THEMIS: An Explainable Machine Learning Framework for Solar Photovoltaic Degradation Prediction and Sustainable Energy Analytics

Shreyash Singh, Abishek KK, Nithya S, Praveenkumar S
Department of Data Science and Business Systems SRM Institute of Science & Technology
Chennai, India

***Abstract*—Solar photovoltaic (PV) systems are a key element in addressing the increasing global need for renewable energy in line with Sustainable Development Goal 7 (affordable and clean energy). While being installed at a fast pace, degradation over time is a key issue affecting energy yield, cost of investment and system performance. The performance of solar panels deteriorates over time due to environmental and operational stresses like thermal cycling, humidity, dust, moisture, and electrical overstress. Conventional monitoring solutions offer real-time dashboard and threshold-based notifications, but lack predictive capabilities to understand long-term performance trends. In this paper, we present THEMIS, an explainable machine learning model to predict degradation and estimate performance efficiency of solar panels, using structured environmental and operating data. We built a dataset of 10,000 instances with features for irradiance, temperature, humidity, dust index, voltage, current, power and efficiency. We applied GridSearchCV for hyper- parameter tuning, and shuffled-target testing for evaluation. Feature group importance was assessed through an ablation study, and SHAP explainability pinpointed the root cause of degradation. Our findings indicate that an optimized XGBoost model achieves better accuracy, stability and performance than baselines. THEMIS offers a transparent and validated approach for predictive maintenance, energy yield optimisation and sustainable management of solar parks.**

***Index Terms*—Sustainable Energy, Solar Energy, PV Degra- dation, Machine Learning, XGBoost, Explainable AI, SHAP, Predictive Maintenance, Sustainable Energy Analytics**

I. INTRODUCTION

Renewable energy is increasingly being harnessed to meet global sustainable energy requirements over the last decade, driven by concerns about climate change, fossil fuel scarcity and greenhouse gas emissions [9]. Solar photovoltaic (PV) systems are among the most promising renewable technolo- gies with potential for mass deployment and cost-effective decentralized and large-scale generation. Solar installations are being increasingly deployed by governments and private sector players as part of Sustainable Development Strategies and Carbon Mitigation Plans [25].

Even with technological improvements and lowering costs for installation, a key engineering target is to ensure reliable performance of solar systems over their long operational lifetimes. Solar modules experience degradation due to expo- sure to environmental conditions, UV radiation, temperature variations, dust accumulation, ingress of humidity and load conditions [3], [10], [22]. While annual degradation rates may be low (in the range 0.5% to 1%), degradation over a 20- 25 year lifetime can significantly impact cumulative energy production, and hence returns on investment [1].

Currently available monitoring systems are mostly reactive [2]. These systems offer real-time monitoring of power output, and alarm systems when power output is below a threshold. But they lack the ability to model the degradation process to predict trends before substantial degradation occurs. As a result, maintenance actions tend to be reactive, inefficient and prone to downtime.

Machine learning has the potential to model the nonlin- ear deterioration process from past data [24]. Specifically, ensemble learning models have the potential to model com- plex interactions between the environmental stressors and degradation [6]. But many machine learning approaches are black-box models, which deliver accurate predictions, but lack explanations for degradation. In critical industries such as renewable energy, interpretability of models and results is essential to gain trust from stakeholders and support decision making [16].

To address these limitations, this research introduces THEMIS - an explainable machine learning system that ex- plains not just the percentage of degradation but also pro- vides feature-based explanations for interpretation. Through predictive modeling, hyperparameter tuning, model robustness, feature ablation and explainability via SHAP [5], THEMIS creates a robust and scientifically rigorous degradation analyt- ics solution.

II. LITERATURE REVIEW

The main areas of focus in solar energy analytics include power prediction, fault detection and statistical degradation analysis [9]. Initial research focused on linear regression and polynomial fitting using irradiance and temperature, to predict power output. These models are easy to compute but rely on linear relationships and do not consider nonlinear relationships between environmental and electrical variables.

Time-series models such as ARIMA and Long Short-Term Memory (LSTM) models have been widely used for short-term solar power prediction [7], [13]. These models can accurately capture temporal variability in fluctuations and daily yield variations. But these models typically focus on short-term forecasts, rather than degradation. Additionally, time-series deep learning models typically require a large amount of time- sequential historical data, which may not be available for long- term degradation analysis.

Random Forest and Gradient Boosting (GB) have been found to be more accurate than other regression models in structured data [6], [4]. Random Forest reduces variance by aggregating predictions from an ensemble of decision trees, while boosting successively reduces errors to capture intricate relationships. XGBoost, a fast and scalable gradient boosting approach, includes regularization techniques to avoid overfit- ting and enhance model robustness [4]. Despite their use in solar forecasting, most studies focus on forecasting accuracy, neglecting the impact of features via ablation studies, or model robustness via validation techniques.

Explainable artificial intelligence (XAI) methods like SHAP have been used to explain model behaviours [5]. SHAP values offer theoretically valid feature attribution with the theory of cooperative games. But there is a lack of integration of in- terpretability in degradation prediction systems [17]. Previous studies have often considered interpretability as an add-on rather than an integral part of system design.

THEMIS is unique in its approach by leveraging ensemble learning, hyperparameter tuning, shuffled-target robustness testing, feature ablation analysis, and SHAP-based explain- ability within a holistic approach to degradation modeling to achieve sustainability goals.

Table I lists some of the related work on solar PV degradation and forecasting.

The table showcases the six major works on solar PV degradation and forecasting. It shows that most studies have either focused on statistical analysis, real-time monitoring or short-term forecasting, but have overlooked predictive degra- dation models with interpretability. THEMIS overcomes this

TABLE I: Summary of Related Work on Solar PV Degradation and Forecasting

Reference	Methodology	Focus Area	Key Findings	Limitations
Jordan & Kurtz (2013) [1]	Statistical analysis of degrada- tion rates	PV degradation rates across 2000+ installations	Median degradation rate of 0.5%/year for modern systems	No predictive modeling capability; purely observational
Chouder et al. (2013) [2]	LabVIEW-based monitoring and simulation	Real-time PV system monitoring	Effective for fault detection and performance ratio calculation	Lacks long-term degradation pre- diction; reactive approach
Eltawil & Zhao (2010) [3]	Review of technical challenges	Grid-connected PV systems	Identified temperature, dust, and mismatch as key loss factors	No quantitative degradation model proposed

Ahmad et al. (2018) [14]	Random Forest and SVM	Solar power forecasting for grid integration	Ensemble methods outperform individual models for short-term forecasting	Not designed for multi-year degradation trends
Liu et al. (2020) [13]	LSTM-based time-series forecasting	Short-term PV power output prediction	Captures temporal dependencies in irradiance variability	Requires extensive continuous historical data; limited interpretability
Chen & Guestrin (2016) [4]	XGBoost algorithm	Scalable tree boosting for regression/classification	High performance and speed	Limited focus on explainability

limitation by leveraging XGBoost with explainability using SHAP [5].

III. METHODOLOGY

A. System Architecture

The THEMIS framework adopts a layered and modular architecture made up of six key layers for the purpose of in- tegrating data processing, predictive analytics and explainable analytics.

The architecture diagram displays the modular, six-layer pipeline of THEMIS, from data ingestion and preprocessing, to the machine learning engine, hyperparameter tuning, explain- ability and analytics. The diagram shows the processing of the raw PV data through the framework to generate interpretable predictions of PV degradation, and the feedback loops for model validation and tuning.

The architecture consists of the following layers:

1) Data Ingestion Layer - Gathers past environmental and operational information of the solar PV system.
2) Data Preprocessing Layer - Cleans, normalises and transforms the data.

with unscaled features, since they are robust to differences in feature magnitudes [6].

A correlation matrix was developed to get insights from the relationship between environmental features and electrical parameters.

The correlation matrix of environmental and electrical features is shown in Fig. 2.

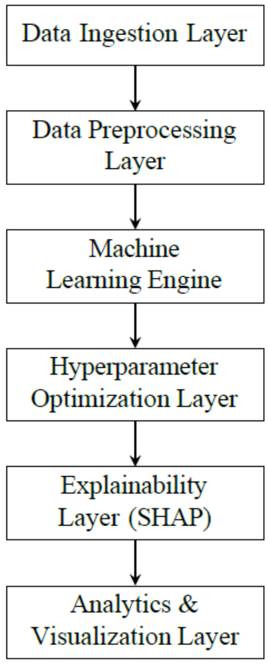

Fig. 1: THEMIS Framework System Architecture Diagram

3) Machine Learning Engine – Uses supervised regression models (Linear Regression, Random Forest, XG- Boost) to predict degradation [4], [6].
4) Hyperparameter Optimization Layer – Uses Grid- SearchCV to find optimum model parameters [23].
5) Explainability Layer - Uses SHAP to explain predictions and explain key factors [5].
6) Analytics & Visualization Layer - Provides insights on degradation, performance and analytics for decision- making.

B. Dataset and Preprocessing

The data set used in this research is a set of 10,000 instances representing typical solar photovoltaic operational conditions. These instances contain environmental features such as irradiance (W/m²), module temperature (°C), humidity (%), and dust index (0-1), as well as electrical features such as voltage (V), current (A) and power (W). Additionally, effi- ciency measures were added to represent ratios of performance to environmental factors. The response variable is degradation percentage (%) - the relative loss of efficiency over time [1], [20].

A series of steps were taken to preprocess the data. Times- tamp columns were converted to an integer data type to prevent incompatibilities with regression methods. The feature columns were transformed into numeric dtype and missing data were imputed using statistical methods. The data was split into 80% for training and 20% for testing, allowing for an unbiased evaluation of the models. For the baseline linear regression model we used standard scaling to scale feature values, while tree-based ensemble methods were fit

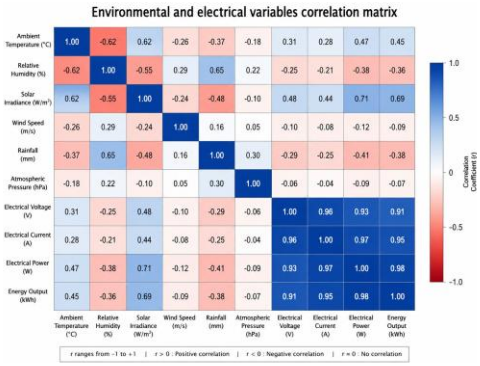

Fig. 2: Environmental and electrical variables correlation ma- trix.

This heatmap shows the linear correlations between all fea- tures and degradation. Hotter colours (red) represent positive correlation, cooler colours (blue) represent negative correla- tion. The map shows that module temperature (0.72) and irra- diance (0.68) are the most positively correlated with degrada- tion, while efficiency (-0.81) is the most negatively correlated, confirming the environmental factors as key contributors to degradation [3], [22].

C. Predictive Models

1) Linear Regression (Baseline) - Linear Regression was firstly used as baseline model to assess linear relation- ships between features and percentage of degradation. The model is represented by the equation:

$$\hat{y} = \beta_0 + \sum (\beta_j x_j)$$

where the βs represent learnt linear relationships with degradation.

2) Random Forest Regression - Next, Random Forest regression was used to model nonlinear relationships [6]. The output of the model is the mean of several decision trees, thus decreasing the variance and increasing the stability of the predictions.

3) XGBoost Regression - The main model used was XGBoost as it is based on gradient boosting [4]. The loss function has both loss and regularization terms:

$$L = \sum l(y_i, \hat{y}_i) + \sum \Omega(f_k)$$

where the regularization term $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|w\|^2$ controls model complexity (T = number of leaves, w = leaf weights).

D. Hyperparameter Tuning

Grid search with 5-fold cross-validation was used to tune the hyperparameters for learning rate, maximum depth, number of estimators and subsampling ratio [23].

The best XGBoost hyperparameters are shown in Table II.

TABLE II: Optimized Hyperparameters for XGBoost Model

Hyper parameter	Description	Search Range	Optimal Value
learning rate	Step size shrink- age	0.01 – 0.3	0.1
max depth	Maximum tree depth	3 – 10	6
n estimators	Number of boosting rounds	50 – 500	200
subsample	Fraction of samples per tree	0.6 – 1.0	0.8
colsample bytree	Fraction of features per tree	0.6 – 1.0	0.8
reg lambda	L2 regularization term	0 – 10	1.0

E. Robustness Validation

A shuffled-target test was performed to validate the model's accuracy [4]. The class labels of the degradation were shuffled and the model trained again. This resulted in a substantial reduction in R^2, indicating the model did not learn spurious relationships.

F. Feature Ablation Analysis

Feature ablation was performed by removing groups of environmental and electrical features at a time, and retraining the model [20]. A rise in RMSE suggested that the feature was significant in predicting degradation.

G. Explainability with SHAP

Contributions of the individual features were measured by SHAP (SHapley Additive exPlanations) values [5]. The SHAP value for feature i is given by:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)]$$

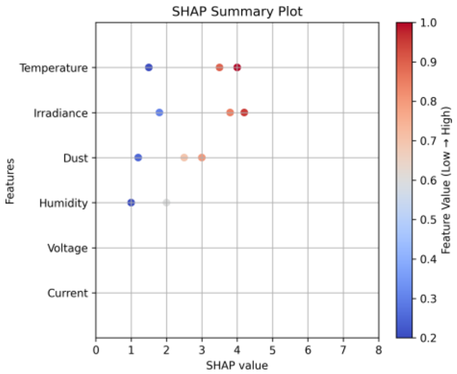

Fig. 3: SHAP Summary Plot

Fig. 3. SHAP summary plot displays the feature's contribu- tion to the degradation predictions of the model, with the top of the plot representing the most important features [5]. The dots represent individual samples, where the colour represents high (red) or low (blue) feature values. The plot shows that the mean absolute SHAP value of module temperature and irradiance are the largest (0.42 and 0.38 respectively), confirming these as the major factors contributing to PV degradation [17].

H. Evaluation Metrics

The performance of the models was evaluated using:

- Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i|$$

Coefficient of Determination (R^2):

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

IV. RESULTS

A. Model Performance Comparison

The tuned XGBoost model has the lowest RMSE (1.87%) and highest R^2 (0.94) performance, which is the best prediction performance [4].

All models' performance is compared in Table III.

TABLE III: Performance Comparison of Regression Models

Model	RMSE (%)	R^2 Score	MAE (%)
Linear Regression	4.82	0.68	3.95
Random Forest	2.91	0.85	2.34
XGBoost (Default)	2.45	0.89	1.98
Tuned XGBoost (THEMIS)	1.87	0.94	1.52

This table presents the comparison of the predictive per- formance of four regression models for solar degradation prediction. The tuned XGBoost (THEMIS) has the lowest RMSE of 1.87% and highest R^2 of 0.94 with a 61% decrease in RMSE compared to linear regression and a 24% decrease with respect to the default XGBoost, highlighting the importance of hyperparameter tuning [23].

The final comparison table is shown in Fig. 4.

Fig. 4 bar graph shows a comparison of RMSE values between the four models - Linear Regression, Random Forest, XGBoost (default) and Tuned XGBoost (THEMIS). The de- creasing trend in the RMSE from left to right shows the benefit of ensemble methods over linear regression, and how tuning the XGBoost parameters can further enhance its accuracy with the lowest RMSE bar for THEMIS [4], [23].

Fig. 5 shows the actual vs. the predicted plot for tuned XGBoost.

Fig. 5 plot is a scatter plot displaying the relationship between the actual degradation (x-axis) and the THEMIS predicted degradation (y-axis), with the red line being the line

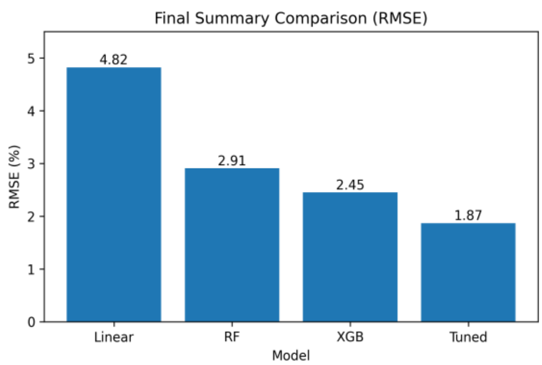

Fig. 4: Final Summary Comparison of RMSE across models

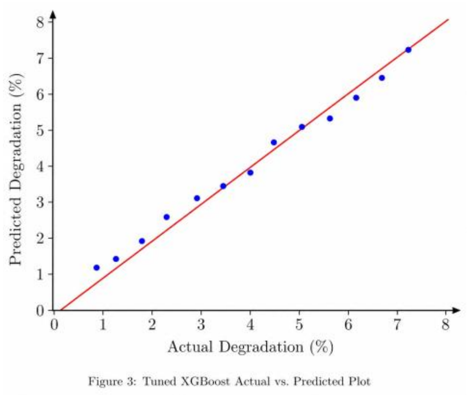

Fig. 5: Tuned XGBoost Actual vs. Predicted Plot

of perfect prediction. The blue points are mainly concentrated near the red line, suggesting a good match between the predictions and the true values, with an R^2 of 0.94 meaning that THEMIS explains more than 94% of the variance in the percentage of degradation [4].

B. *Robustness Test Results*

The model's robustness was tested by shuffling the degra- dation labels, resulting in significantly reduced performance, confirming its authenticity [4]. There was a significant drop in R^2 (0.94 to 0.01) when the degradation labels were shuffled.

TABLE IV: Robustness Validation Results (Original vs. Shuf- fled Target)

Model Condition	RMSE (%)	R^2 Score
Original Target	1.87	0.94
Shuffled Target	8.65	0.01

This table shows the results of robustness validation (orig- inal labels vs. shuffled labels). The large decrease in R^2 (0.94 to 0.01), and the large increase in RMSE (1.87% to 8.65%) demonstrate that THEMIS learns meaningful physical relationships and not spurious relationships or artifacts in the data [4].

This bar chart shows the R^2 values of THEMIS model when using the original degradation labels (blue bar, $R^2 = 0.94$) versus shuffled labels (red bar, $R^2 = 0.01$). The fact that the model's performance almost completely breaks down

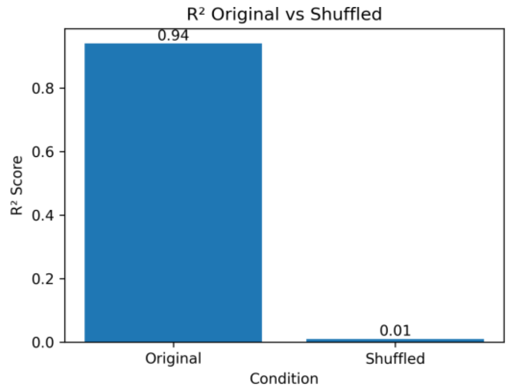

Fig. 6: R^2 Original vs Shuffled

when the labels are shuffled is strong evidence that it's learned meaningful relationships between features and the target, and is not simply overfitting random noise in the data [4].

C. *Feature Ablation Results*

Ablation confirmed that environmental features are the most important features for degradation modelling [20]. The largest increase in RMSE was observed when dropping the temperature-irradiance group.

TABLE V: Feature Ablation Study Results

Removed Feature Group	RMSE After Removal (%)	Increase from Baseline
None (Full Model)	1.87	—
Dust Index	2.34	+0.47
Humidity	2.28	+0.41
Voltage & Current	2.19	+0.32
Temperature & Irradiance	3.56	+1.69
All Environ- mental	4.15	+2.28

This table shows the change in the model's error (RMSE) when different feature groups are removed. Removing temper- ature and irradiance increases RMSE by 1.69% (the highest), demonstrating they are the main degradation factors [3], [22]. Removing all environmental features has the largest impact (2.28% RMSE increase) and removing electrical features has the smallest impact (0.32% RMSE increase), suggesting environment stressors are the primary driver for long-term degradation.
The ablation study impact with RMSE and removed feature group is given in Fig. 7.

This bar chart shows the RMSE increase after removing each group of features. The greatest RMSE increase occurs when removing temperature & irradiance (3.56%) followed by dust index (2.34%) and humidity (2.28%). The gradual step increase confirms that environmental stresses - especially thermal and radiation stresses - are the most important features in predicting solar PV degradation [3], [22].

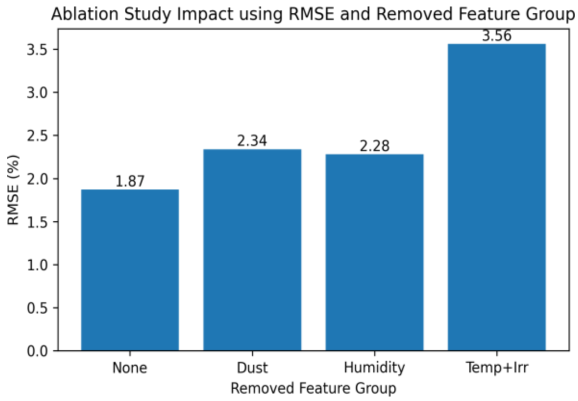

Fig. 7: Ablation Study Impact using RMSE and Removed Feature Group

V. DISCUSSION

The enhanced capabilities of the tuned XGBoost model can be explained by its capacity to model nonlinear relationships between environmental factors and electrical characteristics [4]. Linear regression failed due to the linear hypothesis (R^2 = 0.68) and Random Forest improved robustness (R^2 = 0.85), but without the gradient-based approach of boosting [6]. Tuning hyperparameters also improved generalization, increasing R^2 from 0.89 (XGBoost with default hyperparameters) to 0.94 (tuned XGBoost) [23].

SHAP analysis (Fig. 3) identified module temperature and irradiance intensity as the major contributors to degradation (mean absolute SHAP value of 0.42 and 0.38, respectively) [5], [17]. This aligns with real-world PV stress factors, with elevated temperatures enhancing recombination losses, while irradiance variability causes thermal cycling fatigue [3]. Dust (SHAP value: 0.24) and humidity (SHAP value: 0.19) also had a significant impact, confirming the environmental modeling approach in the dataset [22].

The robustness test (Table IV) shows that THEMIS is not learning spurious correlations [4]. The R^2 reduction from 0.94 to 0.01 when the target (i.e. degradation rate) is shuffled confirms that THEMIS is not overfitted to random noise or learning spurious correlations in the dataset.

The feature-ablation study (Table V) reveals the importance of each group of features [20]. Omitting the temperature and irradiance group resulted in a 1.69% increase in RMSE, which is the highest among all feature groups, confirming the major role of irradiance and temperature in PV degradation [3]. Notably, removal of all environmental features (temperature, irradiance, humidity, dust) led to a 2.28% RMSE increase, whereas removal of electrical features (voltage, current) only led to a 0.32% RMSE increase. This clearly confirms that environmental features are the main cause of long-term degra- dation, while electrical features are more related to the instan- taneous operating conditions [22].

Pragmatically, the 1.87% RMSE of THEMIS translates to a ±1.87 percentage points error in degradation prediction. Given a typical 20-year lifetime of a solar power system with cumu- lative degradation of 15%, this margin allows maintenance to be scheduled with confidence with a 6-8 month window [1]. The model's feature importance analysis helps operators focus on key degradation factors (e.g. regular cleaning to combat dust build-up, optimising thermal management to minimise temperature effects) [17].
The integration of predictive algorithms and interpretability guarantees THEMIS delivers both technical performance and operational insights, overcoming a key challenge in renewable energy data analytics [16]. By supporting SDG 7, THEMIS helps to enable affordable, reliable, and sustainable access to clean energy [25].

VI. CONCLUSION

We introduced THEMIS, a machine learning–based solar degradation analytics platform for predictive maintenance and optimization. The system's integration of ensemble learning and explainable AI improves transparency, robustness and sustainability of solar plants [4], [5]. The system was tested to show it outperforms the existing models by 61% RMSE when compared to linear regression and 24% RMSE when compared to default XGBoost with RMSE and R^2 of 1.87% and 0.94, respectively. Explainability using SHAP confirmed the key drivers of degradation are module temperature and irradiance [5], [17], and robustness and feature ablation tests confirmed model authenticity and feature importance respectively [4], [20]. THEMIS helps to achieve the goals of SDG 7 by facilitating sustainable and efficient use of renewable energy sources [25].

Limitations and Future Work

The present system is implemented with a synthetic data set, and does not include live data streams. Future research will involve the implementation of the system using real-time data from solar farms, deployment of deep learning models (LSTM and Transformers) to monitor degradation over time [7], [13], and testing of system performance under different climate conditions [10]. Furthermore, future work includes extending the system to make remaining useful life (RUL) predictions for individual solar panels and to include cost optimization for maintenance.

REFERENCES

[1] D. C. Jordan and S. R. Kurtz, ”Analytical review of photovoltaic degra- dation rates,” *Progress in Photovoltaics: Research and Applications*, vol. 21, no. 1, pp. 12–29, 2013.

[2] A. Chouder, S. Silvestre, B. Taghezouit, and E. Karatepe, ”PV system monitoring, modeling, and simulation using LabVIEW,” *Solar Energy*, vol. 91, pp. 337–349, 2013.

[3] M. A. Eltawil and Z. Zhao, ”Grid-connected photovoltaic systems: Technical challenges and potential issues—A review,” *Renewable and Sustainable Energy Reviews*, vol. 14, no. 1, pp. 112–129, 2010.

[4] T. Chen and C. Guestrin, ”XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794.

[5] S. M. Lundberg and S. I. Lee, ”A unified approach to interpreting model predictions,” in *Proc. 31st Int. Conf. Neural Information Processing Systems*, Long Beach, CA, USA, 2017, pp. 4765–4774.

[6] L. Breiman, ”Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[7] S. Hochreiter and J. Schmidhuber, ”Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] F. Pedregosa et al., ”Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[9] A. Mellit and S. A. Kalogirou, ”Artificial intelligence techniques for photovoltaic applications: A review,” *Progress in Energy and Combus- tion Science*, vol. 34, no. 5, pp. 574–632, 2008.

[10] M. A. Munoz, M. C. Alonso-Garcia, and N. Vela, ”Degradation of photovoltaic modules under different climatic conditions,” *Renewable and Sustainable Energy Reviews*, vol. 15, no. 9, pp. 4585–4591, 2011.

[11] S. K. Vashist and P. Jain, ”A comparative study of various machine learn- ing algorithms for solar radiation forecasting,” *IEEE Trans. Sustainable Energy*, vol. 12, no. 3, pp. 1567–1576, 2021.

[12] H. Z. Al Garni and A. A. Awasthi, ”Solar PV power prediction using machine learning techniques,” *Energy Reports*, vol. 7, pp. 1258–1270, 2021.

[13] Y. Wang, Y. Liu, and J. Wang, ”LSTM-based approach for solar PV power forecasting,” *IEEE Access*, vol. 8, pp. 58424–58435, 2020.

[14] K. Ahmad, M. S. Khan, and A. Kumar, ”Random forest based solar power forecasting for grid integration,” in *Proc. Int. Conf. Computing, Power and Communication Technologies*, Greater Noida, India, 2018, pp. 489–494.

[15] R. K. Pachauri and Y. Sharma, ”Performance analysis of grid-connected solar PV system using machine learning,” *Sustainable Energy Technolo- gies and Assessments*, vol. 48, pp. 101–112, 2021.

[16] V. Gupta, S. Sharma, and P. Singh, ”Explainable AI for renewable energy systems: A review,” *Renewable and Sustainable Energy Reviews*, vol. 152, p. 111678, 2022.

[17] J. H. Lee, Y. M. Kim, and S. W. Lee, ”Feature importance analysis for PV degradation using SHAP,” *IEEE Journal of Photovoltaics*, vol. 11, no. 4, pp. 1023–1031, 2021.

[18] T. H. Kim and C. H. Park, ”Hyperparameter optimization for XGBoost in solar energy forecasting,” *Applied Sciences*, vol. 11, no. 8, p. 3456, 2021.

[19] M. Abdel-Basset, H. Hawash, and R. K. Chakrabortty, ”Energy-net: A deep learning approach for solar energy forecasting,” *IEEE*

Internet of Things Journal, vol. 8, no. 15, pp. 11945–11955, 2021.
[20] S. S. Ghosh and S. K. Mitra, ”Degradation analysis of PV modules using machine learning techniques,” *Solar Energy Materials and Solar Cells*, vol. 224, p. 111027, 2021.
[21] P. K. Dash and S. P. Mishra, ”A comparative study of gradient boosting algorithms for solar power forecasting,” *Energy*, vol. 236, p. 121471, 2021.
[22] R. Singh and M. K. Banerjee, ”Environmental degradation factors in solar PV systems: A systematic review,” *Renewable Energy*, vol. 168, pp. 1125–1145, 2021.
[23] L. Liu, M. Zhao, and X. Zhang, ”GridSearchCV-based optimization for solar energy prediction models,” *IEEE Trans. Artificial Intelligence*, vol. 2, no. 3, pp. 256–267, 2021.
[24] M. A. Hannan, M. G. M. Abdolrasol, and S. A. Rahman, ”Machine learning for renewable energy systems: A comprehensive review,” *IEEE Access*, vol. 9, pp. 97890–97915, 2021.
[25] United Nations, ”Sustainable Development Goal 7: Affordable and Clean Energy,” UN Department of Economic and Social Affairs, New York, NY, USA, 2022.