Predictive Maintenance in Photovoltaic Systems Using Ensemble ML Empirical Analysis

Raihan Bin Mofidul*, Syed Samiul Alam*, Arbil Chakma*, ByungDeok Chung** and Yeong Min Jang*
*Dept. of Electronics Engineering, Kookmin University, Seoul, Korea
**ENS. Co. Ltd, Ansan, Korea
Email: raihanbinmofidul@ieee.org; syed.samiul.alam@ieee.org; arbilchakma@ieee.org;
bdchung@ens-km.co.kr; yjang@kookmin.ac.kr

***Abstract*—This paper aims to enhance the effectiveness and sustainability of photovoltaic (PV) systems by employing ensemble machine learning empirical analysis (EMLEA) to predict regular maintenance schedules using a minimal set of features. The research utilizes a 99.9kW PV system dataset, with a diverse set of features such as DC voltage, DC current, instantaneous power generation, power factor, and frequency. Ensemble machine learning (ML) algorithms including RandomForest, XGBoost, CatBoost, and LightGBM were deployed to forecast the regular maintenance needs of the PV system. The proposed EMLEA system identified DC voltage and DC current as key determining factors for the system's regular maintenance needs. Using these essential features, the maintenance schedule was estimated, and the performance metrics were evaluated by the EMLEA. Out of all models tested, XGBoost emerged as the superior model with an impressive accuracy of 98.62% and a precision of 94.37%.**

***Index Terms*—Predictive Maintenance, Photovoltaic Systems, Ensemble Learning, Feature Importance, Anomaly Detection**

I. Introduction

Predictive maintenance has emerged as a vital strategy to optimize the performance and reliability of diverse systems, such as photovoltaic (PV) systems, production lines, and equipment. The integration of machine learning (ML) techniques in predictive maintenance has shown encouraging outcomes in detecting failures, estimating remaining useful life (RUL), and enhancing system efficiency.

A. Literature Review:

This literature review endeavors to provide a comprehensive overview of recent research in the domain of predictive maintenance, concentrating on specific aspects and methodologies. The authors in [1] highlighted the need for a system-level predictive maintenance tool for PV systems to prevent failures. In the era of artificial intelligence, remarkable advancements have been made in harnessing its potential across various sectors from security and healthcare to energy management systems [2]–[4]. In [5], [6], researchers provided insights into machine learning (ML) methods used in predictive maintenance. According to [7], ensemble methods based on tree structures perform well in predicting PV power generation. Different feature engineering approaches and their impact on Remaining Useful Life (RUL) prediction accuracy was a key focus in [8]. A novel methodology addressing the challenges of constant power load (CPL) test beds was introduced in [9]. The importance of regular maintenance strategies in PV systems was emphasized in [10] in a comprehensive manner. A Decision Support System (DSS) for handling PV system faults was developed according to [11]. The article [12], [13] proposed a fault detection method suitable for maintenance in rural communities and anomaly detection of domestic home appliances. A Convolutional Neural Network (CNN) framework for predictive maintenance with impressive accuracy rates was presented in [14].

B. Research Gap:

Despite extensive research on predictive maintenance for PV systems, there is a research gap concerning the use of ensemble machine-learning techniques and minimal feature sets for predicting regular maintenance schedules. This paper addresses these gaps by applying ensemble machine learning algorithms to a PV system dataset, emphasizing the importance of features like DC voltage and DC current in maintenance requirements prediction. The resulting findings contribute to optimizing maintenance schedules for PV systems, which could lead to improved long-term system performance.

C. Our Novel Approach:

This paper proposes a method to optimize photovoltaic (PV) systems' efficiency and longevity by predicting regular maintenance schedules using ensemble machine learning algorithms and a minimal set of features from a 99.9kW PV system dataset. In summary, the primary contribution of our research comprises:

- Development of an exclusive system based on ensemble machine learning empirical analysis (EMLEA), emphasizing evaluation metrics and feature importance.
- Subsequently, identification of the most critical features and their utilization in predicting the maintenance schedule.
- Ultimately, recognition of the most efficient ensemble ML model based on an array of performance metrics.

The structure of the remainder of this paper is as follows: The methodology is presented in Section II. The experiment and its evaluation are detailed in Section III, while Section IV concludes the paper and discusses future work.

II. Methodology

The purpose of Algorithm-1 is to analyze photovoltaic (PV) system data using EMLEA including various evaluation metrics and determination of feature importance for predicting the maintenance schedule of the PV system. It aims to generate a set of trained models and comparison data for evaluating their performance. The input dataset contains various features, including Date-Time ($timestamp$), PV states (σ), DC voltage (V_{dc}), DC current (I_{dc}), DC power (P_{dc}), RS line voltage (V_{rs}), ST line voltage (V_{st}), TR line voltage (V_{tr}), R phase current (I_r), S phase current (I_s), T phase current (I_t), PV output power (P_o), power factor (pf), frequency (f), maximum power generation (P_{max}), and daily power generation (ΣP_{day}).

Algorithm 1: Ensemble ML Empirical Analysis with Evaluation Metrics and Feature Importance

Input : $PVSystemData$
/* $PVSystemData$: */
/* $timestamp, \sigma, V_{dc}, I_{dc}, P_{dc}, V_{rs}, V_{st}, V_{tr},$ */
/* $I_r, I_s, I_t, P_o, pf, f, P_{max}, \Sigma P_{day}$ */
Output: $TrainedModels, ComparisonData$
/* $ComparisonData$: */
/* $EvaluationMetrics, FeatureImportance$ */
Function `Main()`:
- $df(dataframe) \leftarrow PVSystemData$
- `/* Selecting only PV Features */`
- $X \leftarrow PVSystemData$ without $timestamp$
- `/* Periodic Maintenance Schedule */`
- $y(MaintenaceDuration) \leftarrow \begin{cases} 1 & \text{if } \sigma = 4 \\ 0 & \text{otherwise} \end{cases}$
- `/* Test and Train ratio is 4:1 */`
- $(X_{train}, X_{test}), (y_{train}, y_{test}) \leftarrow (X, y)$
- Define Ensemble ML $models \leftarrow$ $RandomForest, XGBoost,$ $CatBoost, LightGBM$
- **foreach** *model* ***in*** *models* **do**
  - Train *model* on $X_{\text{train}}, y_{\text{train}}$
  - Save $TrainedModels$
  - Predict on X_{test} to obtain y_{pred}
  - $ComparisonData \leftarrow$ $Accuracy, Precision, Recall, F1Score, AUC-ROC$
  - $FeatureImportance \leftarrow$ Compute using either Gini Importance or Permutation Importance
  - `/* For a feature F, Gini Importance (FI_G) can be calculated as */`
  - $FI_G(F) = \frac{1}{N}\sum_{t=1}^{N}(\text{Impurity Split}_t - \text{Impurity Left}_t - \text{Impurity Right}_t)$
  - `/* Permutation Importance (FI_P) can be calculated as */`
  - $FI_P(F) = \frac{1}{N}\sum_{i=1}^{N}(\text{Model Score}_{\text{original}} - \text{Model Score}_{\text{permuted}})$

The algorithm initiates with the establishment of a data frame, dubbed *df*, from the supplied PV system data. Only PV features, with the exclusion of timestamp, are chosen and delegated to *X*. The target variable, *y*, symbolizes maintenance duration as a binary figure - '1' implies a maintenance state ($\sigma = 4$), and '0' depicts all other states. The data and target variable are split into training and testing sets at a 4:1 ratio to evaluate model performance. Next, the algorithm declares four ensemble machine learning models - Random Forest, XGBoost, CatBoost, and LightGBM. Each model is trained on the training dataset and saved for later use. These models are then applied to make predictions on the test dataset. Finally, the algorithm computes various performance metrics, including accuracy, precision, recall, F1 score, AUC-ROC, and feature importance, for each model. The feature importance is calculated either via Gini Importance, a measure of a feature's contribution to class separation across all trees, or Permutation Importance, a measure of the impact of the feature's permutation on model performance. These metrics form the basis for a comparative evaluation of the model's capability to predict PV system maintenance durations. The proposed methodology entails three primary steps. Initially, the algorithm analyses Evaluation Metrics and Feature Importance across all available features. The second step involves the identification and application of the most influential features in maintenance schedule prediction. Finally, the algorithm determines the most effective ensemble machine learning model based on a comprehensive range of performance metrics.

III. Experiments

A. Feature Importance:

In Figure-1, the significance of different features was quantified using Algorithm-1. Two features stood out - DC voltage (V_{dc}) and DC current (I_{dc}), consistently receiving the highest scores across all models. XGBoost, in particular, attributed the highest scores of 0.36 and 0.26 to these features, thereby highlighting their essential role in predicting the maintenance schedules of PV systems.

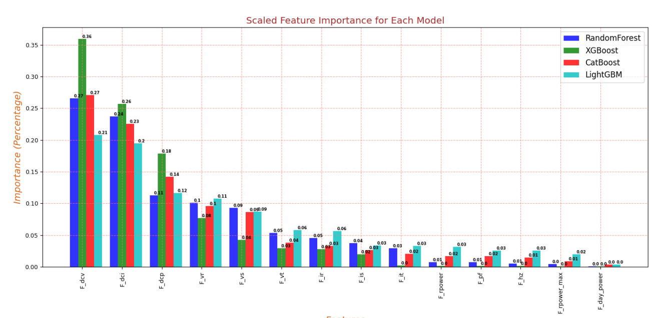

Fig. 1. Feature Importance

B. Performance Evaluation:

Algorithm-1 was redeployed using only two features, the DC voltage (V_{dc}) and DC current (I_{dc}). The performance

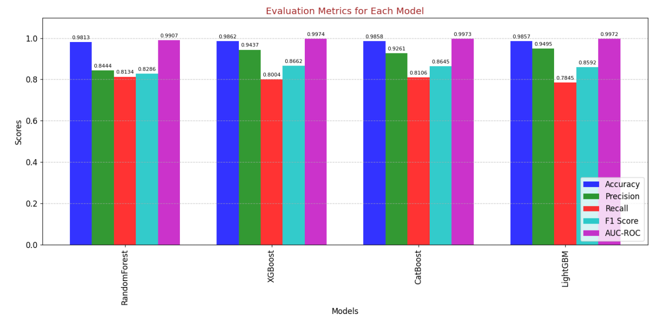

Fig. 2. Performance Evaluation

metrics shown in Figure-2 reveal that XGBoost has the highest accuracy (98.62%), precision (94.37%), and AUC-ROC score (99.73%) among the tested ensemble machine learning models. In terms of recall and F1 scores, XGBoost also demonstrates competitive performance, thus showcasing its effectiveness in predicting PV system maintenance schedules.

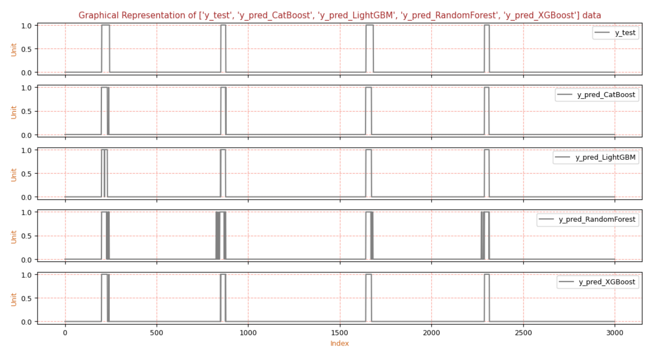

Fig. 3. Real-time Observation

Additionally, a visual comparison of the actual and predicted periodic maintenance schedules is provided in Figure-3 for a more intuitive understanding.

IV. Conclusion and Future work

This research underscores the significance and efficiency of ensemble machine-learning techniques in estimating the maintenance schedule for Photovoltaic (PV) systems. Through the utilization of these methods, potential issues were proactively identified, thereby enhancing the system's long-term efficiency and energy production. Through the application of Algorithm-1, the most important features were identified and exclusively used for predicting periodic maintenance. Among the tested models in EMLEA, XGBoost demonstrated superior performance, showcasing exceptional prediction accuracy (98.62%), precision (94.37%), and AUC-ROC (99.73%).

Future research could delve into other advanced machine learning or deep learning techniques for more insights into PV system maintenance. Additionally, plans include developing an XGBoost-based automated alert system to optimize PV system performance and longevity.

Acknowledgment

This work was partly supported by the Technology Development Program (S3098815) funded by the Ministry of SMEs and Startups (MSS, Korea) and the Ministry of Science and ICT (MSIT, Korea), under the ITRC (Information Technology Research Center) support program (IITP-2022-2018-0-01396) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

References

[1] L. B. Bosman, W. D. Leon-Salas, W. Hutzel, and E. A. Soto, "PV System Predictive Maintenance: Challenges, Current Approaches, and Opportunities," *Energies*, vol. 13, no. 6, p. 1398, Jan. 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/6/1398

[2] S. S. Alam, A. Chakma, M. H. Rahman, R. Bin Mofidul, M. M. Alam, I. B. K. Y. Utama, and Y. M. Jang, "Rf-enabled deep-learning-assisted drone detection and identification: An end-to-end approach," *Sensors*, vol. 23, no. 9, p. 4202, 2023.

[3] S. B. Shuvo, S. S. Alam, S. U. Ayman, A. Chakma, P. D. Barua, and U. R. Acharya, "Nrc-net: Automated noise robust cardio net for detecting valvular cardiac diseases using optimum transformation method with heart sound signals," *arXiv preprint arXiv:2305.00141*, 2023.

[4] M. M. Alam, R. B. Mofidul, and Y. M. Jang, "Community energy storage system: Deep learning based optimal energy management solution for residential community," *Journal of Energy Storage*, vol. 64, p. 107100, 2023.

[5] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. d. P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers & Industrial Engineering*, vol. 137, p. 106024, Nov. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360835219304838

[6] Y. Ran, X. Zhou, P. Lin, Y. Wen, and R. Deng, "A Survey of Predictive Maintenance: Systems, Purposes and Approaches," Dec. 2019, arXiv:1912.07383 [cs, eess]. [Online]. Available: http://arxiv.org/abs/1912.07383

[7] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression," *Energy*, vol. 164, pp. 465–474, Dec. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544218317432

[8] O. E. Yurek and D. Birant, "Remaining Useful Life Estimation for Predictive Maintenance Using Feature Engineering," in *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Oct. 2019, pp. 1–5.

[9] V. Le, X. Yao, C. Miller, and B.-H. Tsao, "Series DC Arc Fault Detection Based on Ensemble Machine Learning," *IEEE Transactions on Power Electronics*, vol. 35, no. 8, pp. 7826–7839, Aug. 2020.

[10] A. Abubakar, C. F. M. Almeida, and M. Gemignani, "A Review of Solar Photovoltaic System Maintenance Strategies," in *2021 14th IEEE International Conference on Industry Applications (INDUSCON)*, Aug. 2021, pp. 1400–1407.

[11] A. Livera, M. Theristis, L. Micheli, E. F. Fernández, J. S. Stein, and G. E. Georghiou, "Operation and Maintenance Decision Support System for Photovoltaic Systems," *IEEE Access*, vol. 10, pp. 42 481–42 496, 2022.

[12] W. B. Richardson, J. Meyer, and S. Von Solms, "Towards Machine Learning and Low Data Rate IoT for Fault Detection in Data Driven Predictive Maintenance," in *2021 IEEE World AI IoT Congress (AIIoT)*, May 2021, pp. 0202–0208.

[13] R. Bin Mofidul, M. M. Alam, M. H. Rahman, and Y. M. Jang, "Real-time energy data acquisition, anomaly detection, and monitoring system: Implementation of a secured, robust, and integrated global IIoT infrastructure with edge and cloud AI," vol. 22, no. 22, p. 8980. [Online]. Available: https://www.mdpi.com/1424-8220/22/22/8980

[14] W. Silva and M. Capretz, "Assets Predictive Maintenance Using Convolutional Neural Networks," in *2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Jul. 2019, pp. 59–66.