PSO achieving accuracy of 93.7%. While the proposed model TEDNN (GBR) based on feature engineering and PSO, achieved an optimal accuracy of 94.1%.

V. Conclusion

The forecasting of solar system output power is crucial for assessing system performance and meeting market demands while avoiding instability. This research utilized TEDNN architecture to predict solar power generation. Three optimization methods GS, BO, and PSO were used to optimize HPs of regression techniques. The findings revealed that PSO outperformed other used optimization methods in terms of optimizing HPs values. Additionally, four regression algorithms including DTR, RFR, GBR, and XgbR combined with DNN were used to forecast AC power generation. GBR+DNN and XgbR+DNN achieved optimal results with R^2 values of 94.1% and 93.3% respectively. Furthermore, feature engineering improves performance, and models combined with DNN show significantly better results than their non-DNN substitutes.

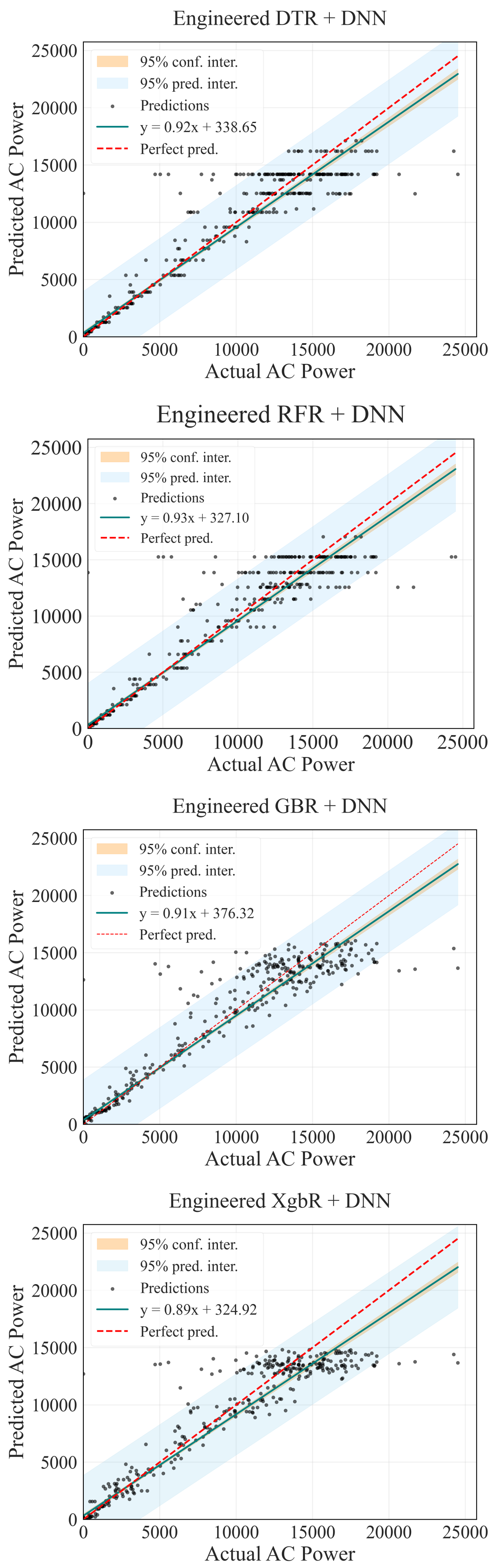
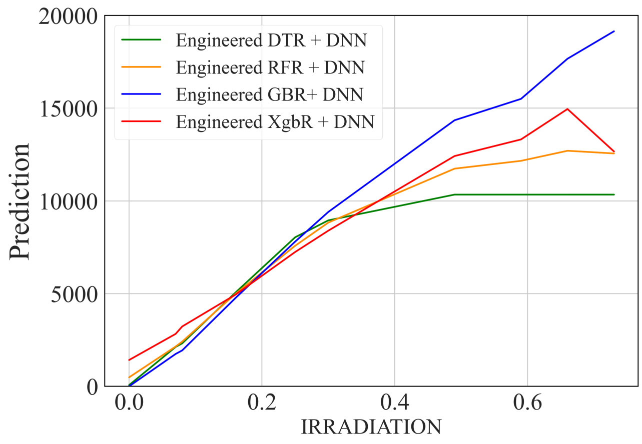
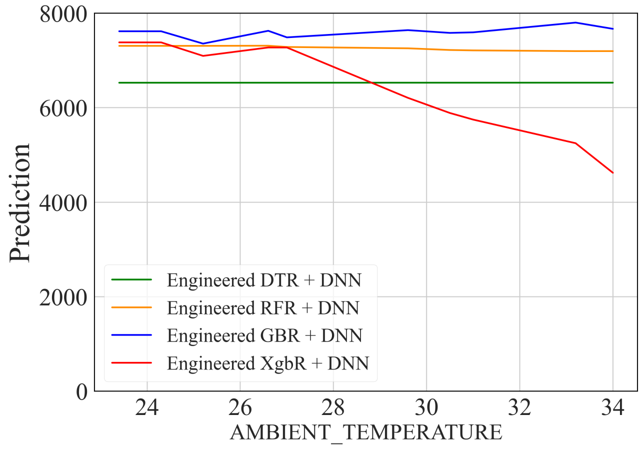


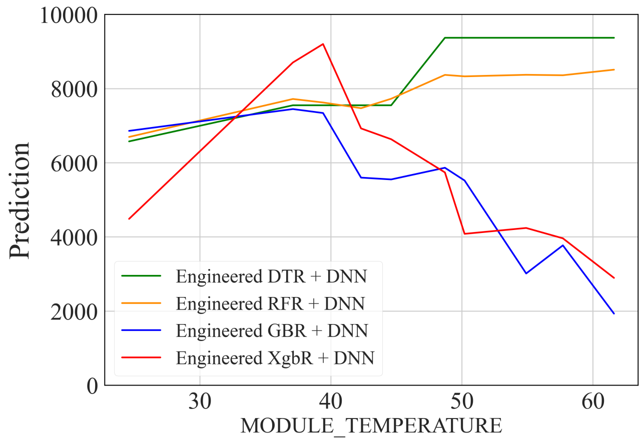
Fig. 5: Relationships between predicted and actual values for engineered regression models + DNN based HPs of PSOs



(a)



(b)



(c)

Fig. 6: Parametric Analysis of different parameters in the model (Engineered regression algorithms + DNN)

References

[1] N. S. L. Holechek, H. Geli and R. Valdez, “A global assessment: Can renewable energy replace fossil fuels by 2050?” *Sustainability*, vol. 14, no. 8, 2022.

[2] M. K. P. K.Hasan, S.Yousuf and S.Islam, “Effects of different environmental and operational factors on the pv performance: A comprehensive review,” *Energy Science & Engineering*, pp. 656–675, 2021.

[3] R. E. R. U. F. J. M.-d.-P. J. Antonanzas, N. Osorio and F. Antonanzas-Torres, “Review of photovoltaic power forecasting,” *Solar Energy*, vol. 136, pp. 78–111, 2016.

[4] M. Ryan and S. N. El-Semary, “A. energy-efficient aodv using super-increasing knapsack,” *Journal of Al-Azhar University Engineering Sector*, vol. 19, no. 72, pp. 102–120, 2024.

[5] Y. J. H. et. al, “The design and implementation of a solar tracking generating power system,” *Engineering Letters*, vol. 17, no. 4, pp. 252–256, 2009.

[6] I. H. Sarker, “Machine learning: Algorithms, real-world applications and research directions,” *SN COMPUT*, vol. 2, no. 3, 2021.

[7] M. Feurer and F. Hutter, “Automated machine learning: methods,systems, challenges,” *Springer Nat.*, pp. 1–219, 2019.

[8] H. C. E. Gültekin and H. Kaynak, Eds., *Decision Tree Regression Method Application For Prediction Of Woven Fabrics Tear Strength*. Turky: 3rd Int. Conf. of Materials and Engineering Technology (TICMET’21), 2021.

[9] B. L. S. Chowdhury, Y. Linb and L. Kerbyd, “Evaluation of tree based regression over multiple linear regression for non-normally distributed data in battery performance,” *Int. Conf. on Intelligent Data Science Technologies and Applications (IDSTA)*, pp. 17–25, 2022.

[10] M. S. Z. A. Ismail, M. Elpeltagy and K. Eldahshan, “A new deep learning-based methodology for video deepfake detection using xgboost,” *Sensors*, 2021.

[11] M. A. M. F. kamel and M. M. Al-Gazzar, “Energy management in smart inter-connected micro-grids using a hybrid sperm swarm optimization and gravitational search algorithm,” *Journal of Al-Azhar University Engineering Sector*, vol. 18, no. 66, pp. 98–115, 2023.

[12] H. Z. L. D. X. H. L. J. Wu, X. Y. Chen and S. Deng, “Hyperparameter optimization for machine learning models based on bayesian optimization,” *Journal of Electronic science and technology*, vol. 17, no. 1, 2019.

[13] I. Yang and A. Shami, “On hyperparameter optimization of machine learning algorithms: Theory and practice,” *Neurocomputing*, vol. 415, pp. 295–316, 2020.

[14] A. S. Desuky and H. M. Harb, “Feature selection on classification of medical datasets based on particle swarm optimization,” *International Journal of Computer Applications*, vol. 104, no. 5, pp. 14–17, 2014.

[15] X. Jiang and C. Xu, “Deep learning and machine learning with grid search to predict later occurrence of breast cancer metastasis using clinical data,” *journal of clinical medicine*, vol. 11, no. 19, 2022.

[16] P. M. et. al, “Direct search optimization application programming interface with remote access,” *IAENG International Journal of Applied Mathematics*, vol. 40, no. 4, pp. 251–261, 2010.

[17] T. B. Blume and D. Schramm, “Hyperparameter optimization techniques for designing software sensors based on artificial neural networks,” *Sensors*, vol. 21, no. 24, 2021.

[18] R. R. Y. C. J. Wang, P. Li and Y. Zhou, “A short-term photovoltaic power prediction model based on the gradient boost decision tree,” *applied sciences*, vol. 8, no. 5, 2018.

[19] J. J. S. Kim and M. Sim, “A two-step approach to solar power generation prediction based on weather data using machine learning,” *sustainability*, vol. 11, no. 5, 2019.

[20] M. AlShafeey and C. Csáki, “Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods,” *Energy Reports*, vol. 7, pp. 7601–7614, 2021.

[21] A. B. Rahul, A. Gupta and K. Roy, “Solar energy prediction using decision tree regressor,” *5th Int. Conf. on Intelligent Computing and Control Systems (ICICCS)*.

[22] O. E. A. G. I. Abdelmoula, S. Elhamaoui and M. E. Aroussi, “A photovoltaic power prediction approach enhanced by fea- ture engineering and stacked machine learning mode,” *Energy Reports*, vol. 8, pp. 1288– 1300, 2022.

[23] J. A.-A. H. J.-H. C. G. Villegas-Mier, J. Rodriguez-Resendiz and A. Odry, “Optimized random forest for solar radiation prediction using sunshine hours,” *micromachines*, vol. 13, no. 9, 2022.

[24] M. Vijay and M. Saravanan, “Optimized solar irradiance forecasting using bayesian optimization based machine learning algorithm to determine the optimal size of a residential pv system,” *Int. Conf. on Sustainable Computing and Data Communication Systems (ICSCDS)*.

[25] L. T.-M. B. M. Chaibia, M. Benghoulama and A. El-Hmaidid, “Machine learning models based on random forest feature selection and bayesian optimization for predicting daily global solar radiation,” *Journal of Renewable Energy Development (IJRED)*, vol. 11, no. 1, pp. 309–323, 2022.

[26] M. I. E. M. M. H. C. Mohamed, M. Reda and H. Seoudy, “A comparative study of module efficiency and sizing of the inverter’s lcl filter for photovoltaic systems a case study,” *Journal of Al-Azhar University Engineering Sector*, vol. 19, no. 72, pp. 183–201, 2024.

[27] M. W. D. Chicco and G. Jurman, “The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation,” *PeerJ Computer Science*, vol. 7, pp. 1–24, 2021.

[28] M. A. A. et. al, “Performance evaluation and comparative analysis of different machine learning algorithms in predicting cardiovascular disease,” *Engineering Letters*, vol. 29, no. 2, pp. 731–741, 2021.

[29] N. V. P. Zlatko, “Notes on topsis method,” *International Journal of Research in Engineering and Science*, vol. 1, pp. 2–15, 2013.

[30] J. R. G. M. G. Villalva and E. R. Filho, “Comprehensive approach to modeling and simulation of photovoltaic arrays,” *IEEE Transactions on Power Electronics*, vol. 24, no. 5, 2009.

[31] E. Skoplaki and J. A. Palyvos, “Operating temperature of photovoltaic modules: A survey of pertinent correlations,” *Renewable Energy*, vol. 34, no. 1, pp. 23–29, 2009.

[32] A. C. C. Bentejac and G. M. Munoz, “A comparative analysis of gradient boosting algorithms,” *Artificial Intelligence Review*, vol. 54, p. 1937–1967, 2021.

[33] [Online], “https://www.kaggle.com/code/pertetotale/solarpowergeneratin,” *last access March*, 2025.

[34] J. V. N. Vázquez, “Photovoltaic system conversion,” *Power Electronics Handbook (Fourth Edition)*, pp. 767–781, 2018.

[35] M. S. K. Kotasthane, M. Kasture and Vashney, “A novel approach to develop energy prediction model for large industries using artificial intelligence,” *International Research Journal of Modernization in Engineering Technology and Science*, vol. 3, no. 6, 2021.

[36] V. D. Gil-Vera and C. Q.-L. Pez, “Predictive modelling of photovoltaic solar power generation,” *Pwseas Transactions on Power Systems*, vol. 18, 2023.

[37] M. A. S. A. El-aal and A. Naiem, “Forecasting solar photovoltaic energy production using linear regression-based techniques,” *JATIT*, vol. 101, no. 9, 2023.

Shereen A. El-aal is a Lecturer at Faculty of Science, Al-Azhar University, Cairo, Egypt. She received the B.Sc. degree in computer science from Al-Azhar University in 2007. In 2008, she joined Information Technology Institute (ITI), Cairo, Egypt, and in 2009, she joined Wipro Company as a developer. She received her M.Sc and Ph.D from Al-Azhar University in 2016 and 2021, respectively. Her research interests: Artificial Intelligence, Machine Learning, Deep Learning, Computational intelligence, and Internet of Things. She has published several research papers related to fields of expertise.

Yomna M. Elbarawy received her B.Sc. and M.S. degrees in computer science from the Faculty of Sciences (girls branch) at Al-Azhar University in Cairo, Egypt, in 2008 and 2014, respectively. Received her computer science Ph.D. in 2020 at the same university. Her research areas are social network analysis, computational intelligence, machine learning, and deep learning technologies.

Asmaa Mohamed received the B.Sc. degree in science, in 2011, the M.Sc. degree in computer science, in 2017, and the Ph.D. degree in computer science, in 2023. She is currently a Lecturer in computer science at the Mathematics Department, Faculty of Science, Al-Azhar University, Cairo, Egypt. She has published a number of research papers in the field of AI, machine learning, IOT, and data mining.

Sahar Abd Elrouf Gomaa is a lecture of Computer Science at Al-Azhar University(girls) in Cairo, Egypt. She was awarded a B.Sc. degree in Computer Science and pure mathematics by Faculty of Science, Menoufia University in 1995. She held an M.S. in 2001 in Computer Science from Computer and Mathematics Department Faculty of Science, Menoufia University, Egypt, and a Ph.D. in 2008 in Computer Science from faculty of Science, Al-Azhar University, Cairo, Egypt.

Nahed El Desouky is an associate professor of Computer Science and Information Systems at Al-Azhar University(girls) in Cairo, Egypt. Bachelor of Science from the faculty of engineering, Cairo University in 1982, Master of communication and Electronics from the faculty of engineering Cairo university 1990, communication and Electronics Ph.D. 1999 from the faculty of Engineering at Cairo University. Associate Professor of computer science from 2010. She Published 23 papers in different branches of computer science ; Security, cloud computing and optimization.