Optimizing PV Array Performance: A^2 LSTM for Anomaly Detection and Predictive Maintenance based on Machine Learning

Ali Hamza, Zunaib Ali*, Sandra Dudley
School of Engineering,
London South Bank University (LSBU)
London SE1 0AA, UK
hamzaa10@lsbu.ac.uk, aliz29@lsbu.ac.uk, dudleyms@lsbu.ac.uk

Komal Saleem
Department of Engineering and Construction,
University of East London (UEL)
London, UK
ksaleem@uel.ac.uk

***Abstract*—Photovoltaic (PV) energy is considered one of the most promising renewable sources. Detecting and monitoring faults in PV systems ensures optimal efficiency and prevents safety and hazards. Predictive maintenance (PdM) is the prominent anomaly prediction strategy that predicts health conditions with machine learning (ML) algorithms. However, existing algorithms overlook the importance of attribute consideration and fail to account for temporal dependence in final results. To address such issues, this paper proposes the implementation of Attribute Attention (A^2)-based Long short-term memory (LSTM) for general PdM framework based on clustering and anomaly detection in PV array data. The A^2-LSTM model is complemented by an unsupervised *KMeans* clustering technique to identify patterns within the data. The attention mechanism in the attribute attention-based LSTM model is used to identify the most relevant attributes in PV array electrical data for each cluster, allowing the model to focus on the information that is most pertinent to predicting the behavior of the PV arrays within that cluster. The results indicate that the proposed model identified anomalies in the predicted data of the PV array more accurately. The proposed model could help the plant operator perform Remaining Useful Life (RUL) for PdM to carry out PV array maintenance. To the best of our knowledge, the A^2 method is not been used for the PdM problem of PV plants.**

***Index Terms*—Photovoltaic System, Anomaly detection, $KMeans$, Attribute attention, Long-short term memory**

I. Introduction

Predictive maintenance (PdM) for solar farms has become a crucial element in effectively operating and maintaining the plant [1]. PdM can be utilized in the PV plant to anticipate the probable failure of components such as PV modules, inverters, and battery systems. Anomaly detection is one of the key stages in PdM to detect abnormalities within the system that may result in future system failures [2]. Detecting anomalies in PV plants poses challenges due to the vast number of string modules and extensive data volumes. However, Machine Learning (ML) algorithms offer faster, cost-effective, and more accurate analysis, facilitating anomaly detection amid the substantial data collected by the system. The increasing adoption of ML in the engineering sector, especially in PV, has led to diverse applications for PdM and anomaly detection. Researchers in [3] proposed a $KMeans$ and LSTM-based prediction algorithm for electricity load prediction by clustering high and low temperature, humidity, and other load characteristics. Authors in [4] implemented a prediction model by combining $KMeans$ clustering of training set and prediction day data and LSTM for short-term power prediction in PV plants. A data-driven fault prediction based on LSTM and auto-encoder has been proposed in [5] for the solar array. Work in [6] addressed anomaly detection in the PV components by evaluating the performance of autoencoder LSTM, isolation forest, and Facebook prophet. It helped in identifying the PV system's healthy and abnormal behavior. Isolation forest-based anomaly detection for solar power plants has been implemented in [7]. It also used a rule-based fault localization technique to locate the abnormal behavior of the solar plant. Several approaches have been proposed such as [8] about survey-related anomaly detection techniques, [9] for monitoring PV systems by identifying anomalies by One-Class Support Vector Machine and K nearest Neighbor, and [10] for PdM and anomaly detection approach for PV designs. Anomaly detection using a semi-supervised ML model to predetermine solar panel settings, as suggested by [11], to prevent PV component failure. Another study [12] introduced anomaly detection for PdM of large-scale solar PV plants using $KMeans$ and LSTM. Specifically, the study developed a clustering model focusing on the modules's output current for anomaly detection.

While PdM is effective in fault detection, anomaly detection in PdM faces challenges related to feature relevance and the dynamic nature of data. PV systems generate vast amounts of data (Voltage, current, temperature, yield measurements), etc. Traditional anomaly detection algorithm usually fails to select and prioritize relevant features at each time step. Deficiencies in attribute features and the presence of redundant attributes contribute to poor performance. Also, the PV system exhibits high variability due to environmental conditions e.g., temperatures and irradiance level. Another significant challenge is the nonlinear relationship between features and their impact on the system's behavior, making it difficult to accurately capture and analyze anomaly patterns. These issues, make it more difficult for traditional anomaly detection techniques to accurately identify anomalies.

To solve this problem, this paper proposed an $KMeans$ based $A^2 - LSTM$ for anomaly detection based on electrical data of the PV array. $KMeans$ clustering preprocess the input data by grouping similar instances (time series segments) based on their attribute values. This segmentation create more meaningful input representations for LSTM models. $KMeans$ simplifies the learning task for A^2-LSTM, allowing it to focus on learning patterns within homogeneous clusters rather than dealing with the entire dataset at once. A^2-LSTM employs an attention mechanism to dynamically select which features (attributes) are most relevant at each time step. Unlike traditional LSTMs which treat all input features equally, A^2-LSTM assigns higher weights to attributes that are more informative for anomaly detection at a particular time. This includes three distinct parts namely feature extraction and fusions, attribute attention mechanism, and anomaly detection. Varying window sizes, representing the length of the sequence, have been taken into account as input to the model for prediction purposes. Starting with $KMeans$ clustering to group instances of data, distinct clusters based on features like power, current, voltage, solar irradiance, and total yield are created. The attribute attention mechanism based on Global average pooling (GAP) is applied to the features (F), condensing temporal features into a fixed-length vector representation. This pooled output is subsequently fed into Fully Connected Neural Network (FCNN) layers to refine the attribute impact (reweighted attributes - Q). Finally, the Q are input into the anomaly detection layer including LSTM with a Fully connected (FC) layer to create a robust anomaly detection system capable of effectively detecting anomalies in the data.

The paper is organized as follows: Section II presents the PV system and data description. In Section III, the methodology has been demonstrated. Section IV presents the results and discussion. Section V focuses on the conclusion and future recommendations

II. PV System and Data Description

This section presents two PV systems: Grid-Connected Photovoltaic System (GCPVS) model and real PV plant location in Greece. This system serves as the basis for anomaly detection and PdM study. The detailed description and data characteristics of each system are described below:

A. Grid Connected Photovoltaic System (GCPVS)

A historical dataset including current, voltage, solar irradiance, and power has been synthetically generated using a MATLAB Simulink model of a GCPVS. The dataset spans 480 days, with measurements recorded at 10-minute intervals throughout the recorded period. A GCPVS has been modeled shown in Fig.(1) to construct the current dataset to be used for anomaly detection. The modeled system involves 36 PV modules interconnected within a 6×6 PV array TCT configuration. Using the historical weather data, the current dataset is generated shown in Fig.(3).

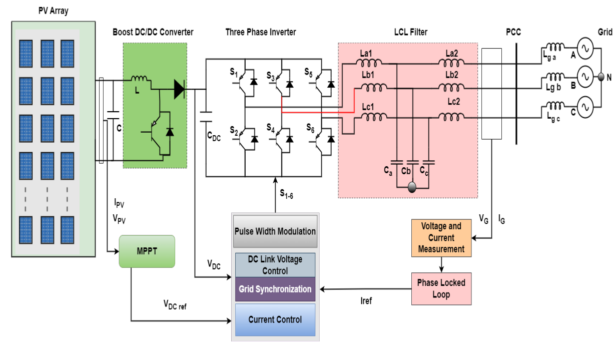

Fig. 1. PV Grid Connected System

*1) **PV Array configuration: TCT**:* The configuration under consideration is Total-Cross-Tied (TCT) and MATLAB simulation is employed to model this configuration, allowing for a comprehensive analysis and performance evaluation. The simulation involves 36 PV modules interconnected within a 6×6 PV array configuration. The aforementioned configurations are operated under normal as well as faulty conditions.

The TCT configuration is modeled as follows: initially, PV modules in strings are arranged in parallel as rows, and these rows are then connected in series. In this configuration, each row has the same voltage as each module, and the PV array's output voltage is the sum of individual row voltages. The array's output current is the sum of individual module currents in a row. The 6×6 TCT PV configuration is depicted in Fig. (2). Inter and intra line-to-line faults and Partial Shading faults are induced individually as anomalies one by one into the PV arrays, and the voltage, current, and total yield were calculated and compared with the normal condition of the PV array as shown in Fig.(3). Fig.(4) depicts the distribution of the normal and anomaly clusters based on features comparison (V vs I, V vs P, P vs G), etc. It clearly distinguished the normal region denoted by 0 and the anomaly region denoted by 1.

B. Real PV plant

The used data was collected at the PV plant located in Greece over 365 days, with 20 min intervals. The plant sensors measure the generation rate, DC and AC powers known as internal factors that could cause anomalies. The inverters at the plant level measured external factors such as ambient and module temperature, windspeed, and solar irradiance. The inverter DC power is sourced from the PV array. The correlation matrix depicted in Fig.(6) represents the linear correlation between internal and external variables. The value of the linear correlation ranges from -1 to 1, where -1 indicates a strong negative correlation between variables, and 1 indicates a strong positive correlation. It shows both the strength and direction of the correlation relationship which is determined by the Spearman's rank correlation [13] as follows:

$$\varrho = 1 - \frac{6\sum \kappa_i^2}{\tau(\tau^2 - 1)} \tag{1}$$

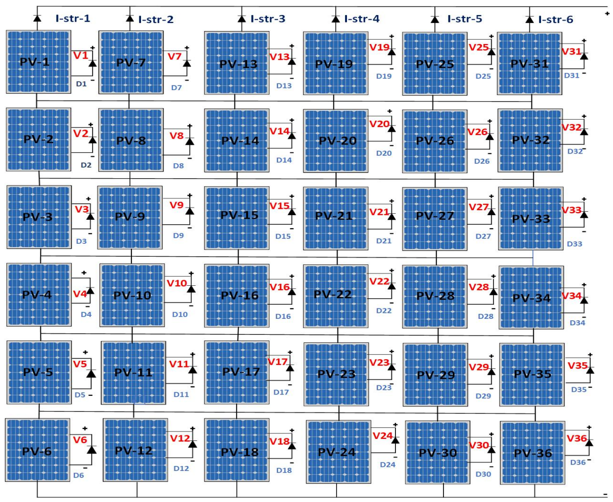

Fig. 2. TCT configuration of PV array.

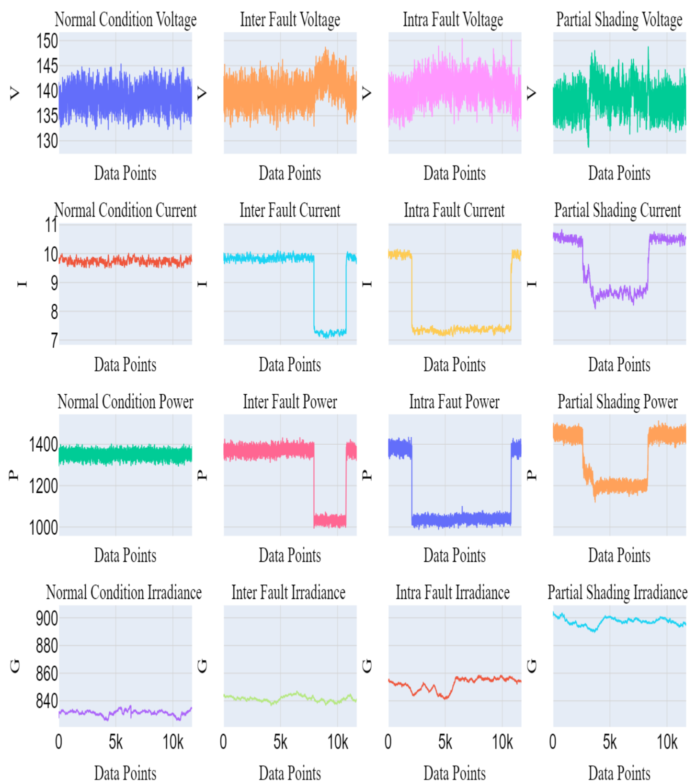

Fig. 3. Normal and Fault data distributions.

Where, ϱ is the correlation coefficient, κ is the difference between two ranks for every data point and τ is the number of the data points. Fig.(6) depicts that PV production, inverters' energy, irradiance, and module temperature are highly correlated except for the ambient temperature and digital windspeed. The PV production power is the sum of all power provided by the 23 inverters (Energy - Fronius IG TL 3.6_i and Energy - Fronius IG TL 4.0_i). It displays 6 out of 23 inverters'

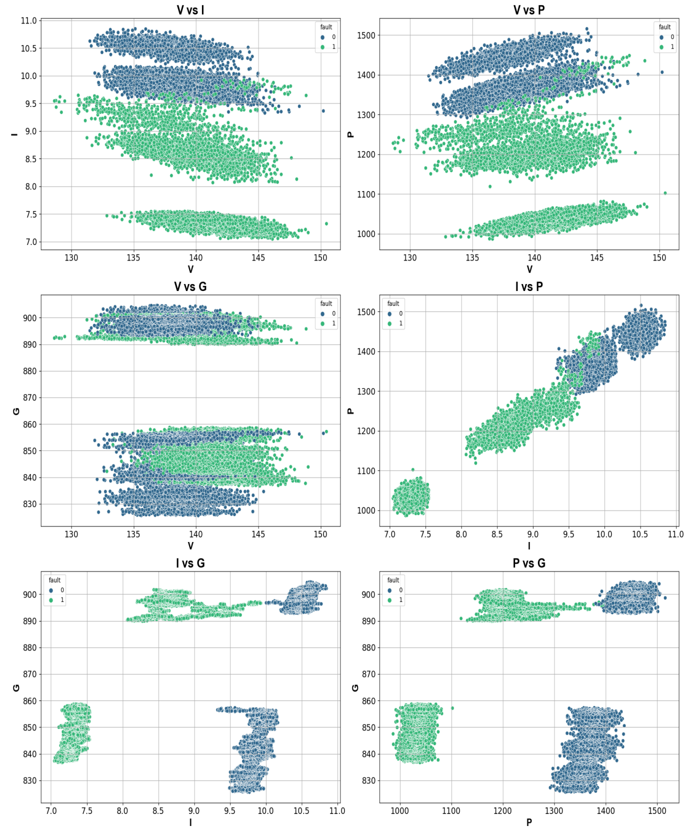

Fig. 4. Dataset for A^2 attribute selection..

DC power because all the inverters show highly correlated data.

III. Proposed Anomaly Detection and PdM Framework

The proposed methodology involves data preparation, *K Means* clustering, attribute selections, and anomaly detection with LSTM, depicted in Fig. (5). A^2-based LSTM with *K*-Means has been employed which is well suited for handling large datasets including historical as well as current data. The clustering by *K*-Means is followed by the attribute attention layer where GAP and FCNN are added to adjust the importance of attributes automatically. After reweighting attributes, it is input to the LSTM and FC layer for anomaly detection.

A. Feature Extraction and Fusion

This study uses $KMeans$ to cluster PV array electro-power matrics with solar irradiance and module temperature. The $KMeans$ method divides a dataset into K clusters, with each cluster distinguished by its centroid. Each data point is iteratively assigned to the nearest centroid by the algorithm after choosing K initial centroids. Applying KMeans clustering to the dataset to group similar points:

$$\gamma_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i \forall k = 1, 2, ..., K \tag{2}$$

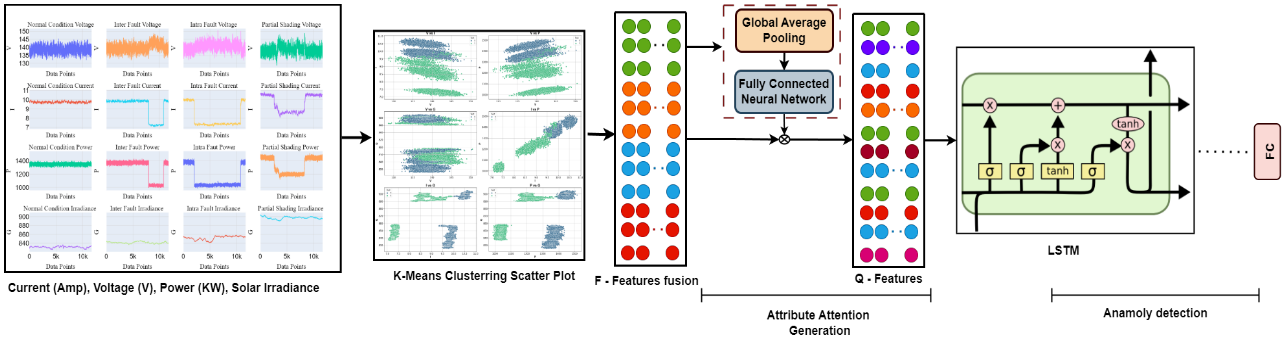

Fig. 5. A^2 LSTM based Anamoly detection overview

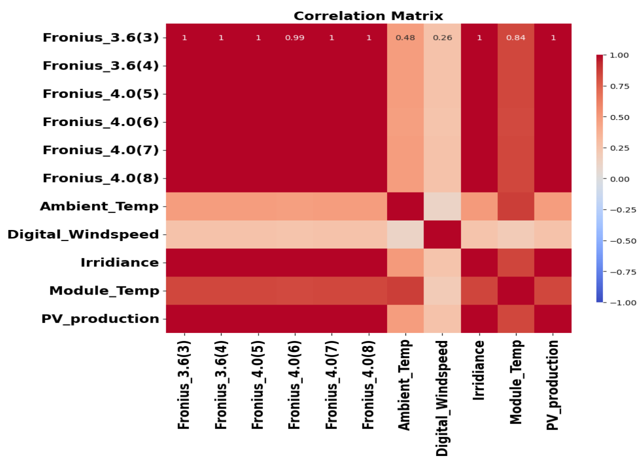

Fig. 6. Correlation matrix illustrating linear correlation among the internal and external elements for PV plant

Where γ_k is the centroid of cluster k and C_k represents a set of points in the cluster. Extract features from each cluster k: F_k = Feature extraction$*(C_k)$ and combining, we have:

$$F_{fusion} = Fusion(F_1, F_2, ...F_k) \tag{3}$$

B. Attribute attention Mechanism

The A^2 architecture comprises multiple layers: shrinking, learning, and reweighting. Further, within A^2 layer, The shrinking step corresponds to the GAP, applied to fused features within each cluster for data dimensionality reduction. The GAP operation computes the average of the features for all data points assigned to a specific cluster, resulting in a fixed-length vector representation for that cluster. It takes fused features (F_{fusion}) as input and is implemented as a timestamp.

$$F_{avg} = \frac{1}{n}\sum_{i=1}^{n} F_{fusion}[i] \tag{4}$$

Where n is the number of features in F_{fusion}. The fixed-length vectors representing each cluster are then fed into an FCNN for further processing. The learning step is to take the pooled F_{avg} as input which is processed through FCNN layers to generate $Q - features$.

The attention score A as output through multiple layers is stated as:

$$A = softmax\left(\sigma(G_{fn1} * F_{avg} + y_{fn1})\right) \tag{5}$$

Where G_{fn1} and y_{fn1} are the parameters of the the first layer of FCNN. The second layer's reweighting step corresponds to the multiplication process of A and F_{fusion} as input. The reweighted Q-feature as output, formulated as:

$$Q_t = F_{fusion} * A = F_{fusion} * softmax \\ \left(\sigma(G_{fn1} * F_{avg} + y_{fn1})\right) \tag{6}$$

C. Anamoly detection

Following the FCNN, the output is passed into the final component of anomaly detection. It consists of two units, i.e., LSTM and FC. LSTM network, which is a type of recurrent neural network (RNN) well-suited for modeling sequential data. The LSTM network captures temporal dependencies within the data, allowing it to learn patterns and detect anomalies over time. The LSTM takes reweighted features Q_t as input as shown in Fig.(7) and the FC takes these features to identify deviations from normal behavior indicative of anomalies. Each unit is discussed as follows:

LSTM module consists of a cell state, hidden state, and three

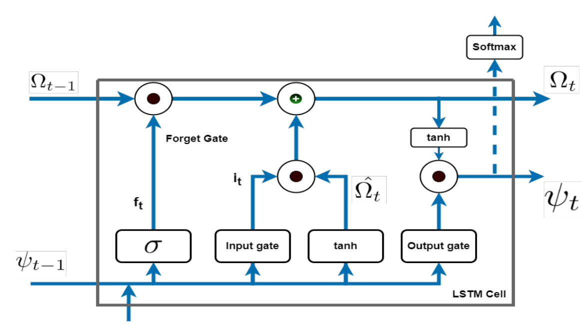

Fig. 7. LSTM Network for anomaly detection

gates (forget, Input, and output gate). The information between states and gates to be carried is automatically managed by the LSTM unit which adds or removes information as needed. The information dropped from the cell state is managed by the forget gate to determine what information should be removed. This decision is made by sigmoid function denoted by σ and its value is in between '0' and '1'. '1' represents completely keeping this information and '0' shows removing the information completely. The forget gate is formulated as:

$$f_t = \sigma\left(G_f \cdot [\psi_{t-1}, Q_t] + y_f\right) \tag{7}$$

Where G_f and y_f are the forgat gate's parameters. The next step is to determine what new information going to store in the cell state Ω_t. This step has two parts: First, the sigmoid layer called the 'Input gate layer' denoted by 'i_t' will update the value, and next, '$tanh$' creates a new vector candidate $\hat{\Omega}_t$, that could be added to the state, as:

$$i_t = \sigma\left(G_i \cdot [\psi_{t-1}, Q_t] + y_i\right) \tag{8}$$

$$\tilde{\Omega}_t = \tanh\left(G_C \cdot [h_{t-1}, Q_t] + y_C\right) \tag{9}$$

where, G_i, y_i, G_C and y_C are the parameters of the input gate. Equations.(8 and 9) are combined to update the state and can be represented as:

$$\Omega_t = f_t * \Omega_{t-1} + i_t * \tilde{\Omega}_t \tag{10}$$

The old state Ω_{t-1} is multiplied by the forget state f_t to forget the things, that decided to be forgotten earlier. then new candidate value $i_t * \hat{\Omega}_t$ is added to decide how many updates it should be for each state value. LSTM output gate updates the hidden state and determines which portion of the cell state is to be output, formulated as:

$$o_t = \sigma\left(G_o\left[\psi_{t-1}, Q_t\right] + y_o\right) \tag{11}$$

$$\psi_t = o_t * \tanh\left(\Omega_t\right) \tag{12}$$

Where, G_o, y_o are the output gates's parameters. The $tanh$ represents the tanh activation function (values to be between -1 and 1).

Given the output from the LSTM, FC performs a linear transformation followed by a nonlinear activation function formulated as:

$$\epsilon = G_{FC} * \psi_t + y_{FC} \tag{13}$$

Where G_{FC} and y_{FC} are the parameters of the FC layer. α is an activation function (sigmoid or Rectified Linear Unit (RLU)).

$$\alpha = \vartheta(\epsilon) \tag{14}$$

The output of the FC layers α can be used in a regression-based approach to detect the anomaly. The value $\hat{y}$ predicted by the model, $\hat{y} = \alpha$ and the anomaly score is computed based on the difference between the actual and predicted one. If the score crosses a certain limit, the anomaly is detected.

$$Anamoly\, score = |y - \hat{y}| \tag{15}$$

The algorithm for A^2-LSTM is summarized in Table.(I) for anomaly detection:

TABLE I
STEP-WISE EXECUTION OF $KMeans$ BASED A^2-LSTM

Algorithm: A^2 - LSTM
1: Perform $KMeans$ clustering on the dataset to get cluster assignment γ_k.
2: Extract features F_k and fuse them in a fixed-length vector representation F_{fusion}.
3: Apply GAP F_{avg} on fused features.
4: Feed the GAP-fused features into FCNN to extract Q_t.
5: For sequential processing, give Q_t as input to the LSTM.
6: Feed the LSTM output ψ_t to FC layer for anomaly detection.
7: Perform Anomaly score.

IV. Results and Discussion

In this section, we present the experimental evaluation of the proposed algorithm for two distinct scenarios: The Simulink model of the GCPV System and a real PV plant. The primary focus is to assess the effectiveness of the proposed algorithm for anomaly detection in both simulated and real-world PV system applications.

A. *Case 1: GCPV System*

In the first case, we evaluate the algorithm performance using the simulink model of the GCPV system. *K Means* has been applied to the simulation data to form the clusters and identify patterns indicative of normal and faulty behavior depicted in Fig.(8). The elbow technique has been used to determine the optimal number of clusters in the dataset which are 5 here. The elbow method heuristic is used to determine the optimal number of clusters in a clustering algorithm. It involves plotting the sum of squared distances (SSE) from each data point to its assigned centroid for different values of *K* and then identifying the "elbow" or bend in the plot where the rate of decrease in SSE slows down. To assess the quality of clusters, we examined their purity which quantifies the degree to which each cluster consists of data points belonging to a normal denoted by (0) or anomaly region denoted by fault (1). The purity analysis provides valuable information on the clustering accuracy and the algorithm's efficacy in detecting anomalies. The cluster purity for clusters 0 to 4 is 0.99, 1.00, 1.00, 0.98, and 1.00 respectively.

The effectiveness of anomaly detection can be enhanced by deploying clusters for the training of A^2-LSTM, enabling feature extraction and dynamically weighting critical features such as module electrical data and solar irradiance. The processed data is fed into the pooling layer followed by FCNN to abstract and reduce dimensionality. After training the A^2-LSTM, the model was used to detect and predict anomalies by establishing a threshold by identifying the instance where the error between actual and predicted was notably high. Fig.(9) depicts A^2-LSTM detected anomalies and classified as fault '0' for different clusters based on learned patterns and attribute weighted features.

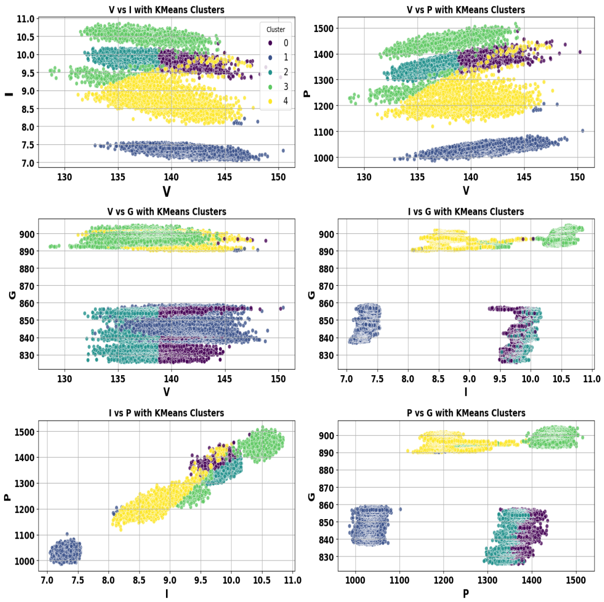

Fig. 8. *K Means* Clusters on attributes

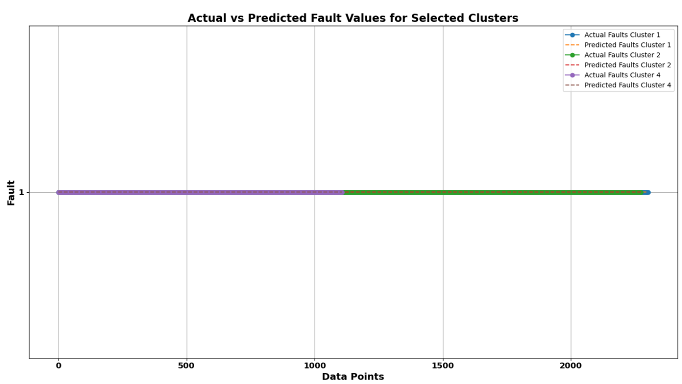

Fig. 9. Actual vs Predicted Fault Values for Selected Clusters

B. Case 2: Real PV Plant

The study detailed an analysis of the power production from 23 inverters and the aggregate PV production output, highlighting an anomaly in the dataset. Fig.(10) illustrates the relationship between the PV plant power production, module temperature, and solar irradiance for January month. Covering the period from January 1, 2012, to December 31, 2012, the analysis spans the entirety of the PV plant's operational timeline. The initial examination contrasts total PV production with irradiance. Notably, on August 18, between 15:30 and 16:00, a significant irregularity emerges. During this interval, there is a discernible glitch in PV production, as depicted in Fig.(11). This glitch occurs simultaneously in all the provided inverter data (Energy — Fronius IG TL 3.6 (i) series and 4.0(i) series. This should be investigated and A^2-LSTM is being developed to detect such anomalies. By leveraging these predictions, the maintenance team can schedule proper maintenance activities to address and remove these anomalies for successful PdM proactively.

To achieve this, the first phase involved investigating different parameters, with a particular focus on those that received high attention scores such as total PV power production, module temperature, and solar irradiance. Total Power production is a most holistic measurement of the plant performance. Similarly, module temperature and solar irradiance more directly impact the efficiency of the plant power production as compared to windspeed and ambient temperature. These feature selections can improve the A^2-LSTM Performance by removing the redundant attributes. The network parameters $G_{f_{n1}}$, $G_{y_{n1}}$, G_f, y_f, G_i, G_c, y_c, y_i, G_o, y_o, G_{FC} and y_{FC} are randomly initialized. The second phase focuses on the training of the proposed model. The A^2-LSTM has trained 120 epochs, with a batch size of 32. To access the overall performance of the A^2-LSTM, a confusion matrix has been used to quantify the anomaly detection performance. Accuracy, Precision, F1 score, and Recall based on the confusion matrix have been used as evaluation metrics.

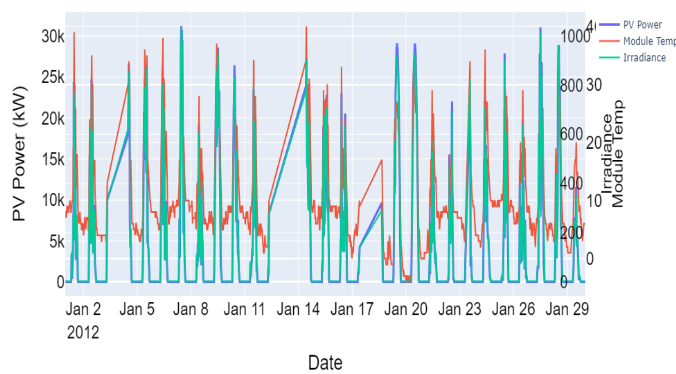

Fig. 10. Plant power production Vs Irridance Vs Modular temperature

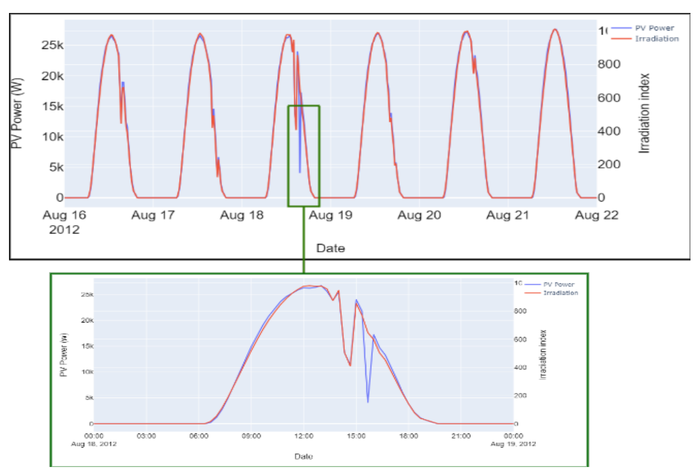

Fig. 11. Plant anomaly behavior

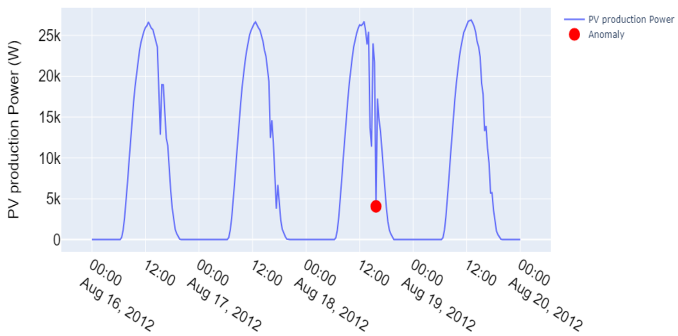

Fig. 12. A^2-LSTM Anomaly detection outcome for PV production

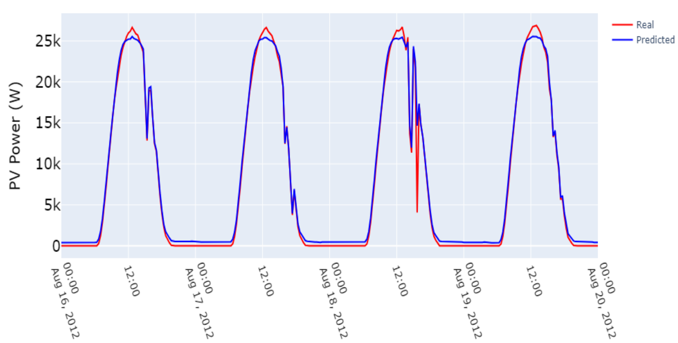

Fig. 13. A^2-LSTM PV Power production prediction

The A^2-LSTM successfully detected a significant power glitch that happened between August 18 to 19. This anomaly is marked by the sudden drop in PV power production, which was correctly identified and depicted in Fig.(12), demonstrating its ability to identify deviation from the normal operational pattern of the PV plant. Although other anomalies are present, successfully detecting this malfunction demonstrates the robustness of the A^2-LSTM in real-world PV systems. The anomaly was detected as a notable deviation from the forecasted values by comparing the actual PV production power data with forecasted power by the proposed model as shown in Fig.(13).

The A^2-LSTM has been evaluated using different window lengths: 1, 2, 4,8, 16, 32, and 64-time stamps. This allowed the assessment of the influence on the anomaly detection capabilities by the A^2-LSTM. This comparison showed how sensitive the model is to window length and how it affects the detection and prediction of anomalies and the optimal window size for effective anomaly detection. Fig. (14). shows the comparison based on different window sizes for accuracy, precision, recall, and F1-score of the A^2-LSTM with LSTM provides valuable insights into the performance of both models. A^2-LSTM demonstrate impressive accuracy, precision, recall, and F1-score as compared to LSTM across various window length. With large window lengths (8 and 16), A^2-LSTM showed noticeable improvement in precision and F1-Score, indicating its improved ability to balance false positive and true positive rates. On the other hand, LSTM performed consistently but less across performance matrics and different windows. Handling computational resources for A^2-LSTM becomes challenging when window length exceeds 32, as computational complexity increases significantly, and also, the anomaly detection becomes irrelevant to the data from longer intervals. Finally, the window length is chosen to be 8, since it covers the most relevant anomaly detection data for the PV System. A^2-LSTM showed better performance in handling anomaly detection due to its advanced architecture.

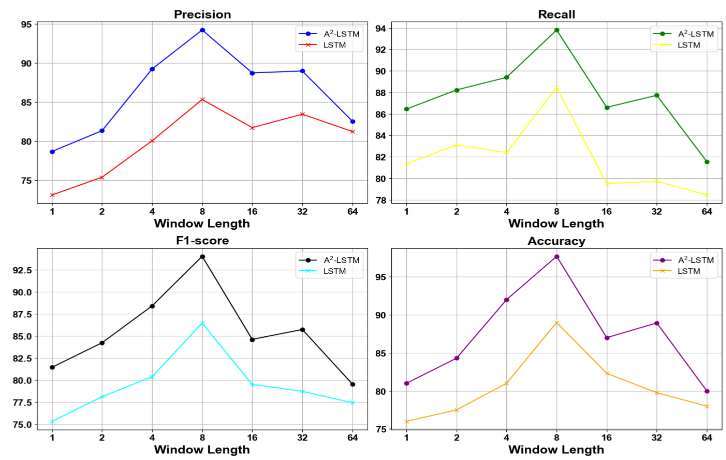

Fig. 14. Evaluation metrics of the $A^2 - LSTM$ vs. $LSTM$ at different lengths of window size.

V. Conclusion

This paper proposed $KMeans$ based A^2-LSTM framework for anomaly detection in PV array for PdM. During the experimental evaluation, it was observed that the clustering of data, the attribute attention mechanism for feature extraction, and window size choice significantly impacted the final performance of the model for anomaly detection. The effectiveness of the proposed algorithm was demonstrated through a real-world application (a real PV plant), and comparison results suggested that the A^2-LSTM achieved superior performance. Future work includes gathering real-time data using a 3×3 indoor PV setup, Remaining Useful Life (RUL) insights into the health and condition of PV plant equipment, classifying detected anomalies into specific fault classes, and utilization of a combination of expert knowledge with generational models such as variational autoencoders to improve the feature extraction.

References

[1] M. De Benedetti, F. Leonardi, F. Messina, C. Santoro, and A. Vasilakos, "Anomaly detection and predictive maintenance for photovoltaic systems," *Neurocomputing*, vol. 310, pp. 59–68, 2018.

[2] Y. I. A. A. Rahim, H. Zainuddin, E. A. Setiawan, *et al.*, "Performance of grid-connected photovoltaic systems in northern and southern hemispheres under equatorial climate," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 15, no. 2, pp. 858–873, 2024.

[3] D. Geng, H. Zhang, and T. Xu, "Effective lstm with k-means clustering algorithm for electricity load prediction," in *Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence*, 2019, pp. 476–481.

[4] T. Dong, S. Liao, W. Li, Z. Li, and Y. Lv, "Aggregate distributed photovoltaic power joint prediction method based on lstm," in *2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, IEEE, 2021, pp. 155–158.

[5] Q. Xue, Y. Cheng, B. Jiang, and X. Han, "Fault prediction for solar array based on long short-term memory and autoencoder," in *2021 40th Chinese Control Conference (CCC)*, IEEE, 2021, pp. 4567–4572.

[6] M. Ibrahim, A. Alsheikh, F. M. Awaysheh, and M. D. Alshehri, "Machine learning schemes for anomaly detection in solar power plants," *Energies*, vol. 15, no. 3, p. 1082, 2022.

[7] S. Kabir, A. Shufian, and M. S. R. Zishan, "Isolation forest based anomaly detection and fault localization for solar pv system," in *2023 3rd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, IEEE, 2023, pp. 341–345.

[8] A. Toshniwal, K. Mahesh, and R. Jayashree, "Overview of anomaly detection techniques in machine learning," in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, IEEE, 2020, pp. 808–815.

[9] M. Benninger, M. Hofmann, and M. Liebschner, "Online monitoring system for photovoltaic systems using anomaly detection with machine learning," in *NEIS 2019; Conference on Sustainable Energy Supply and Energy Storage Systems*, VDE, 2019, pp. 1–6.

[10] W. Chine, A. Mellit, V. Lughi, A. Malek, G. Sulligoi, and A. M. Pavan, "A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks," *Renewable Energy*, vol. 90, pp. 501–512, 2016.

[11] C.-W. Tsai, C.-W. Yang, F.-L. Hsu, H.-M. Tang, N.-C. Fan, and C.-Y. Lin, "Anomaly detection mechanism for solar generation using semi-supervision learning model," in *2020 Indo–Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN)*, IEEE, 2020, pp. 9–13.

[12] I. A. Zulfauzi, N. Y. Dahlan, H. Sintuya, and W. Setthapun, "Anomaly detection using k-means and long-short term memory for predictive maintenance of large-scale solar (lss) photovoltaic plant," *Energy Reports*, vol. 9, pp. 154–158, 2023.

[13] G. W. Corder and D. I. Foreman, *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons, 2014.