Research Article

Machine Learning-Based Relative Performance Analysis of Monocrystalline and Polycrystalline Grid-Tied PV Systems

Asfand Yar (1), Muhammad Yousaf Arshad (2,3), Faran Asghar (4), Waseem Amjad (1), Furqan Asghar (1), Muhammad Imtiaz Hussain (5), Gwi Hyun Lee (6), and Faisal Mahmood (1)

1Department of Energy Systems Engineering, University of Agriculture, Faisalabad, Pakistan
2Sitara Chemical Industries Pvt. Ltd., Faisalabad, Pakistan
3Department of Chemical Engineering, University of Engineering and Technology (U.E.T) Lahore, Pakistan
4School of Information Management, Nanjing University, Nanjing, China
5Agriculture and Life Sciences Research Institute Kangwon National University, Chuncheon 24341, Republic of Korea
6Interdisciplinary Program in Smart Agriculture, Kangwon National University, Chuncheon 24341, Republic of Korea

Correspondence should be addressed to Gwi Hyun Lee; ghlee@kangwon.ac.kr and Faisal Mahmood; faisal.mahmood@uaf.edu.pk

Received 1 March 2022; Revised 17 June 2022; Accepted 11 July 2022; Published 12 August 2022

Academic Editor: Alberto Álvarez-Gallegos

Copyright © 2022 Asfand Yar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this research study, the design and performance evaluation of grid-tied photovoltaic systems has been carried out through experimentation, HelioScope simulation, and black-box machine learning methods for data-driven artificial intelligence system performance assessment and validation. The proposed systems are based on 15 kWp of monoperk and polyperk, which are separately installed in the industrial sector of Faisalabad, Pakistan. The experimental evaluation of the installed PV modules was performed from November 2020 to August 2021. The performance of the PV modules was evaluated by determining the annual average daily final yield (If), performance ratio (PR), and capacity factor (CF). The study showed that the annual average of daily final yield, performance ratio, and capacity factor for 15 kW polyperk was estimated to be 61.94 kWh, 84.17%, and 19.12, respectively. The annual average of daily final yield, performance ratio, and capacity factor for 15 kW monoperk was estimated to be 58.32 kWh, 81.42%, and 18.13, respectively. A comparison of final yield is obtained from simulation and real-time systems obtained from polyperk PV and monoperk. A significant mean error exists between the experimentation and simulation results which lie within the range of 1250 to 1470 kWh and 1600 to 1950 kWh, respectively. Substantial differences between both aforementioned results were initially tested and highlighted by statistical values; i.e., the standard error lies in-between 5 and 45% in polyperk crystalline and 5 and 25% in monocrystalline PV grid-connected module. Machine learning logistical regression evaluated that monoperk crystalline grid-connected system, experimental work was found to be more reliable with error difference reduces in off-peak months as compared to corresponding simulation study and vice versa for polyperk crystalline grid-connected system. Model accuracy after training and testing produced resulted up to 99.5% accuracy for either grid-connected experimentation or simulation outcomes with validation.

1. Introduction

In any country, energy is one of the key factors for the smooth and faster upgradation of the socioeconomic activities, and it has become a universal fact now. One of the major drivers of economic growth is energy as identified by the government of Pakistan. Therefore, Pakistan requires adequate supplies of energy for the generation of healthy economic activities. The country currently is passing through the worst energy crisis in over 70 years. The energy crisis has led to the hindrance of socioeconomic progress below the level of critical sustainability and tolerance of the people. Pakistan has a land area of 881,913 km² consisting of a population of 211.117 million [1]. In 2019, Pakistan’s electricity
generation was recorded to be 87.3 TWh [1]. The energy mix of Pakistan’s energy scenario depicts that thermal energy is the widest energy resource upon which Pakistan has been relying in recent years. To warrant adequate, secure, and cheap energy supply to industries, transportation, domestic needs, and agriculture availing available resources in such a manner that it minimizes its losses should be the main objectives of the energy sector. For the proper growth of renewable energy sectors, they must be given the chance in the present energy mix.

A photovoltaic system is not a new phenomenon as a plethora of literature is available on the topic. It has been in practice for the past few decades in our everyday life producing renewable energy, which is not only efficient, but it is carbon emission-free which is beneficial for the environment. Today, our technology has shifted to renewable resources, such as biomass, wind, and solar. In the development of countries, renewable energy plays a contentious role. Nowadays, the resource in trend is solar power. There are various kinds of photovoltaic (PV) power plants that can be integrated with existing power systems. PV energy plants generally use various types of PV engineering such as monocrystalline, polycrystalline, and amorphous silicon. They come as thin-film PV panels using copper indium di-selenide, copper indium gallium selenide, and cadmium telluride.

Various studies have evaluated the performance parameters of installed PV power plants in different locations, with the varying climatic condition, such as Europe [2–5], tropical regions [6, 7], Malaysia [8, 9], Oman [10], and Asian countries, India, Korea [11–15], and China [16–22]. It was observed that the performance analysis is affected by system losses, module quality, inverters, shading effect, losses in wiring, array tilt angle, and type of grid connections. The effect of these factors was confirmed by literature where a 190 kWp grid-connected PV power plant in India was evaluated [16]. In another study, it was found that these factors played a crucial role in reducing the annual average daily final yield. The efficiency of the PV module, inverter, and system was found to be 14.9%, 89.2%, and 12.6%, respectively. The capacity factor and performance ratio were found to be 10.1% and 81.5%, respectively [2]. The impact of dust is one of the significant factors which reduces the efficiency of PV modules. As compared to previously mentioned factors, the effect of dust is inferior to the performance of the PV systems [20]. Research about the performance evaluation of PV systems installed in Serbia was carried out to investigate the effects of ambient temperature and compare it with similar studies around the world. The study reports an analysis of 2 kW monocrystalline silicon (mc-Si) PV power plant tied to the grid, which was built on the rooftop in Niš, Republic of Serbia. The values of PR and CF of the reported PV power plants were articulated to be 93.6% and 12.88%, respectively [23]. Performance evaluation of polycrystalline silicon- (p-Si-) based PV systems was carried out at various locations, namely, Singapore, Turkey, and Greece, with installed capacities of 142.5 kWp, 2.73 kWp, and 171.36 kWp, respectively. From these studies, it can be observed that local varying parameters such as CF, PR, and final yield affect the performance of PV systems. PV systems have also been installed to meet the energy demands of local office buildings in Singapore. The first zero-energy office building used a 142.5 kWp p-Si grid-connected system to meet its energy demands by using inverter factors such as efficiency and losses of heat. The first performance evaluation of the PV system was carried out over 18 months, which showed a good overall PR of 81% [24]. A similar study was published where performance analysis of a 2.73 kWp mc-Si PV power station located in the Muglia climatic conditions in Turkey was performed wherein winter thy efficiency of mono PV module is increased. According to a study carried out in Greece, a performance analysis of a fully monitored 171.36 kWp p-Si grid-connected PV system was done on an hourly, daily, and monthly basis [4].

The claim regarding the change in the tilt angle of PV modules and the change in efficiency can be supported by a report coming from Italy. This article focused on the influence of climatic conditions on a 960 kWp mc-Si-based PV power plant. The installed systems were divided into two subfields with varying tilt angles and nominal powers (i.e., 606.6 kWp and 353.3 kWp). The results revealed that the performance ratio varied between 79% and 86.5% from March to October 2012 [24]. Another study from Italy (northern), articulated the functioning of two grid-tied PV systems (i.e., 17.94 kWp and 15.9 kWp) with the same factories but possessing varying PV technology, their rated power, and overall efficiency. The performance ratio was observed, over a year, to be 89.1% for the first mc-Si wafer-based PV system and 82.7% for the second mc-Si-based PV system [25].

Although PV systems have been around for quite some time but there is no articulated study from this region that evaluates the performance of installed systems according to the region’s environmental conditions, to the best of our knowledge, there has not been a study conducted that takes advantage of machine learning algorithms for the difference in performance evaluation between installed and simulated PV systems. This study has been carried out to evaluate the performance of installed PV systems located in the outskirts of Faisalabad on the same rooftop of an industry. The objectives of this study were to investigate the difference between power obtained from installed PV systems and via simulation software. Further, a machine learning algorithm was applied to analyze the performance of installed systems and via simulation software. Further, a machine learning algorithm was introduced. Machine learning algorithm contains the error to a minimum and provides useful insight, corelation, and evaluation with class distinction according to subjective weather and time horizon.

2. Designing of Solar Photovoltaic System

The PV systems are located at a latitude of 31.47° N, a longitude of 73.22° E, and an altitude of 186 m above sea level in the outskirts of Faisalabad city in Pakistan (Figure 1).
The monthly average daily values of solar radiation for the region range from 5.5 to 5.8 kWhm\(^{-2}\). Figure 2(a) displays the aerial view of the installed PV systems. The PV systems are installed on the rooftop of a switchgear manufacturing firm, where the top row is designed with monoperk PV modules, and the lower row is designed with polyperk PV modules as depicted in Figure 2(b). These PV plates are supported by a steel stand having tilt angles of 30°. To mitigate the effects of dust accumulation, it is necessary to regularly clean the plates at least once a week [12].
2.1. Types of PV Modules. The capacity of the installed grid-tied PV system is 15 kWp which is composed of two independent monocrystalline and polycrystalline segments. Each segment is installed in a series parallel and has a capacity of 15 kWp. The systems are installed facing south having an optimum monthly tilt angle to enhance their efficiency. In this study, thirty-eight monocrystalline modules (model CS3400WP) and thirty-four polycrystalline modules (model CS3400WP) of similar maximum power output and efficiency ratings are used. Properties of the installed modules are summarized in Table 1. There are four mounting structures of galvanized steel frames holding six panels for each type of PV module. The design of the support structure is to keep a variable inclination with the horizontal plane with the reinforced concrete foundation to withstand the weight of modules and maximum wind speeds at the site as shown in Figure 2(b).

The following formulae can be used for calculating parameters related to the designing of a PV system. Equations (1) and (2) can be used to calculate the required numbers of PV panels.

\[ PV_{\text{Max}} = \frac{Le \times Ip}{H_{\text{Avg}} \times \eta \times \eta_{f}}, \]  

\[ PV_{N} = \frac{PV_{\text{Max}}}{P_{m}}, \]  

where \( PV_{\text{Max}} \) is the peak power of PV array (kWp), \( L_e \) is the electrical load (kWh/day), \( I_p \) is the peak solar intensity, \( H_{\text{Avg}} \) is the average available radiation in kWh/m²/d, and \( T_{\text{cf}} \) is the temperature correction factor (0.4-0.5% °C for crystalline silicon) and the efficiency of inverter.

\[ P_{m} = \frac{V_{M} \times I_{M}}{T_{\text{cf}}}, \]  

where \( P_{m} \) is the rated power of the selected panel and \( PV_{N} \) is the number of PV panels.

\[ \text{Size of inverter} = \text{Total Load} + \frac{(1 + A_{f})}{I_{V}}. \]  

where \( A_{f} \) is the additional further load expansion (20%), \( I_{V} \) is the efficiency of inverter, and \( V_{A} \) is the volt-ampere.

3.2. Reference Yield. Referenced yield can be defined as the total irradiance \( (H_{t}) \) divided by irradiance \( (G_{\text{STC}}) \), which, primarily, is 1 kW/m² (Equation (7)). This constitutes the solar energy that is available for a specifically mentioned time frame at any place where the PV power system has been deployed. It constitutes the hours per day that are necessary to obtain solar radiation to be at reference irradiance level to produce the same incident energy as received by the sun [25].

\[ \text{Reference yield} = \frac{H_{t}}{G_{\text{STC}}}, \]  

where \( H_{t} \) is the total irradiance and \( G_{\text{STC}} \) is the irradiance.

3.3. Performance Ratio. Performance ratio (PR) is a significant parameter that depends on the PV power plant’s geographical location. It is represented by the ratio of energy outputs obtained actually and theoretically over the time of a month or a year as depicted by Equation (8). A high PR value for a grid-connected PV is an indication that the system is efficient and reliable. The overall losses from various
factors such as inverter losses, module mismatch, wiring, temperature, and dust or snow on the rated output power are less. The overall efficiency of the solar photovoltaic power system is closely related to the value of the performance ratio, i.e., closer to 100% [25].

\[
\text{Performance ratio} = \frac{Y_f}{Y_r}, \tag{8}
\]

where \(Y_f\) is the actual energy output and \(Y_r\) is the theoretical energy output.

3.4. Capacity Factor. The capacity factor is mainly the ratio of produced electricity capacity factor. Geothermal energy and biomass among the clean and renewable energies perform at higher capacity factors compared to intermittent renewable sources which tend to have lower capacity factors. For PV plants, the capacity factor ranges from 10% to 30%, as their outputs fluctuate resulting from weather conditions [25]. It can be calculated by using

\[
\text{Capacity factor} = \frac{Y_f}{\text{hours}}, \tag{9}
\]

where \(Y_f\) is the actual energy output.

3.5. Meteorological Data. To examine the performance of the PV system, it is required to study the weather data recorded by the weather station. This may include data set comprising temperature and wind speed. The ambient and module temperature fluctuated during the nominated period from November 2020 to August 2021 individually. Undesirable module temperature was usually observed during nighttime. During the measured period, wind velocity varied from 1 m s\(^{-1}\) in January and 2 m s\(^{-1}\) in June on average, whereas relative humidity reached almost 87% level during the monsoon season (i.e., July and August), and the maximum value of global-tilt solar irradiation of 1225 W\(\text{m}^{-2}\) was recorded for Faisalabad. It has been observed that on any sunny day, the module performance is affected by three meteorological parameters. These factors include temperature, solar irradiation, and wind speed, whereas other parameters are shown in Table 3.

### Table 1: Electrical and mechanical parameters of the installed PV modules.

| Parameter                                | mc-Si     | p-Si      |
|------------------------------------------|-----------|-----------|
| Maximum power (\(P_{\text{max}}\))       | 400 Wp    | 440 Wp    |
| Module efficiency (\(\eta\))             | 18.11%    | 19.7%     |
| Maximum power point voltage (\(V_{\text{mp}}\)) | 38.7 V    | 40.3 V    |
| Current at maximum point (\(I_{\text{mp}}\)) | 10.34 A   | 10.92 A   |
| Open-circuit voltage (\(V_{oc}\))        | 47.2 V    | 48.7 V    |
| Short circuit current (\(I_{sc}\))        | 10.9 A    | 11.4 A    |
| Module temperature at NOCT (\(T_{\text{NOCT}}\)) | 42\(\pm\)3°C | 41\(\pm\)3°C |
| Temp. coefficient of short circuit current (\(\mu\)) | 0.05%/°C | 0.05%/°C |
| Temp. coefficient of open-circuit voltage | -0.29%/°C | -0.28%/°C |
| Temp. coefficient of maximum power         | -0.37%/°C | -0.36%/°C |

### Table 2: Electrical datasheet of three-phase inverters.

| Parameter                                | Value     |
|------------------------------------------|-----------|
| Input (DC)                               |           |
| Max. Dc power (\(\cos \theta = 1\))     | 15 KV     |
| Max. input voltage                       | 1000 V    |
| MPP voltage range/rated input voltage    | 200-800 V |
| Min. input voltage/initial input voltage | 350 V     |
| Max. input current input A/input B       | 18 A+18 A |
| Output (AC)                              |           |
| Rated power (230 V, 50 Hz)               | 15 kW     |
| Max. apparent AC power                   | 33 kVA    |
| Rated power frequency                    | 50/60 Hz  |
| Max. output current                      | 21.7 A    |
| Power factor at rated power              | 0.8 (min.)|
| Maximum efficiency                       | 97.5%     |
| General data                             |           |
| Weight                                   | 30 KG     |
| Operating temperature range              | -26°C–60°C|
| Noise emission (typical)                 | <30 dBA   |
| Self-consumption (night)                 | <1 watt (night) |
| Topology                                 | Transformer less |
| Cooling concept                          | Natural convection |

Table 2 shows the electrical and mechanical parameters of the installed PV modules.
Table 3: Meteorological data for Faisalabad.

| Parameter measured          | Value obtained          |
|----------------------------|-------------------------|
| Measuring range            | 0.0-1400.0 Wm⁻²         |
| Resolution                 | <1 Wm⁻²                 |
| Spectral range             | 300-2800nm              |
| Sensitivity range          | 5-20 μV/Wm⁻²            |
| Operating temperature rate | -40°C to +80°C          |
| Maximum operational irradiance | 2000 Wm⁻²              |
| Relative humidity range    | 4.35-17.4 Psi           |
| Relative humidity accuracy | ±2%RH                   |
| Air temperature range      | -50°C to +60°C          |
| Air temperature accuracy   | 0.5°C                   |
| Wind speed range           | 0.8-40 m⁻²              |
| Wind speed measuring accuracy | ±0.5%                  |

4. Performance Analysis and Result Discussion

This comparative study visualizes that in winter, the efficiency of monoperk is better as compared to other types of modules. In summer, polyperk performed with better efficiency because it is highly efficient in the high-temperature range. The maximum production of kWh in winters was more than 2,100 kWh for polyperk. The data obtained displayed that the annual average of daily final yield, performance ratio, and capacity factor for 15 kW polyperk was estimated to be 61.94 kWh, 84.17%, and 19.12, respectively. The annual average of daily final yield, performance ratio, and capacity factor for 15 kW monoperk was estimated to be 58.32 kWh, 81.42%, and 18.13, respectively. The reported results are comparatively better than a similar study conducted in Iran [26].

On the other hand, a maximum yield of around 1,700 kWh was obtained from monoperk as summer ambient temperature is high leading to reduced efficiency. At the start of November, the ambient temperature began to decrease, and thus, the efficiency of monoperk also increased, whereas the efficiency of polyperk started decreasing till December as the highest temperature of the season was monitored this month. In March, the yield of monoperk plummeted, whereas for the polyperk modules, the yield improved. In April, May, June, and July, the ambient temperature stayed at a peak level. Therefore, the efficiency of polyperk increased, and consequently, the efficiency of monoperk decreased. In August, temperature variation was observed due to the monsoon season. On warm days, the efficiency of polyperk increased whereas, on the other hand, the efficiency of monoperk increased on relatively cool days as depicted by simulation results in Figure 3(b). So, it can be concluded that polyperk modules outperform monoperk type modules when the ambient temperature increases. The common material used in solar cells, crystalline silicon, does not help to prevent them from getting hot either. As a great conductor of heat, silicon actually speeds up the heat building up in solar cells on hot summer days. Mono PV modules are made from a single crystal of silicon whereas poly is made from several fragments of silicon melted together which prevents it from heating up.

As it can be seen in Figure 3(a), HelioScope-based system generated higher power in comparison to a real-time system as the software considers ideal conditions and ignores various factors that affect the real system output performance such as dust losses and shadow factor.

Scatter plot in Figure 4 displays two hotspots and distinct regions highlighted by two circles. The upper-end circle has a wide range of experimental mono- and polycrystalline systems and polyperk simulation yield results. The lower circle has a keen distribution of yield results in experimental monocrystalline systems and polyperk simulated systems. Only peculiar results in raw data for the distinctive monocrystalline solar energy material produce oblique and relatively lesser in midseason.

Quantified data and behavior in tail line circles need to be categorized using learning, training, and testing data sets for relative comparison, which is a new theme in solar energy system classification; further details are defined in Section 4.

Ideal conditions and real-time working systems classify two different data sets and elucidated graphs. Statistical graphs show the relatively higher difference in total yield over the time scale length in real time due to undistinguished and random errors in calculations and measurement. Figure 4 shows the hotspot regions using the estimated data varying from lower tail to high end. Figures 5 and 6 depict raw data sets in peak slope form. Figures 4 and 5 show distinct groups and variations in estimated and calculated values. In polyperk system experimentation, Figure 6 shows the error doping in mean values for polyperk crystalline modules (simulation and experiment result comparison). A greater mean error exists in the experimentation results with the range lying in 1250 to 1470 kWh in comparison to simulation results with the least mean error in the upper bound of 1950 kWh. Significance of new advance predictive and artificial intelligence tools in the realm of statistical-based machine learning produces a counterproductive approach using both real-time and ideal simulation patterns and data sets for classification and decision-making for applied applications while using monoperk and polyperk solar systems. The apparent results are statistically plotted for heat mapping to understand the relationship between the distinctive classes of monoperk and polyperk crystalline modules of simulation and experimental data sets.

Figures 7 and 8 display futuristic parameters that are independent and dependent in nature on each other. Plasma-type heat map of monoperk crystalline modules shows maximum nonlinearity in the experimental and simulation results up to -0.86. The corresponding increase in module yield is inferred from the heat map but linearity is minimum in experimental and simulation results with values as low as 0.14 and 0.16. In polyperk crystalline modules, greater amount of synergy is accounted for in the experimental and simulation results. The corresponding results are accounted for and validated with heat map nonlinearity values greater than monoperk crystalline modules up to ~0.98. However, in relationship with months and
method classes, a reverse effect has been observed with negative values of as low as -0.061 and -0.96 for experimental and simulation data sets.

Heat maps in Figures 7 and 8 depict the statistical difference with nonlinearity and class distinction. An efficient technique must be applied for validating results, consolidating decisions for which type of crystalline modules and which method needs to be implemented for desirable calculation of yield measurement according to our requirements. Machine learning algorithms are statistically powerful
Figure 4: Comparison of hotspots generated by mono- and polyperrk on monthly basis using real-time installed systems and simulations.

Figure 5: Peak form data of experimental and simulation results for polycrystalline material energy systems (1, simulation; 0, experimentation).
algorithms that over the period of using artificial intelligence learning update their sets and make the logical decision of either using experimental means or simulation for time and cost-saving.

4.1. Machine Learning-Based Relative Performance Investigation of Mono- and Polyperk Modules. As previously defined, the nonlinearity and class distinction in the available raw data sets are classically classified into the categorical division between experimental and simulation results. Parametric results of machine learning-based logistical regression have been introduced as a strategic machine learning algorithm due to its robust nature and easiness with mixed data set performance. Using this technique, the algorithm will
| Method (1-Simulation, 0-Experimentation) | Month | Experimentation results | Helio-scope simulation results |
|----------------------------------------|-------|-------------------------|-------------------------------|
|                                        | 1     | -0.095                  | -0.16                         |
|                                        | -0.061| 1                       | 0.98                          |
|                                        | -0.16 | 0.98                    | 1                             |
|                                        | 0.14  | -0.96                   | -0.96                         |

**Figure 8:** Heat map between each independent and dependent variable in the case of polyperk crystalline modules.

**Figure 9:** Machine learning algorithms and flow chart.
evaluate the reasons behind the performance gap between simulation and real-time system outputs as shown in Figure 3 and help to contain the error to a minimum; it also provides useful insight, correlation, and evaluation with class distinction according to subjective weather and time horizon.

4.1.1. Machine Learning-Based Logistic Regression Modeling. Machine learning-based regressions are keenly a new method for inference of error and hyperparameter deduction. For efficient future predictive system installation, result synchronization, and error reduction between the simulation model and real-time system, machine learning-based logistic regression modeling has been performed in this section. ML classification has been done on the following three algorithms:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

Flowchart classification of machine classes, input and output data set nature, and algorithms are presented in Figure 9.

The categorical data in ML lies mainly in the domain of logistic regression. Linear and polynomial regressions are commonly used for conventional regressive weight...
parameters and their correlations based on $p$ values. ML differentiates logistical regression from linear regression based on the continuous and descriptive data set, respectively. The regular fit in linear regression mainly highlights continuous values in a data set. Logistic regression predicts whether something is either true or false, yes or no, and 1 or 0, instead of predicting the weight or performance based on the mean absolute, $R^2$, etc., values. Due to

Figure 11: Pair plot of polyperk module data sets in grid-connected formation for correlation between each independent and dependent variable.
significant differences between the simulation results and experimental results performed earlier and as highlighted by statistical values, i.e., the standard error lies between 5 and 45% for polyperk crystalline and 5 and 25% for Monocrystalline PV grid-connected module. ML-based logistic regression modeling has been carried out. A pair plot drawn in Python using the Seaborn library assimilates the finest set of attributes and features for a relative explanation of relationships between two and more variables. The plot displays clusters and separation of offset data points along with hyperparameters that results in decision-making and classification of either supervised or unsupervised algorithm for machine learning. Pair plots in Figures 10 and 11 show class representations in monoperk and polyperk module data sets with bar plots showing energy yield produced over a time series.

Monoperk crystalline grid-connected system shows the sigmoidal plot and two hotspot regions mainly featuring yield and timeline variable, i.e., month. Figures 10 and 11 also depict method validity for the experimentation and simulation results. The pair plot of monoperk has data sets more aligned statistically for simulation data and interpretations, while the polyperk features data more viable within experimentation data points. The pair plot consists of scatter plots and bar plots. Experimentation data points in the pair plot depict a higher yield for polyperk as compared to the monoperk crystalline grid-connected system.

The under discussion ML-based case study for monoperk and polyperk crystalline module-based grid-tied systems is being conducted first time as no previous literature is available regarding the AI-based performance analysis. This method highlights the significance of implications between simulated and experimental methods with error reduction, consolidation, and the performance evaluation of grid-tied PV solar modules.

Anaconda-Python 3.8 libraries including NumPy, Matplotlib, Seaborn, pandas, and scikit-learn have been utilized for data analysis and machine learning-based decision modeling between experimental and simulation results. The initial case study for the polyperk crystalline module data set shows a higher error doping in mean values during experimental evaluation in comparison to simulation results as shown in Figure 5.

Figure 5 displays the comparison based on error occurrence between simulation and experimental results in the polyperk crystalline module where class 0 represents experimental and class 1 represents simulation results. In both simulated and experimental environments, per month yield maximum output has been calculated. The descriptive statistics of the polyperk and monoperk crystalline PV module performance can be seen in Tables 4 and 5.

It can be seen from the extracted features represented in Tables 4 and 5 that there is a significant difference between simulation and experimental data. This difference is due to the ideal scenario of the software as there is no consideration of temperature variance and its effect on the system performance as well as other factors such as dust factor and shadow factor. Based on the abovementioned features, a machine learning algorithm has been used to signify the error reduction and predictability module.

4.1.2. Logistical Regression Mathematical Representation.

Logistical regression is generally used for two-class problems and can be extended for implications in the multiclass problems but becomes highly unstable in calculations when classes are well separated; the working diagram is represented in Figure 12. The baseline for the mathematical formulation of logistical regression is simple linear regression.

\[ y = b_0 + b_1 x, \] (10)

| Table 4: Statistics (extracted features) based on performance analysis of polyperk modules. |
|------------------------------------------|-----------------|-----------------|
| Features                  | Experimental values | Simulated values |
| Mean                      | 1842.98          | 1733.76         |
| Standard error            | 196.6675         | 112.8729366    |
| Median                    | 2152.85          | 1909.65         |
| Standard deviation        | 621.9171         | 356.93556559    |
| Sample variance           | 386780.9         | 127402.9982     |
| Kurtosis                  | -1.78782         | -1.835603662    |
| Skewness                  | -0.155217059     | -0.435316653    |
| Range                     | 1537.4           | 879.6           |
| Minimum                   | 900.8            | 1243            |
| Maximum                   | 2438.2           | 2122.6          |
| Sum                       | 18429.8          | 17337.6         |
| Count                     | 10               | 10              |
| Largest (1)               | 2438.2           | 2122.6          |
| Smallest (1)              | 900.8            | 1243            |
| Confidence level (95.0%)  | 444.8927         | 255.3363221     |

| Table 5: Statistics (extracted features) based on performance analysis of monoperk modules. |
|------------------------------------------|-----------------|-----------------|
| Features                  | Experimental values | Simulated values |
| Mean                      | 1265.67          | 1849.62         |
| Standard error            | 108.0271232      | 122.1181821     |
| Median                    | 1328.6           | 2037.5          |
| Standard deviation        | 341.6117584      | 386.1715992     |
| Sample variance           | 116698.5934      | 149128.504      |
| Kurtosis                  | -2.063216661     | -1.835603662    |
| Skewness                  | -0.155217059     | -0.435316653    |
| Range                     | 827.7            | 957.9           |
| Minimum                   | 834.6            | 1317.3          |
| Maximum                   | 1662.3           | 2275.2          |
| Sum                       | 12656.7          | 18496.2         |
| Count                     | 10               | 10              |
| Largest (1)               | 1662.3           | 2275.2          |
| Smallest (1)              | 834.6            | 1317.3          |
| Confidence level (95.0%)  | 244.3743305      | 276.2505204     |
where $y$ is the continuous output result while the $b_0$ is the intercept of the line, $b_1$ is the slope of the line, and $x$ is the input or independent variable. In continuous data, we can utilize linear regression, but in distinct classes, we cannot draw the best-fit curve. We introduce a classifier for whether an action/event shall occur or not. Figure 13 shows that the output in logistical regression is binomial. LR has a sigmoid function based on the probability of an event ($P$) and ranges from 0 to 1 and is given as

$$ P = \frac{1}{1 + e^{-y}}. \quad (11) $$

Introducing $P$ in the LR equation and computing at the natural log scale gives the following mathematical expression:

$$ \ln \left( \frac{P}{1 - P} \right) = b_0 + b_1 x. \quad (12) $$

Weighted output is in the range of either 0 or 1; if the output is $<0.5$, then it corresponds to 0th class if not then 1st class. For training and testing data sets after incorporation into machine learning the module, accuracy is defined by

$$ \text{Accuracy} \% = \frac{\text{Correct Predictions}}{\text{Total Predictions}} \times 100\%. \quad (13) $$

Figure 12: Working diagram of the testing and training in logistical machine learning regression modeling.
Figure 13: Sigmoid function behavior at a discrete event time scale length.

Figure 14: Working data sets of monoperk and polyperk modules using logistic regression: (a) training set of monoperk module; (b) testing set of monoperk module; (c) training set of polyperk module; (d) training set of polyperk module.
The accuracy of a system based on the prediction of classification problems is either correct or can be incorrect so a matrix, namely, a confusion matrix, is formed for a relational summary of actual and predicted values of classification problems.

4.1.3. Performance Analysis of Mono- and Polyperk Crystalline Modules.

The total reading count of simulation and experimental work is divided into 80% training and 20% testing data as shown in Figure 14. The classifier is introduced for encoding purposes to have the binary distinction between greater probability events. The timeline horizon consists of 10 months; no data was conceived for the month of September and October.

\[
\log(p) = \ln\left(\frac{p}{(1-p)(1-p)}\right),
\]

where \(p\) is the probability of a discrete event defined earlier.

The training data set selects a total of 80% values from raw data. The algorithm reads the pattern according to the problem statement and improves the probability values. In each run, a class check is run for appointing the new red and green distribution as highlighted in Figure 14. An increase in the iterations increases the width of class boundaries. Based on the extracted features and training/testing of both modules, the data set collected from simulation and experimental work of mono- and polyperk crystalline module for a one-year cycle (except September and October) can be seen in Tables 6 and 7 which depict the difference between them ranging from 1% to 60.5% in monoperk and 6% to 27% in polyperk crystalline.

In logistical regression training, the data set values determine the probable event in advance. The most likelihood and unlikelihood are distinguished for experimental and simulation results. For the monoperk crystalline grid-connected system in the experimental scenario, readings are likelihood, and the probability after training and testing the data set lies about 99% accuracy. The confusion matrix produces an absolute validated summary result for the method and material to be incorporated in each particular scenario. A 20% test data set values in most actual and predicted values are in harmony with each other using the ML module. In further classification, as shown in the heat map and pair plot in Figures 8–11, the class distinction reflects that those major values classified in the experimental work are authentic.

Table 6: Data set of monoperk crystalline module in simulation and experimental work.

| Month   | Monoperk experimental results | Monoperk simulation results | Method/class max. yield output | Difference | Error percentage |
|---------|------------------------------|----------------------------|--------------------------------|------------|-----------------|
| January | 1629.5                       | 1329                       | 0                              | 300.5      | 18.44123964     |
| February| 1591                         | 1561                       | 0                              | 30         | 1.885606537     |
| March   | 1456.5                       | 2003.5                     | 1                              | 547        | 27.3022111      |
| April   | 1200.7                       | 2110.7                     | 1                              | 910        | 43.11365898     |
| May     | 1000.2                       | 2275.2                     | 1                              | 1275       | 56.03902954     |
| June    | 899.1                        | 2254.1                     | 1                              | 1355       | 60.11268355     |
| July    | 865.5                        | 2071.5                     | 1                              | 1206       | 58.21868211     |
| August  | 834.6                        | 2111.6                     | 1                              | 1277       | 60.47546884     |
| November| 1662.3                       | 1462.3                     | 0                              | 200        | 12.03152259     |
| December| 1517.3                       | 1317.3                     | 0                              | 200        | 13.1813089      |

Table 7: Data set of polyperk crystalline module in simulation and experimental work.

| Month   | Experimental polyperk results | Polyperk simulation results | Method/class max. yield output | Difference | Error percentage |
|---------|------------------------------|----------------------------|--------------------------------|------------|-----------------|
| January | 900.8                        | 1245.8                     | 1                              | 345        | 27.69304864     |
| February| 1371.2                       | 1471.2                     | 1                              | 100        | 6.797172376     |
| March   | 2007.1                       | 1881.1                     | 0                              | 126        | 6.27714115      |
| April   | 2311.1                       | 1976.1                     | 0                              | 335        | 14.495262       |
| May     | 2298.6                       | 2122.6                     | 0                              | 176        | 7.656834595     |
| June    | 2405.6                       | 2105.6                     | 0                              | 300        | 12.47090123     |
| July    | 2438.2                       | 1938.2                     | 0                              | 500        | 20.50693134     |
| August  | 2377.5                       | 1977.5                     | 0                              | 400        | 16.82439537     |
| November| 1276.5                       | 1376.5                     | 1                              | 100        | 7.264802034     |
| December| 1043.2                       | 1243                       | 1                              | 199.8      | 16.07401448     |
evaluation results are likely the same and coherent. In the case of polyperk crystalline grid-connected studies, a greater significance and coherence lies within the simulation and experimental studies, and the likelihood of probability is significant as seen in Tables 6 and 7. Model accuracy after training and testing produces results of up to 99.5%. Simulation results are likely most favorable in predicting the performance evaluation of maximum yield in grid-connected systems. During off-peak (i.e., winter season), the error difference increases due to various factors such as module angle, line losses, cell dimension, and temperature invariability.

5. Conclusions

In this research, a comparative study was conducted for two solar systems of 15 kW. The systems primarily included monoperk and polyperk crystalline. From the study, it was observed that maximum yield was produced, in winters, by monoperk, while during summers, polyperk displayed enhanced efficiency in Faisalabad, polyperk has proven to be efficient considering the high rates of temperature whereas monoperk, consequently, reduced the efficiency of the solar system. The annual average of daily final yield, performance ratio, and capacity factor for 15 kW polyperk was estimated to be 61.94 kWh, 84.17%, and 19.12, respectively. The annual average of daily final yield, performance ratio, and capacity factor for 15 kW monoperk was estimated to be 58.32 kWh, 81.42%, and 18.13, respectively. The efficiency can also be attributed to the quality of the system. For example, good quality inverter requires quality plates for enhanced efficiency. Additionally, the production of a 430-watt panel was noted as 400 watts. Furthermore, according to the applied machine learning module, a significant difference was observed in the considered PV module types (i.e., mono- and polyperk). A machine learning modeling study using logistical regression was applied to determine the greater significance of experimentation results in monoperk crystalline with an accuracy of 99.5%, while the results in polyperk using simulation studies are more accurate and recommended in the evaluation of PV-connected grid. Depending upon observed parameters and frequency period, the model elucidated a better understanding of the performed real-time analysis leading to both cost and time saving for the installation of similar projects in the region. Furthermore, better performance can be achieved from installed systems if factors such as dust, tilt angle, and shadow effects are considered before the installation of PV systems.

Data Availability

Data will be available on request. For data-related queries, kindly contact Faisal Mahmood (faisal.mahmood@uaf.edu.pk).

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this research article.

Acknowledgments

The authors acknowledge the support of the Department of Energy Systems Engineering and Punjab Bio-Energy Institute (PBI), University of Agriculture, Faisalabad, for facilitating data collection. This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through the Technology Commercialization Support Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (No. 821048-3).

References

[1] Government of Pakistan Finance Division, “Pakistan economic survey,” 2019, https://www.finance.gov.pk/survey_1920.html.
[2] M. S. Adaramola and E. E. T. Vågnes, “Preliminary assessment of a small-scale rooftop PV-grid tied in Norwegian climatic conditions,” Energy Conversion and Management, vol. 90, no. 2015, pp. 458–465, 2015.
[3] L. M. Ayompe, A. Duffy, S. J. McCormack, and M. Conlon, “Measured performance of a 1.72kW rooftop grid connected photovoltaic system in Ireland,” Energy Conversion and Management, vol. 52, no. 2, pp. 816–825, 2011.
[4] G. C. Bakos, “Distributed power generation: a case study of small scale PV power plant in Greece,” Applied Energy, vol. 86, no. 9, pp. 1757–1766, 2009.
[5] D. Chemisana and C. Lamanotou, “Photovoltaic-green roofs: an experimental evaluation of system performance,” Applied Energy, vol. 119, pp. 246–256, 2014.
[6] P. Ferrada, F. Araya, A. Marzo, and E. Fuentealba, “Performance analysis of photovoltaic systems of two different technologies in a coastal desert climate zone of Chile,” Solar Energy, vol. 114, pp. 356–363, 2015.
[7] T. Khatib, K. Sopian, and H. A. Kazem, “Actual performance and characteristic of a grid connected photovoltaic power system in the tropics: a short term evaluation,” Energy Conversion and Management, vol. 71, pp. 115–119, 2013.
[8] A. H. A. Al-Waeli, M. T. Chaichan, H. A. Kazem, K. Sopian, and J. Safaei, “Numerical study on the effect of operating nanofluids of photovoltaic thermal system (PV/T) on the convective heat transfer,” Case Studies in Thermal Engineering, vol. 12, pp. 405–413, 2018.
[9] A. H. Al-Waeli, H. A. Kazem, K. Sopian, and M. T. Chaichan, “Techno-economical assessment of grid connected PV/T using nanoparticles and water as base-fluid systems in Malaysia,” International Journal of Sustainable Energy, vol. 37, no. 6, pp. 558–575, 2018.
[10] H. A. Kazem, J. Yousif, M. T. Chaichan, and A. H. Al-Waeli, “Experimental and deep learning artificial neural network approach for evaluating grid-connected photovoltaic systems,” International Journal of Energy Research, vol. 43, no. 14, pp. 8572–8591, 2019.
[11] J. Y. Kim, G. Y. Jeon, and W. H. Hong, “The performance and economical analysis of grid-connected photovoltaic systems in Daegu, Korea,” Applied Energy, vol. 86, no. 2, pp. 265–272, 2009.
[12] K. A. Kumar, K. Sundareswaran, and P. R. Venkateswaran, “Performance study on a grid connected 20 kW p solar photovoltaic installation in an industry in Tiruchirappalli (India),”
Energy for Sustainable Development, vol. 23, pp. 294–304, 2014.

[13] D. H. W. Li, K. L. Cheung, T. N. T. Lam, and W. W. H. Chan, “A study of grid-connected photovoltaic (PV) system in Hong Kong,” Applied Energy, vol. 90, no. 1, pp. 122–127, 2012.

[14] K. Padmavathi and S. A. Daniel, “Performance analysis of a 3MWp grid connected solar photovoltaic power plant in India,” Energy for Sustainable Development, vol. 17, no. 6, pp. 615–625, 2013.

[15] V. Sharma, A. Kumar, O. S. Sastry, and S. S. Chandel, “Performance assessment of different solar photovoltaic technologies under similar outdoor conditions,” Energy, vol. 58, pp. 511–518, 2013.

[16] V. Sharma and S. S. Chandel, “Performance analysis of a 190 kWp grid interactive solar photovoltaic power plant in India,” Energy, vol. 55, pp. 476–485, 2013.

[17] Y. Su, L. C. Chan, L. Shu, and K. L. Tsui, “Real-time prediction models for output power and efficiency of grid-connected solar photovoltaic systems,” Applied Energy, vol. 93, pp. 319–326, 2012.

[18] X. Wu, Y. Liu, J. Xu et al., “Monitoring the performance of the building attached photovoltaic (BAPV) system in Shanghai,” Energy and Buildings, vol. 88, pp. 174–182, 2015.

[19] W. Zhou, H. Yang, and Z. Fang, “A novel model for photovoltaic array performance prediction,” Applied Energy, vol. 84, no. 12, pp. 1187–1198, 2007.

[20] M. Mani and R. Pillai, “Impact of dust on solar photovoltaic (PV) performance: research status, challenges and recommendations,” Renewable and Sustainable Energy Reviews, vol. 14, no. 9, pp. 3124–3131, 2010.

[21] A. Y. Appiah, X. Zhang, B. B. K. Ayawli, and F. Kyeremeh, “Review and performance evaluation of photovoltaic array fault detection and diagnosis techniques,” International Journal of Photoenergy, vol. 2019, Article ID 6953530, 19 pages, 2019.

[22] C. Li, “Comparative performance analysis of grid-connected PV power systems with different PV technologies in the hot summer and cold winter zone,” International Journal of Photoenergy, vol. 2018, Article ID 8307563, 9 pages, 2018.

[23] D. D. Milosavljević, T. M. Pavlović, and D. S. Piršl, “Performance analysis of a grid-connected solar PV plant in Niš, republic of Serbia,” Renewable and Sustainable Energy Reviews, vol. 44, pp. 423–435, 2015.

[24] P. M. Congedo, M. Malvoni, M. Mele, and M. G. De Giorgi, “Performance measurements of monocrystalline silicon PV modules in South-Eastern Italy,” Energy Conversion and Management, vol. 68, pp. 1–10, 2013.

[25] D. Micheli, S. Alessandrini, R. Radu, and I. Casula, “Analysis of the outdoor performance and efficiency of two grid connected photovoltaic systems in northern Italy,” Energy Conversion and Management, vol. 80, pp. 436–445, 2014.

[26] S. Edalati, M. Ameri, and M. Iranmanesh, “Comparative performance investigation of mono- and poly-crystalline silicon photovoltaic modules for use in grid-connected photovoltaic systems in dry climates,” Applied Energy, vol. 160, pp. 255–265, 2015.