

A novel multi-layer weather classification-based regression model for PV output power forecasting

Zeina Bahij¹, Bing Yan¹, Katie McConky²

 **Cite this article:** Bahij Z, Yan B, McConky K. *Cybernetics and Intelligence* 2026, 1(2): 9390010. <https://doi.org/10.26599/CAI.2025.9390010>

ABSTRACT: The intermittent nature of photovoltaic (PV) power poses significant challenges to grid operations. However, accurate forecasting is difficult since PV power depends on weather conditions that fluctuate within minutes. To address this, a novel multi-layer weather classification-based regression model (MWCR) is developed to predict the PV output power for the coming 24 h in different datasets with 1-h and 5-min as time steps. The key idea is to classify weather conditions at each time step to capture the rapid PV power fluctuations. To address the absence of irradiance sensors, the power ratio of predicted power obtained by a long-short term memory (LSTM) model over the maximum power in a clear sky is used to reflect sky condition. Classification is based on power ratio, local irradiance, temperature, wind speed, and humidity. Then, for each class, a regression model with the same inputs is developed. To improve the accuracy of the model, an LSTM model predicting power is integrated as an additional input to the regression model. Prediction results of three datasets at 1-h and 5-min time steps demonstrate superior accuracy for both models compared to an LSTM model alone in all testing datasets, especially on a day with high PV power fluctuations.

KEYWORDS: photovoltaic (PV) power forecasting; long-short term memory (LSTM) power; classification-based regression; multi-layer weather classification-based regression (MWCR)

1 Introduction

The penetration of green energy resources is becoming more crucial these days than ever, due to the rapid increase in global average temperature caused by burning fossil fuels [1]. In recent years, countries are investing more in renewable energy since it is identified as a replenishing resource that produces zero emissions [2]. PV energy is one of the most widely used green energy resources, with a worldwide capacity of 1100 GW by the end of 2024, which was sufficient to satisfy 2.1% of total energy consumption worldwide [3]. The main problem associated with PV energy is its intermittent nature. The generation of power from a PV plant depends on the solar radiation at a specific location and time [4]. PV power depends on highly variable weather conditions such as cloud cover, humidity, and temperature [5]. These meteorological factors show high variations within minutes resulting in rapid PV power fluctuations. Different solutions have been utilized to address the intermittent nature of PV output power, such as power forecasting to allow for advanced preparation for upcoming fluctuations.

Different PV power forecasting techniques have been developed as reviewed in Section 2. These techniques are classified as physical and data-driven. For physical approaches [6–9], prior knowledge of PV physical parameters along with climatological factors are needed. However, these parameters may not be available. Data-driven techniques are independent of physical

parameters [10], and some of them can derive PV power from historical data without considering meteorological factors [11]. However, due to the intermittent nature of PV, the power-to-power forecasting approaches cannot catch the rapid fluctuations of weather conditions. To capture the relationship between weather conditions and PV output, regression models were developed [12, 13]. Their forecasting accuracy is high on sunny days, but low on days with varying weather conditions.

In this paper, a multi-layer weather classification-based regression (MWCR) model is developed to forecast 24 h ahead PV power at 1-h and 5-min time steps in Section 3. The key idea is to classify weather conditions based on irradiance, temperature, humidity, and wind speed at each time step to capture the rapid fluctuations in weather conditions. Due to the absence of irradiance sensors, the publicly available local irradiance based on clear sky cannot reflect the actual sky condition. To address this, the ratio of predicted power obtained by a long-short term memory (LSTM) model over maximum power calculated by the physical model is used to replace local irradiance to account for the sky condition. By using the k-nearest neighbor, weather conditions are classified into clear sky, partly cloudy, mostly cloudy, and cloudy. Then, for each weather class, a unique regression model is developed by the general regression neural network (GRNN). Climatological factors used for regression are local irradiance in clear sky, temperature, humidity, and wind speed along with the power ratio. As LSTM model power-to-power prediction gives an estimate of the PV power, it is also

¹ Department of Electrical and Microelectronic Engineering, Rochester Institute of Technology, Rochester 14623, USA. ² Department of Industrial and System Engineering, Rochester Institute of Technology, Rochester 14623, USA.

 Corresponding author. E-mail: zb6781@rit.edu

Received: September 19, 2024; Revised: January 27, 2025; Accepted: April 29, 2025

© The Author(s) 2026. This is an open access article under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0, <http://creativecommons.org/licenses/by/4.0/>).

Nomenclature	
I_0	Solar constant
I_C	Irradiance during the measurement
I_{STC}	Standard test condition irradiance
k	Power temperature coefficient
L	City latitude
n	Number of the day in a year
P_{PV}	Maximum PV output power
$P_{PV,max}$	Maximum PV power in clear sky
P_{STC}	Power at the standard test condition
T_C	Temperature at the operating point
T_{STC}	Temperature at the standard test condition
δ	Declination angle of the sun
ω	Hour angle

added as input to the regression model to develop the enhanced model.

The above models are tested using PV power data collected from a 2-MW solar farm located at Rochester Institute of Technology in the USA for 3.5 years with 5 min as a time step as discussed in Section 4. Three different weather datasets are used for testing. The first dataset is to test the model performance in a real-case scenario based on 1-h historical actual and historical forecasted weather data available for 3.5 years. The second one is to test the model behavior to sudden PV fluctuations based on actual historical 5-min weather data available for 3.5 years. The third one is to test the model behavior to sudden PV fluctuations under weather forecasting error. An error is added to the actual 5-min weather for one month to reflect the effect of the forecasting error. Performance evaluation is based on the common error metrics correlation coefficient (r), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percent error (MAPE). Numerical testing results show our models outperform a standalone LSTM model, especially during days with high weather fluctuations.

The main contributions of this paper are summarized below:

1. Classification of weather conditions at each time step to address the rapid PV power fluctuation.
2. A regression model for each weather class for PV prediction with the absence of irradiance sensors data, and an enhanced MWCR model with an LSTM model power prediction added to inputs for regression.
3. Improved PV power forecasting accuracy compared to single regression, convolutional neural networks (CNN), gated recurrent unit (GRU), and LSTM models during different days of the year, especially days with high weather fluctuations.

2 Literature review

This section gives a summary of the existing work on forecasting the PV output power. Section 2.1 reviews regression models, Section 2.2 shows power-to-power forecasting methods using only PV power data, and Section 2.3 summarizes weather classification methods.

2.1 PV power forecasting using regression

Regression models need previous knowledge of weather data and PV output power. Selecting input data and regression models are discussed in Sections 2.1.1 and 2.1.2, respectively.

2.1.1 Selecting meteorological factors

Selecting the climatological factors affecting the PV output is

controversial. The most widely known parameters are temperature and irradiance [14, 15]. However, more factors showed an influence on PV output power. Zakzouk et al. [16] did an experimental study about the effect of normal radiation, ambient temperature and wind speed on the PV power and efficiency. As the temperature increases, the PV output power and efficiency decrease. They increase with increasing the solar radiation and wind speed as wind cools solar panels preventing electrons from escalating to the excited state and allowing more current to pass. Chaichan and Kazem [17] tested the effect of humidity on the PV panel output over one year, and Park et al. [18] conducted five different tests to test the relationship between moisture variation and PV power output. Both studies showed that humidity reduces the PV efficiency and leads to fast degradation since water droplets found on the PV panel can reflect and refract sunlight beams away from the panel causing less absorption of sunlight by the panel. Jena et al. [19] showed by experiment that snow decreases the PV output power. Hosenuzzaman et al. [20] studied the effect of shading on PV performance; a partial shading of the PV surface can decrease the PV output power by more than 50%. Samal and Bisoyi [21] demonstrated that moisture and dust have a high effect on reducing the PV output power, where dust showed the most significant influence by reducing generation by 2.73% per day.

2.1.2 Selecting regression model

Different regression models for forecasting the PV power have been developed. Mathur et al. [22] did a comparative study of the linear, Poisson and Lasso regression to predict the PV power based on temperature, humidity and luminous intensity. Results showed that linear regression outperformed the other two models. Alfadda et al. [23] used support vector regression to predict one day ahead of PV power forecasting with one hour as a timestep. Features selected for regression models were temperature, humidity, visibility, windspeed, and wind direction. Support vector regression outperformed the polynomial and Lasso regressions. Shaban et al. [24] compared classical and neural network models' performances in PV power forecasting. The classical model was the multiple linear regression and the neural network models were a multilayer perceptron using a radial basis function. Inputs of regression models were temperature and irradiance. Neural network models showed higher accuracy than their linear regression counterparts.

Machine learning models have shown higher accuracy over classical models. Several artificial neural network techniques have been developed for predicting PV output power. Fentis et al. [25] tested the performance of two supervised machine learning techniques, least square support vector regression and feed-forward neural network with the Levenberg–Marquardt algorithm. Inputs of the regression models were solar irradiance and cell temperature. Feed forward neural network outperformed the least square support vector regression. Serfa Juan and Kim [26] developed an algorithm that forecasts the PV power using a generalized regression neural network (GRNN). The developed model inputs are the IV characteristic curve, weather conditions, and temperature parameters. The developed model shows a higher accuracy compared to a feed-forward neural network.

Given the high performance of GRNN model, it was chosen for comparison with the developed models in this paper.

2.2 Power-to-power forecasting

The PV output power should be available to build a regression model for prediction. However, the PV output power is time-

series data and could be used alone to build a power-to-power time series prediction. Gupta et al. [27] built two machine learning models, the Facebook Prophet and XGB model for solar power forecasting. At the end of that study, the XGB model performed better than the Facebook prophet model. Different deep learning neural networks (DLNN) have been used for short and long-term prediction using only historical PV power data. LSTM networks were first introduced in Ref. [28] as a modified recurrent neural network. LSTM model output depends on cell states which are the long-term memory of the network, output data at the previous time step and input data at the current time step. An LSTM layer is able to learn long-term dependencies and is often used for both time series prediction and language model predictions. A simple architecture consists of a set of LSTM units each representing a state at time step t and a dense output layer as shown in Fig. 2. The performance of five different deep learning models has been tested in Ref. [29] where LSTM achieved the highest accuracy. Due to the dependency of PV output on meteorological factors, regression models that relate these factors with PV output have been more accurate than time-series predictions, especially during days with highly variable weather conditions [30]. Kardakos et al. [31] addressed two methods for PV forecasting, the time series method using autoregressive integrated moving average with PV power as input, and the artificial neural network (ANN) regression model with multiple weather inputs, where ANN shows better accuracy.

2.3 Weather data classification

Forecasting PV power is more challenging in winter than in summer, due to the high fluctuations in power caused by the variable weather conditions, such as irradiance, temperature, wind speed, and humidity in winter. Therefore, weather classification methods are reviewed in this section. Zhao and Wu [32] developed a weather classification model using a convolutional neural network to classify visual images into 4 different classes, sunny, foggy, rainy, and snowy. The developed model showed a better accuracy of 5% compared to traditional techniques. Mali et al. [33] tested the performance of support vector machine (SVM) and linear regression in classifying weather conditions based on measured data, temperature, wind speed, humidity, and pressure. Data were classified into four different classes, cloudy, partly cloudy, sunny, and cold. Linear regression performs with an accuracy of 94%, 95%, and 71% using temperature, pressure, and humidity respectively. Support vector machines' classification accuracy was 96%. Naik and Pathan [34] used feed-forward ANN to predict hourly future weather conditions and classify them using the Levenberg–Marquardt algorithm. The inputs of the weather prediction model were temperature, air pressure, and wind direction. The weather was classified as sunny, rainy, or partly cloudy. The Levenberg–Marquardt algorithm showed the best performance over other back propagation algorithms.

2.4 Classification-based regression models

In order to capture the fluctuations of the PV power, several researchers have developed weather classification-based prediction models.

Shi et al. [35] developed a PV power prediction model that is based on weather classification and SVM. Weather conditions (solar irradiance and temperature) are classified into clear sky, cloudy day, foggy, and rainy for the whole day, and the PV power is predicted for the coming 24 h. Results showed a significant improvement in the prediction accuracy. Agarwal et al. [36] classified whole day weather conditions using radial basis function support vector machines into sunny, rainy, and cloudy. Then, the

neural network model corresponding to each weather class has been applied to forecast the solar PV output. Wang et al. [37] divided the daytime into 2-h intervals and added a weather class for each. Weather conditions are classified into 4 different categories based on irradiance, temperature, wind speed, and humidity.

In order to improve the accuracy of PV power forecasting under continuously variable weather conditions, the developed model classifies these conditions at each time step rather than the whole day or large time window as done in prior work.

3 Methodology

This section presents the MWCR model layers, the E-MWCR (enhanced MWCR) model, evaluation criteria, and dataset description. The physical model layer is presented in Section 3.1; LSTM layer in Section 3.2; classification layer in Section 3.3; regression layer in Section 3.4; the enhanced model is presented in Section 3.5; evaluation criteria in Section 3.6; dataset description in Section 3.7 and experimental testing description in Section 3.8.

To address the difficulty of predicting the PV power on days with rapid weather conditions fluctuation, the key idea is to classify weather conditions in each time slot into four different weather classes (clear sky, partly cloudy, mostly cloudy, and cloudy). The classification will be based on temperature, wind speed and humidity as discussed in the third layer in Section 3.3. Irradiance is the most significant input for weather classification, but irradiance sensors are rarely installed in solar farms. So, to address this issue, the novel idea is to add the ratio of PV actual power over the maximum power calculated by the physical model in a clear sky using the local irradiance in clear sky and external temperature (1st layer discussed in Section 3.1) to reflect sky condition. However, the actual power for the next 24 h is not available. So, the LSTM model (2nd layer discussed in Section 3.2) is added to predict the PV power for the next 24 h with 5 min as a time step to give an estimate of the actual PV power data. In the 4th and final layer, a unique regression model is built for each weather class with power ratio, maximum local irradiance, external temperature, wind speed and humidity as inputs as presented in Section 3.4. The developed multi-layer model consisting of four layers is presented in Fig. 1. The enhanced model with LSTM predicted power added to regression inputs is defined. In Section 3.6, the model evaluation criterion is presented.

3.1 First layer: Physical model

Weather data, such as temperature, humidity, and wind speed are available online for certain areas such as a city or town. However, irradiance data are not available for small areas. In addition, most PV panels are not equipped with irradiance sensors. For this reason, the maximum local irradiance is considered in this study. The maximum local irradiance, $I_{\max}$, depends on city latitude and varies from day to day, being higher during the summer months. Also, irradiance changes from one hour to another due to the change in the hour angle that is considered 0° at solar noon where highest solar irradiance is reached. Local irradiance, $I_{\max}$, for a specific area in a clear sky condition is calculated using Eqs. (1) and (2) [38]:

$$I_{\max} = I_0 (1 + 0.033 \cos(2\pi(n-4)/365)) (\sin L \sin \delta + \cos L \cos \delta \cos \omega) \quad (1)$$

$$\delta \approx \varepsilon \sin\left(\frac{(284+n)360}{365}\right) \quad (2)$$

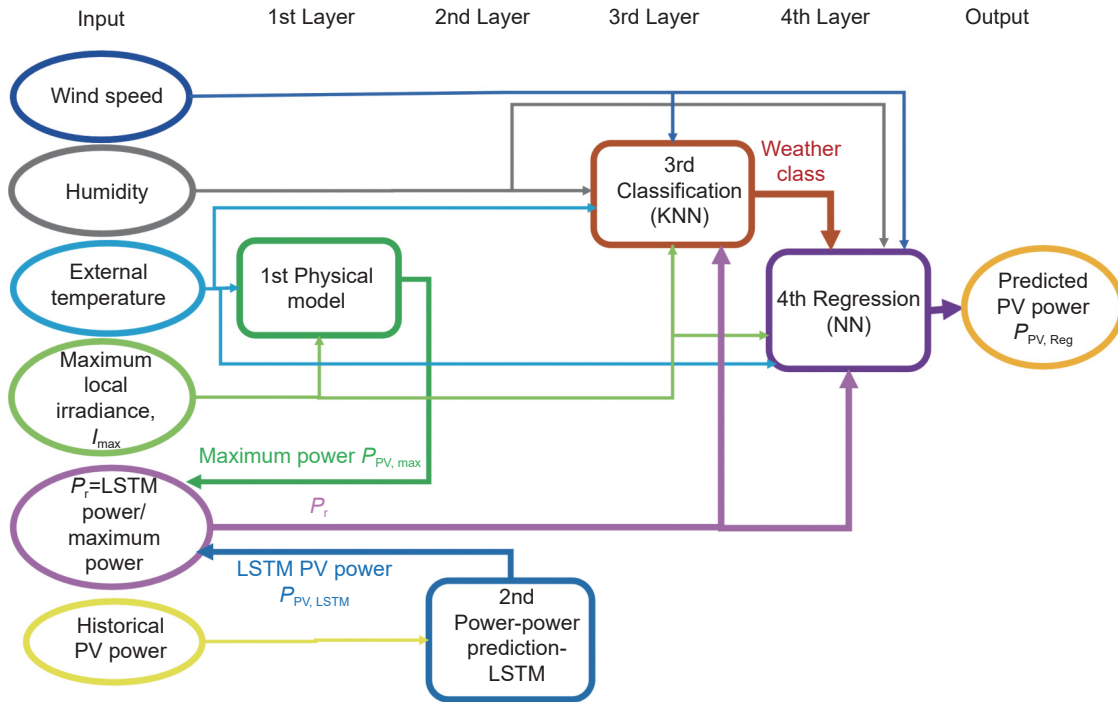


Fig. 1 MWCR model.

where δ is the declination angle of the Sun, n is the sequential number of the day in a year with $n=1$ for the 1st of January, $\varepsilon=23.45^\circ$, L is the city latitude, $I_0=1.366 \text{ kW/m}^2$ is the solar constant, and ω is the hour angle that reflects the variation of the location of the sun in the sky by time. This value is zero at solar noon, negative before, and positive after the solar noon. It is calculated using Eq. (3):

$$\omega = (t_s - 12 \text{ h}) \times 15/\text{h} \quad (3)$$

where t_s is the solar time calculated as in Eq. (4):

$$t_s = t_c + \frac{\lambda}{15/\text{h}} - Z_c + E \quad (4)$$

where t_c is the civil time in hour, λ is the city's longitude, Z_c is the time zone defined as + or - an integer number of hours in relation to Coordinated Universal Time. In addition, E is the "equation of time" used to calculate the difference between the time shown on a standard clock and the time indicated by the Sun's position (as seen on a sundial) [39]. The equation of time can be approximated as shown in Eq. (5):

$$E = 6.24 + 0.017 [365.25 (y - 2000) + d] \quad (5)$$

where d represents the number of days since 1st of January of the current year y .

Irradiance calculated using Eq. (1) is the irradiance that is coming from the Sun in a clear sky condition and not the one hitting the PV panel. So, using Eq. (1) in a PV model will not give the actual output power, instead, it will represent the maximum PV power in a clear sky with a panel moving to face the Sun. The PV power mathematical model, P_{PV} , is presented in Eq. (6):

$$P_{PV} = P_{STC} \frac{I_C}{I_{STC}} [1 + k(T_C - T_{STC})] \quad (6)$$

where P_{STC} is the power at the standard test condition, I_C is the actual irradiance during measurement considering the sky condition which is less than the maximum irradiance in the clear

sky $I_{\max}$ in Eq. (1), I_{STC} is the standard test condition irradiance (1 kW/m^2), k is the power temperature coefficient (0.5%), T_C is the temperature at the operating point, and T_{STC} is the temperature at the standard test condition (25°C).

The maximum irradiance calculated in Eq. (1) is substituted in Eq. (6) to get the maximum PV power in a clear sky as presented in Eq. (7).

$$P_{PV,\max} = P_{STC} \frac{I_{\max}}{I_{STC}} [1 + k(T_C - T_{STC})] \quad (7)$$

The actual irradiance received by the panels is not available due to the absence of irradiance sensors; therefore, the actual PV power cannot be calculated. For this reason, the maximum PV power in a clear sky calculated in Eq. (7) and the LSTM PV power to be introduced in the next subsection are used to get the power ratio that could reflect the sky condition.

3.2 Second layer: Power-to-power prediction by LSTM

An estimate of the PV output power is needed to calculate the power ratio; however, during real-time forecasting, only time series predictions can be used to predict the PV power due to the absence of weather inputs used in regression models. In this layer, LSTM is used for a time series PV power prediction that depends only on PV power historical data. The inputs will be P_1 till P_t which represents the PV power for each 1 h of the input sequence length. P_{t+1} is the power to be predicted one step after the end of the input sequence as shown in Fig. 2.

In this study, the output is the LSTM PV power $P_{PV,LSTM}$ of sequence length of 24 for the 1-h time step data which represents the PV power in 24 h at 1-h time steps where P_{t+24} is the last output element. A sequence-to-sequence prediction is used with an input vector including PV power of the length of 3 months with 1-h time step. The goal is to predict the coming 24 h, so the output is a sequence of PV power for the next 24 h (24 data points) with 1-h as a time step at the end of the input vector. During training, the data will be shifted by 24 h, so the second input vector to be trained will start from the 2nd of January.

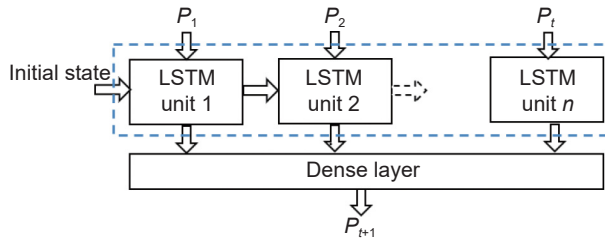


Fig. 2 Single LSTM configuration: one hidden layer with n units and one dense.

LSTM will continue to train the model till the end of the given training data.

The power ratio P_r of predicted power using LSTM over maximum power calculated in Eq. (7) can reflect the sky condition Eq. (8).

$$P_r = P_{PV,LSTM} / P_{PV,max} \quad (8)$$

The closer the ratio to one, the irradiance is closer to its maximum value, and hence the sky is clear. This ratio will address the issue of the missing actual irradiance value received by the PV panel.

3.3 Third layer: Classification by k-nearest neighbor

Due to the high intermittency of PV power within the same day, a single regression model is not sufficient for accurate prediction. Since irradiance has the most significant influence on PV power, meteorological factors along with power ratio are classified in terms of sky condition that reflects the irradiance value, i.e., clear sky, partly cloudy, mostly cloudy, and cloudy classes. k-nearest neighbor (KNN) classifier is used with the ratio of predicted actual power from the 2nd layer over the maximum power from the 1st layer, external temperature, wind speed, and humidity as inputs. KNN assumes that similar labels are close to each other, so it will assign the test point a class label based on the majority vote.

3.4 Fourth layer: Regression by GRNN

For each weather class, a regression model is built based on GRNN which was first introduced by D. F. Specht [40]. It is an improved technique of neural networks based on nonparametric regression. A GRNN is a radial basis feedforward neural network that consists of four layers, input, pattern, summation and output layer as shown in Fig. 3.

For the problem under study the local irradiance, temperature, humidity, wind speed, and power ratio were used for each weather class separately to predict the PV output power. The output is the predicted PV power given the value of the input vector.

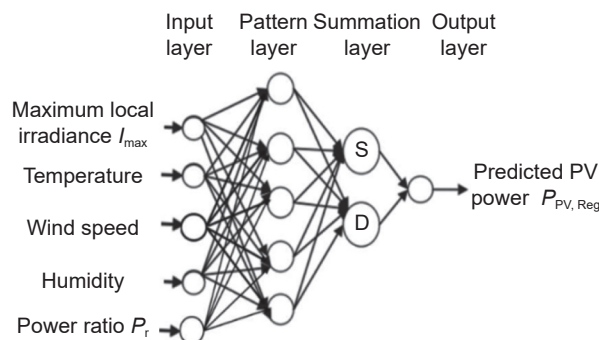


Fig. 3 GRNN model.

3.5 E-MWCR model

During a sunny day, the maximum local irradiance is close to the actual irradiance at the PV panel. During a cloudy day, the PV panel receives the minimum value of irradiance compared to other weather classes. Therefore, the regression model can predict the PV power with good accuracy in a clear or cloudy sky. On the other hand, during a partly or mostly cloudy day, the portion of received local irradiance is uncertain. The received irradiance by the panel could take any value from the maximum local irradiance depending on the cloud cover. Due to the uncertainty of weather conditions on partly and mostly cloudy days, misclassifications are expected to occur during days with highly variable weather conditions, and the MWCR model cannot accurately predict PV power.

To address the above, the LSTM predicted power is added to the regression model to estimate the PV power value at a certain time. Consecutive days are mostly having close average daily power. LSTM average daily predicted power is most probably higher than the actual average daily power, due to the fact that PV power takes an inverted parabola shape on most of the days, so LSTM will be trained and predict that shape in most cases which would be far from the actual PV power with fluctuations due to varying weather conditions. An opposite correlation between predicted power using LSTM and actual power exists. To build the E-MWCR model, the LSTM power is added as an additional input to the regression model in the fourth layer of the MWCR model (Fig. 1), and all other layers of the MWCR model are kept with the same inputs and outputs. The LSTM power input to the regression model is presented as a dashed line in Fig. 4. The classification layer will help in the prediction of PV fluctuations and LSTM power will help in getting the predicted power closer to actual value, for example, actual power will decrease in case of changing the weather class from clear sky to partly cloudy, LSTM cannot rapidly transfer from a higher to a lower value and vice versa, resulting in a higher error in predicting during transition from one weather class to another. So, a regression with LSTM power input gives an estimate of the power that should be dropped after being classified.

3.6 Model evaluation criterion

The performance of layers and multi-layer model is evaluated based on the common error metrics: correlation coefficient shown in Eq. (9), mean absolute error (MAE) in Eq. (10), mean absolute percent error (MAPE) in Eq. (11), and root mean squared error (RMSE) in Eq. (12).

$$r = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (9)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (10)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (12)$$

where x_i and y_i are the actual and forecasted values of the PV output power respectively, and $\bar{x}$ and $\bar{y}$ are the average values of the actual and the forecasted data, respectively.

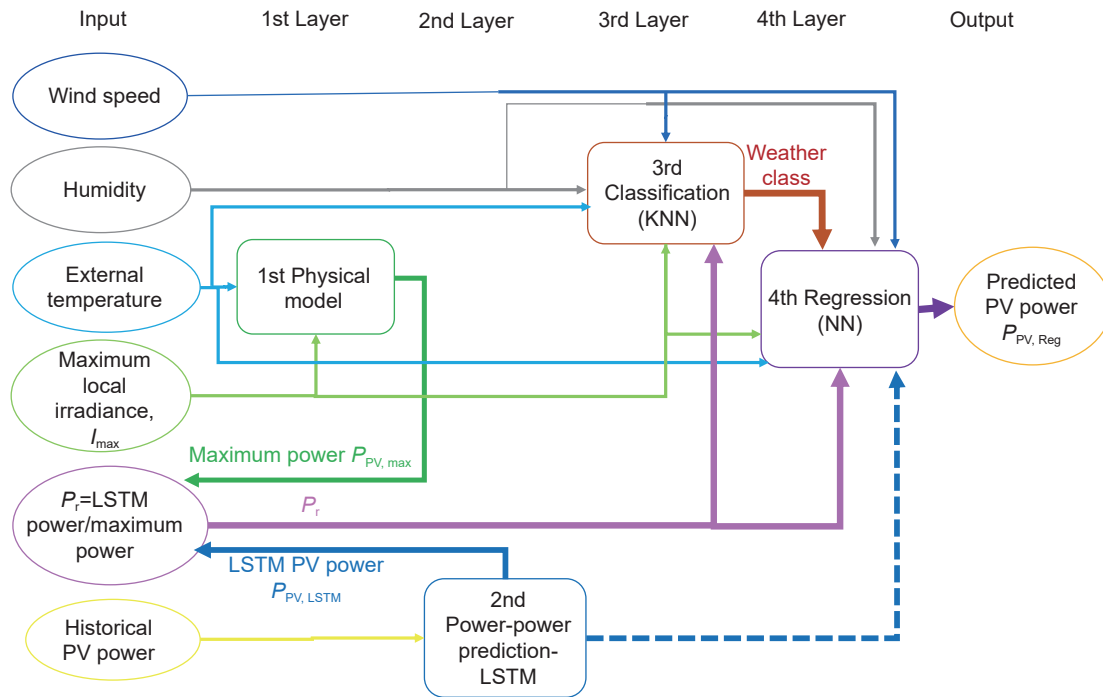


Fig. 4 E-MWCR model.

3.7 Dataset description

The testing data were collected from a 2 MW solar farm located at the Rochester Institute of Technology in New York. The PV output power is from June 13th, 2018, to December 31st, 2021, with 5 min time steps (347,124 data points) with some missing data due to interruption. The solar farm is shown in Fig. 5.

The weather conditions collected from NOAA consisting of external temperature, humidity, wind speed and direction are available from June 1st, 2018, to December 31st, 2021, with 5 min time steps (357,171 data points) with some missing readings due to interruption. However, only 26,006 time slots, that are randomly located over the 3.5 years, have weather classes. The weather classes are classified as 6662 clear sky, 6918 partly cloudy, 5730 mostly cloudy, and a total of 6696 of wintry mix, storm, snow, flurries, drizzle, and cloudy which are combined as cloudy. Figure 6 represents the daily average PV output power, temperature, humidity, and wind speed over the year 2021.

Referring to Fig. 6(a), wind speed records the highest values between 5 and 22 mph for the days between 320 and 356, which correspond to mid-November till the end of December. The lowest wind speed between 1.7 and 5 mph corresponds to days from 165 to 260, which is mid-June to mid-September. Referring to Fig. 6(b), the lowest humidity between 30% and 70% is recorded from days 52 to 92, which corresponds to the end of February till the first week of April. Other days of the year show approximately the same fluctuating humidity between 60% and 90%. Referring to Fig. 6(c), the highest temperature values



Fig. 5 Solar farm at RIT [41].

between 58 and 84 °F are recorded from days 160 to 240, which correspond to June, July, and August. The lowest temperature values between 15 and 29 °F are recorded from days 25 to 50, which correspond to the end of January and mid-February. Referring to Fig. 6(d), the maximum power of 1189 kW is recorded on the 29th of June, and the minimum of 108 kW is recorded on the 1st of January.

3.8 Experimental testing description

Testing MWCR and E-MWCR performance includes testing of each layer in the model structure, testing the effect of omitting a layer from the model and comparing model prediction performance with other models.

Models' performance is compared with the performance of the single regression, CNN, GRU, and LSTM models. The single regression is unclassified regression using GRNN with 5 input features (local irradiance, temperature, humidity wind speed, and power ratio); after hyperparameter tuning, hidden layers are set to 30 and 10 neurons, respectively; and activation of "ReLU". In the CNN, GRU, and LSTM configurations, the number of hidden layers is 6, each with 24 units. Units are equal to the number of time steps in one day (24 h). The epoch size is set to 100, and the batch size is 64 using the "Adam" optimizer. For MWCR and E-MWCR, the number of the input features is 5 (temperature, humidity, wind speed, local irradiance, and power ratio) and 6 (same as in the MWCR model with LSTM predicted power added), respectively, and the activation function and the number of hidden layers are the same as those of the single regression.

The dataset was split as the first 80% for training, second 10% for validation, and the last 10% for testing, to maintain the data sequence to be used for time series prediction models for fair comparison. Days of missing information are omitted from the training and testing datasets to make sure that data are kept sequential. Also, testing days are eliminated from the training dataset.

MWCR, E-MWCR, and single regression inputs should be the forecasted weather conditions (temperature, wind speed, and

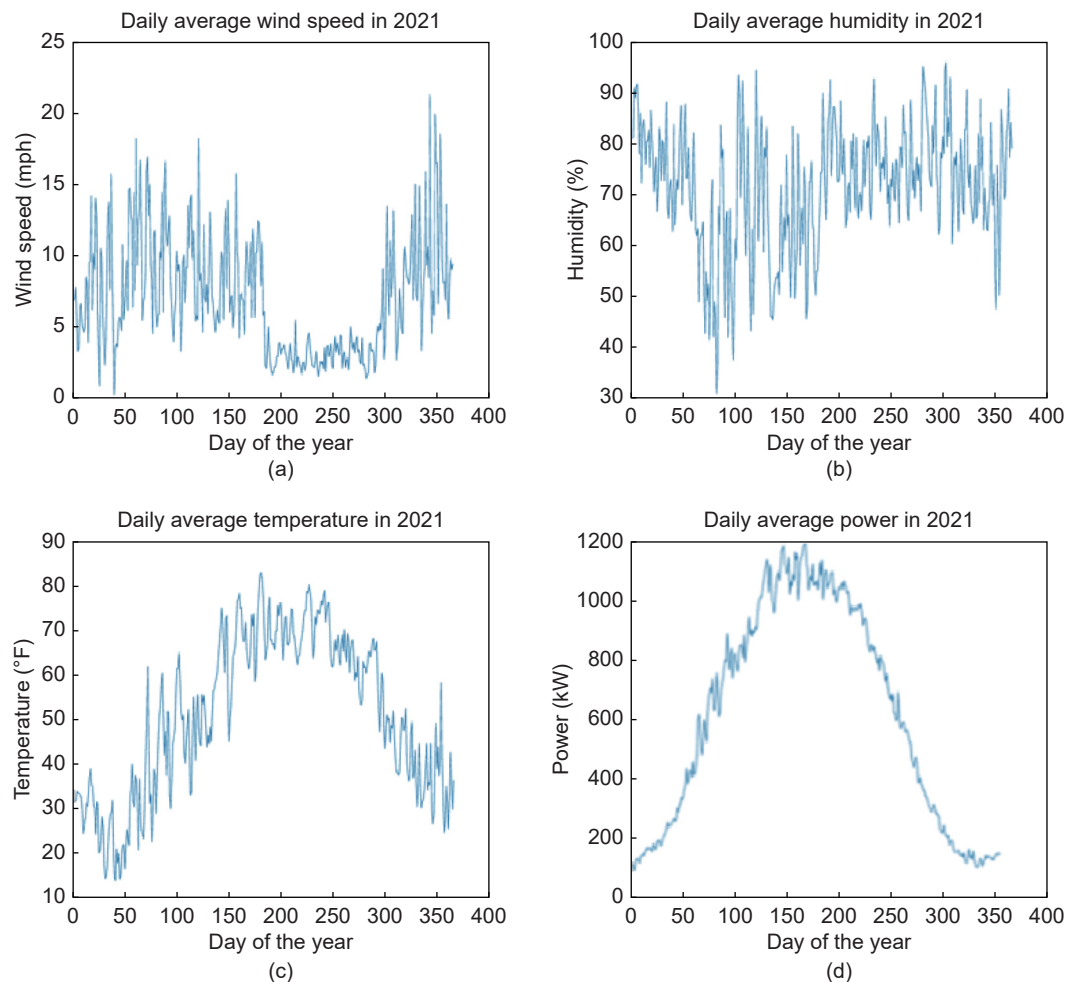


Fig. 6 Daily average in 2021: (a) wind speed (mph); (b) humidity (%); (c) temperature (°F); (d) power (kW).

humidity) for the next day and not the actual data. The MWCR model accuracy is tested using 3 different datasets. The first is to test the prediction performance in a real-life environment using the historical actual and historical forecasted weather data with 1 h as a time step. The data are collected from NOAA for training and validation for 3 years and 54 days (80% of the available 3.5 years) and the historical forecasted weather data are collected from Ref. [42] for 4 months and 6 days (10% of the available 3.5 years). The model has also been evaluated using 5-min weather data to address the rapid fluctuations in PV power that are not apparent in hourly data. So, the second test is to test the model performance during rapid PV fluctuations within minutes assuming perfect weather forecasting in the future with 5 min as a time step. The data are the actual historical weather conditions since the historical forecasted data is not available collected from NOAA with 5 min as a time step for training, validation and testing (at 80%, 10%, and 10% data split, respectively) for a total of 3.5 years. In order to perform a more realistic test, a noise is added to the 5-min actual weather data for one month in the third test to reflect the forecasting error in the weather data since forecasted 5-min weather data are not available. The forecasted 15-min weather data are available for the coming day. However, PV power data are not recorded at a fixed time steps after December 2021 due to technical issues. So, testing the model is not possible at the 15-min time step since forecasted weather data and PV power data are not at the same time steps in June 2024. The main goal is to create a noise in the 5-min weather data in a realistic way that matches variations in weather conditions in the 15-min data on days with

close PV power fluctuations for a month when 5-min weather data and PV data are available. The 15-min weather actual and forecasted data were collected manually for June 2024 from Ref. [43]. A probability distribution is created for each day separately using MATLAB, and the deviation in the actual PV power in each day is measured. The noise is added to the 5-min weather data of June 2021 for each day based on the probability distribution created in a day with the closest PV power deviation. The cumulative distribution function is measured at each 5-min based on the 15-min probability distribution to get the estimate of the forecasted weather. The training and validation datasets are the same historical actual weather data with 5 min as a time step used in the second testing case for 3 years and 54 days. The testing data are the weather data with 5 min as a time step with the added noise for June 2021.

4 Results

This section presents a summary of testing results of 1st and 2nd layers in Section 4.1, testing results of 3rd and 4th layers in Sections 4.2 and 4.3, respectively, the effect of each layer in Section 4.4, and comparison of MWCR and E-MWCR performance to other models' results in Section 4.5.

First, layers are tested individually in a disconnected mode, so that in case the model shows a low accuracy the cause of the error will be known. No connection between layers is established and inputs of each layer are taken from the weather dataset presented in Section 3.6. The physical layer dataset inputs are taken from the

weather dataset; the LSTM layer inputs are taken from historical PV data; the classification and regression layers inputs are taken from the weather dataset without considering the weather class in the regression layer. After testing each layer separately, all layers are connected together as shown in Fig. 1 to form MWCR. Testing accuracy is then evaluated using the common error metrics mentioned in Section 3.6. The models have been implemented in MATLAB 2022 Deep Learning toolbox.

4.1 Mathematical model layer and LSTM results

Local irradiance values of Rochester city are calculated using Eq. (1), where latitude $L = 43.16103^\circ \text{N}$, ω is the hour angle, and solar constant $I_0 = 1.366 \text{ kW/m}^2$. There is no prior knowledge of the physical parameters of the PV panels. Using the historical power dataset, power at standard test condition P_{STC} at temperature of 25°C and irradiance of 1 kW/m^2 is found to be 1636 kW . Using the physical model presented in Eq. (4), the maximum PV power $P_{\text{PV,max}}$ is calculated in a clear sky condition at irradiance calculated using Eq. (1). Figure 7 represents actual PV power, PV power calculated using the physical model, and the ratio of actual power over maximum power calculated from the physical model for the year 2021.

According to Fig. 7, the difference between the actual PV power output and the PV power output calculated using the physical model is large during the winter since the physical model considers the local irradiance in clear sky conditions. In addition, the ratio is closer to one during the summer when most of the days have clear sky, and the local irradiance is close to the actual irradiance. The maximum power ratio is 0.976 at day 165 which is mid-June, and the minimum ratio is 0.586 on the 2nd of January.

4.2 Classification results

KNN classifier inputs are temperature, humidity, wind speed, and the ratio of LSTM power over maximum power with a k (number of neighbors) value of 5. The available 26,006 classified time slots are divided sequentially into 80% for training, 10% for validation, and 10% for testing. Table 1 represents the confusion matrix of classifying the 4 different weather classes of the testing data.

As shown in Table 1, the highest classification accuracy is recorded for clear sky and cloudy since weather conditions of these two classes come at extremes. Also, the misclassification occurs between two consecutive weather classes since weather conditions are close and varies smoothly between two close sky conditions. For clear sky, only 0.04% was wrongly classified as

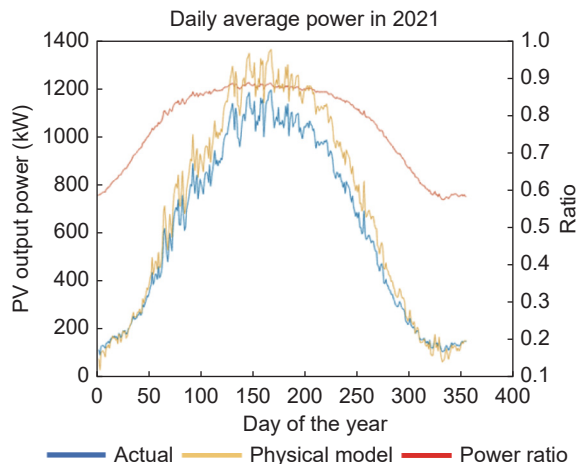


Fig. 7 Daily average actual power, physical model power, and power ratio in 2021.

Table 1 Classification accuracy

Actual value	Predicted value			
	Clear sky	Partly cloudy	Mostly cloudy	Cloudy
Clear sky	6633	29	0	0
Partly cloudy	21	6707	165	25
Mostly cloudy	5	52	5484	189
Cloudy	0	22	78	6596

partly cloudy. Cloudy class was misclassified at 1.16% and 0.03% between mostly cloudy and partly cloudy, respectively. Partly cloudy was misclassified between mostly cloudy and cloudy with 1.7% and 0.25% respectively. Mostly cloudy recorded the lowest accuracy where 0.93% and 3.2% were wrongly classified into partly cloudy and cloudy respectively. The values of temperature, power ratio, humidity, and irradiance should be highest during clear sky day and lowest during cloudy day, so these two extreme classes are classified more accurately than the two others (partly and mostly cloudy). After testing the KNN classifier and getting good accuracy, the remaining unclassified weather conditions are filled.

4.3 Regression results

In this section, GRNN was tested to check its performance as a single layer. No connection between previous layers is established. The aim is to find the relationship between meteorological factors, local irradiance in a clear sky, temperature, humidity, and wind speed, with PV output power for each weather class. The 26,006 data points with assigned weather classes are used. Testing results for each weather class are shown in Table 2. Figure 8 shows the predicted vs. actual power for clear sky, partly cloudy, mostly cloudy, and cloudy days.

The highest recorded PV power in the available 26,006 classified weather is 1636, 1475, 860, and 456 kW in clear sky, partly cloudy, mostly cloudy, and cloudy classes, respectively, as shown in Fig. 8. GRNN for a clear sky class shows the best performance, r is 99.8% due to the close values of actual irradiance and local irradiance. Cloudy days shows also a good r of 95.65% since in cloudy days when the sky is completely covered with clouds, the actual irradiance is a specific portion of the local irradiance. Partly cloudy and mostly cloudy classes recorded the lowest accuracy, r of 88.45% and 87.89% since these are classes of wide range of weather conditions.

4.4 The MWCR model test: Effect of each layer

In order to test the effect of each layer on the performance of the whole MWCR model, each layer is omitted from the model and the training and testing performance using evaluation criteria mentioned in Section 3.6 is evaluated. The 1-h actual weather data and actual PV data for the first 80% of available data are used for training and validation. The 1-h forecasted weather conditions collected from Ref. [42] for the testing data in the dataset along with the PV power collected at the same time from solar farms are used for testing.

The first and second layers are used to calculate the power ratio

Table 2 GRNN accuracy

Testing data	r (%)	MAE (kW)	MAPE (%)	RMSE (kW)
Clear sky	99.86	75.25	24.96	170.56
Partly cloudy	88.45	75.64	38.50	169.79
Mostly cloudy	87.89	70.45	47.79	83.48
Cloudy	95.65	49.16	26.45	57.55

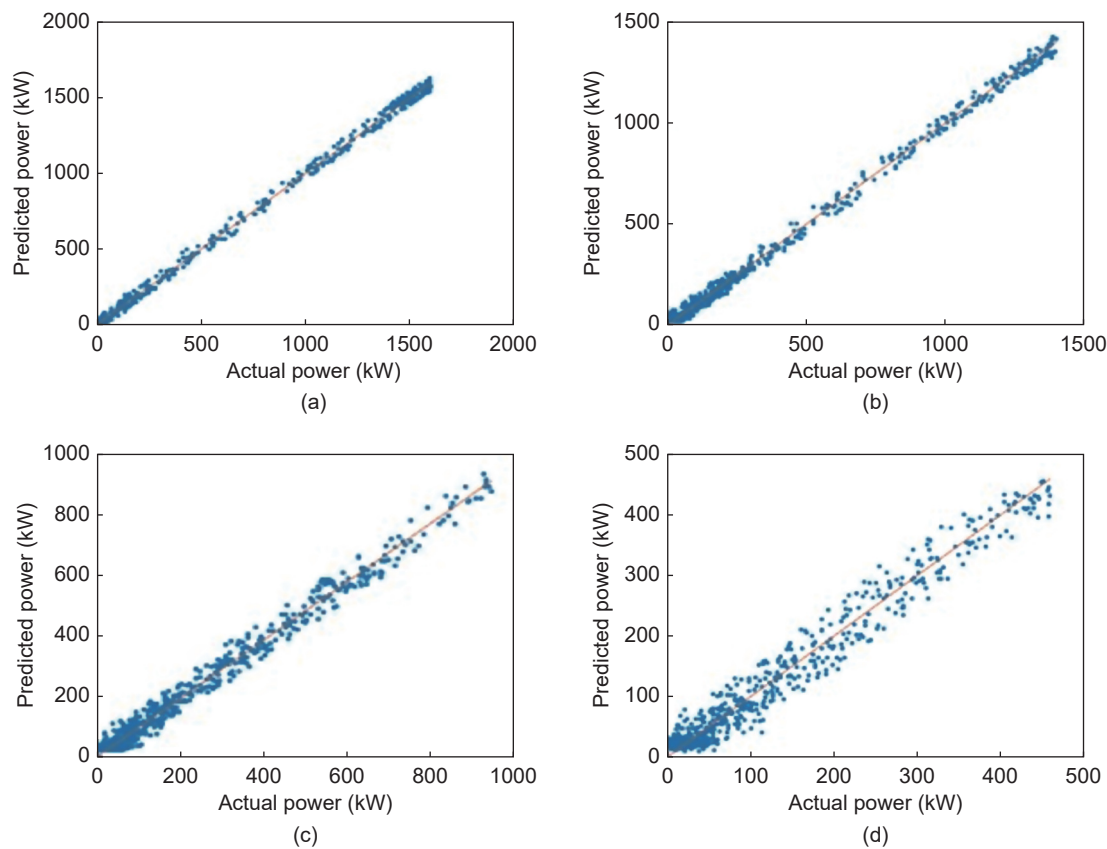


Fig. 8 Actual and predicted power for different weather classes: (a) clear sky, (b) partly cloudy, (c) mostly cloudy, (d) cloudy.

to be used as an input in both the classification and regression layers (3rd and 4th layers, respectively). The 1st and 2nd layers are omitted together to test the effect of the power ratio on the prediction accuracy. The third layer is omitted to test the effect of the classification layer, so the PV power is predicted using single regression without classification. Results are shown in Table 3.

Omitting 1st and 2nd layers results in removing the power ratio P_r from the MWCR inputs. The power ratio is used to replace the irradiance which is the most significant factor affecting the PV power [20]. This results in a high prediction error, and r is only 46.29% in the testing datasets, respectively. Omitting the classification layer results in a single regression model that will predict the PV power without classification. All of the model inputs (power ratio, maximum local irradiance, temperature, wind speed, and humidity) are kept as inputs for the single regression model which increases the prediction accuracy over omitting the power ratio input. The correlation r is 79.4% on the testing dataset. MWCR model with all inputs and layers records the best accuracy since it is a classification-based regression model that uses power ratio to estimate the actual irradiance data.

4.5 Testing MWCR and E-MWCR accuracy

The MWCR and E-MWCR models are tested using three

different datasets: hourly actual weather data, 5-min actual weather data, and 5-min actual weather data with added noise. The first is to test the prediction performance in a real-life environment using the historical actual and historical forecasted weather data with 1 h as a time step. The second one is to test the model performance during rapid PV fluctuations within minutes assuming perfect weather forecasting in the future with 5-min as a time step. The third one is to test the model performance during rapid PV fluctuations within 5-min considering weather forecasting with errors. A distribution of error is formed based on the error that occurred in the 15-min weather data. The training and validation datasets are the same historical actual weather data with 5-min as a time step used in the second testing dataset. The testing data are the weather data with 5-min as a time step with the added noise for June 2021. Model performance is compared with the performance of the single regression, CNN, GRU, and LSTM models.

4.5.1 Prediction results using the 1-h actual weather data

In this test, the MWCR and E-MWCR are tested in a real-life environment since both the historical actual and historical forecasted weather data are available at a 1-h time step for previous years. The actual 1-h weather and PV data will be

Table 3 Effect of each layer on the MWCR model performance

Missing layer	r (%)		MAE (kW)		MAPE (%)		RMSE (kW)	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
First and second	46.29	41.24	288.39	301.46	102.80	285.92	160.20	191.30
Third	81.12	79.40	115.82	123.60	51.10	198.41	112.20	157.23
MWCR	90.11	86.67	51.94	82.15	24.67	40.23	78.87	88.95
E-MWCR	94.89	91.06	49.52	76.30	21.67	38.20	57.56	74.02

extracted from the whole available dataset discussed in Section 3.7 for training and the forecasted weather data for Rochester city are taken from Ref. [42] for testing. Training will be done using the actual weather conditions, temperature, humidity, wind speed along with the local irradiance calculated using Eq. (1) and the power ratio calculated from the 1st and 2nd layers.

During testing, the forecasted weather conditions along with the calculated local irradiance and power ratio will be used as inputs for the MWCR model starting after the end of training and validation data. Other prediction models that are used for comparison are time series models that require only time series data of the PV power for the last 3 months at 1-h time step. Results shown in Table 4 demonstrate the accuracy of E-MWCR model compared to all others.

To test model performance in different weather conditions, prediction accuracy is calculated on six days over the year as shown in Table 5 and Fig. 9. One sunny day (1st of August) where PV prediction is easy with no unpredicted variations in weather conditions, a spring day (1st of March) when small weather variations occur, two rainy days (1st of January and 1st of October) when the weather class is cloudy for most of the times and PV power output is low, two days during the transition from one season to another where weather conditions show rapid fluctuations (1st of May and 1st of October). On a sunny day with no PV fluctuations like 1st of August all models show a high r with the highest recorded by the E-MWCR. During that day the sky is clear with no highly variable weather conditions, so the PV power takes a perfect inverted parabola shape making it easier for

Table 4 Performance comparison between single regression, CNN, GRU, LSTM, MWCR, and E-MWCR using the 1-h actual data

Model	r (%)		MAE (kW)		MAPE (%)		RMSE (kW)	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Single regression	81.12	79.40	115.82	123.60	51.10	198.41	212.20	157.23
CNN	82.90	80.30	116.40	120.50	48.98	177.42	195.59	149.26
GRU	83.76	80.50	110.22	117.50	46.21	160.54	154.67	136.49
LSTM	86.74	80.91	93.41	112.40	35.40	111.90	168.23	127.70
MWCR	90.11	86.67	51.94	82.15	24.67	40.23	98.87	88.95
E-MWCR	94.89	91.06	49.52	76.30	21.67	38.20	57.56	74.02

Table 5 Performance comparison between LSTM, MWCR, and E-MWCR on different days using the 1-h actual data

1st day	r (%)			MAE (kW)			MAPE (%)			RMSE (kW)		
	LSTM	MWCR	E-MWCR	LSTM	MWCR	E-MWCR	LSTM	MWCR	E-MWCR	LSTM	MWCR	E-MWCR
January	87.71	92.48	93.28	26.97	22.08	20.34	65.37	51.92	48.57	49.74	39.16	34.23
March	85.73	95.36	96.34	74.58	44.28	40.34	66.53	49.79	47.16	137.66	84.61	80.35
May	83.38	92.58	94.27	73.40	44.93	40.23	41.93	40.41	37.45	127.76	76.13	70.26
August	97.49	98.39	98.98	78.41	73.28	70.23	49.54	39.66	31.45	134.22	119.833	99.23
October	79.83	84.78	85.28	35.74	29.72	25.37	82.72	54.02	48.23	65.29	54.98	46.29
December	82.47	95.15	97.36	34.32	18.93	15.48	64.87	51.44	45.34	85.19	32.13	30.67

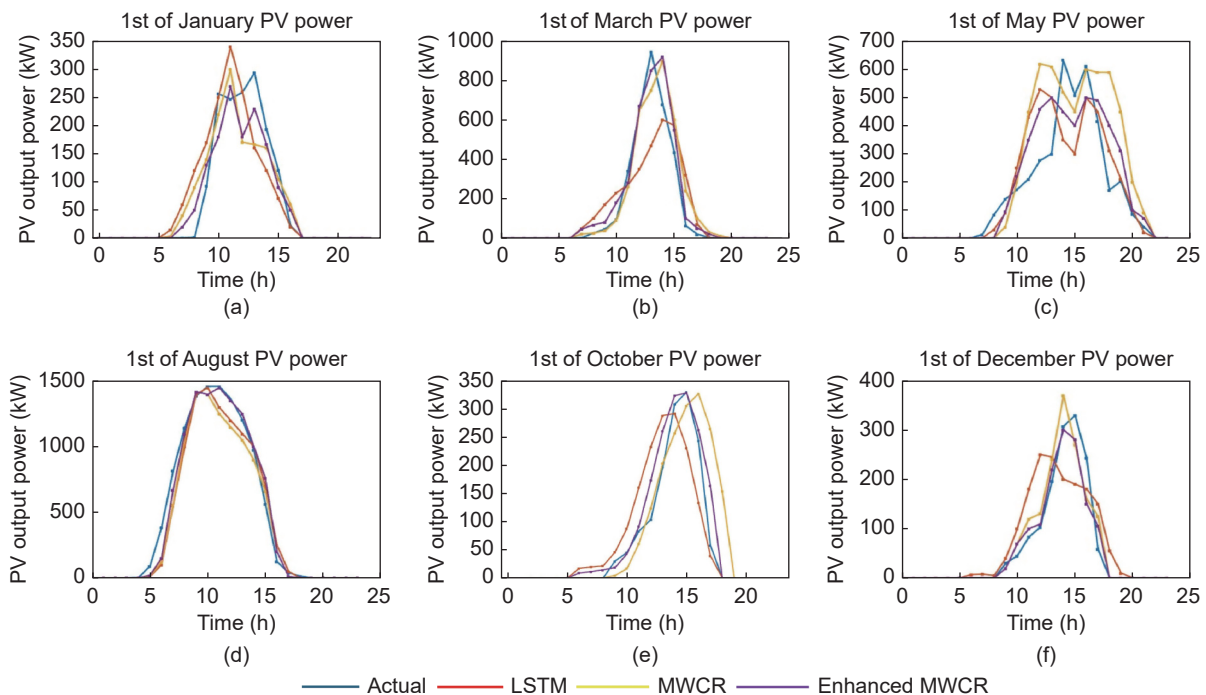


Fig. 9 Actual and predicted power in the first day of (a) January, (b) March, (c) May, (d) August, (e) October, and (f) December in the year of 2021 using LSTM and MWCR at 1-h data.

a sequence-to-sequence prediction to predict a “routine” shape PV power as shown in Fig. 9(d). On days with some PV fluctuations like on the 1st of January, 1st of March, and 1st of December, LSTM records less accuracy as r is 87.71% and 85.73%. The MWCR model improves accuracy significantly: r increases to 92.48% and 95.36%, respectively, and the E-MWCR further enhances r to 93.28% and 96.34%. As shown in Figs. 9(b) and 9(f), PV power shows some fluctuations, and the sky is mostly cloudy or cloudy with weather conditions.

The E-MWCR improved prediction accuracy over both models on all days, especially on days with high fluctuations in PV power. In days with seasonal changes such as in the first days of May and October when there are more fluctuations in the weather conditions and hence more fluctuations in the PV output power as shown in Figs. 10(c) and 9(e). The weather forecasting error is higher during these days that will affect the forecasting error of both MWCR and LSTM which can be resolved by the E-MWCR since it considers both weather condition and LSTM power as inputs to reduce the effect of weather forecasting error and misclassification of weather class.

4.5.2 Prediction results using the 5-min actual data

The available 1-h weather conditions data fail to capture many sudden variations in the PV power and the historical forecasted

weather conditions data at 5-min time steps is not available. For this reason, this test uses 5-min actual weather data collected from NOAA for training, validation, and testing, assuming a perfect weather prediction, to test the model performance during sudden PV power fluctuations within minutes. In this test, the MWCR and E-MWCR models performance are compared with single regression, CNN, GRU, and LSTM. Accuracy of models prediction using the training and testing data are shown in Table 6. E-MWCR outperformed all other models using both data. Also, MWCR and E-MWCR are compared to LSTM on different days of the year, and results are presented in Table 7 and Fig. 10. Figure 11 shows the plots of the predicted power using MWCR and E-MWCR in 1st days of May and October when the enhanced model shows the most significant improvement. In the LSTM configuration, units are set to the number of time steps on one day (288 5-min slots in 24 h). For MWCR and E-MWCR, the same inputs as in the 1-h data are used but at 5-min time steps.

Testing at the 5-min time step increased the forecasting error in all testing days because of the sudden PV fluctuations occurring within minutes. These fluctuations are difficult to capture in time series prediction like LSTM. Also, they make weather conditions forecasting more difficult which will affect the performance of MWCR model. MWCR outperforms LSTM, and E-MWCR shows improved accuracy over the MWCR on all testing days. On

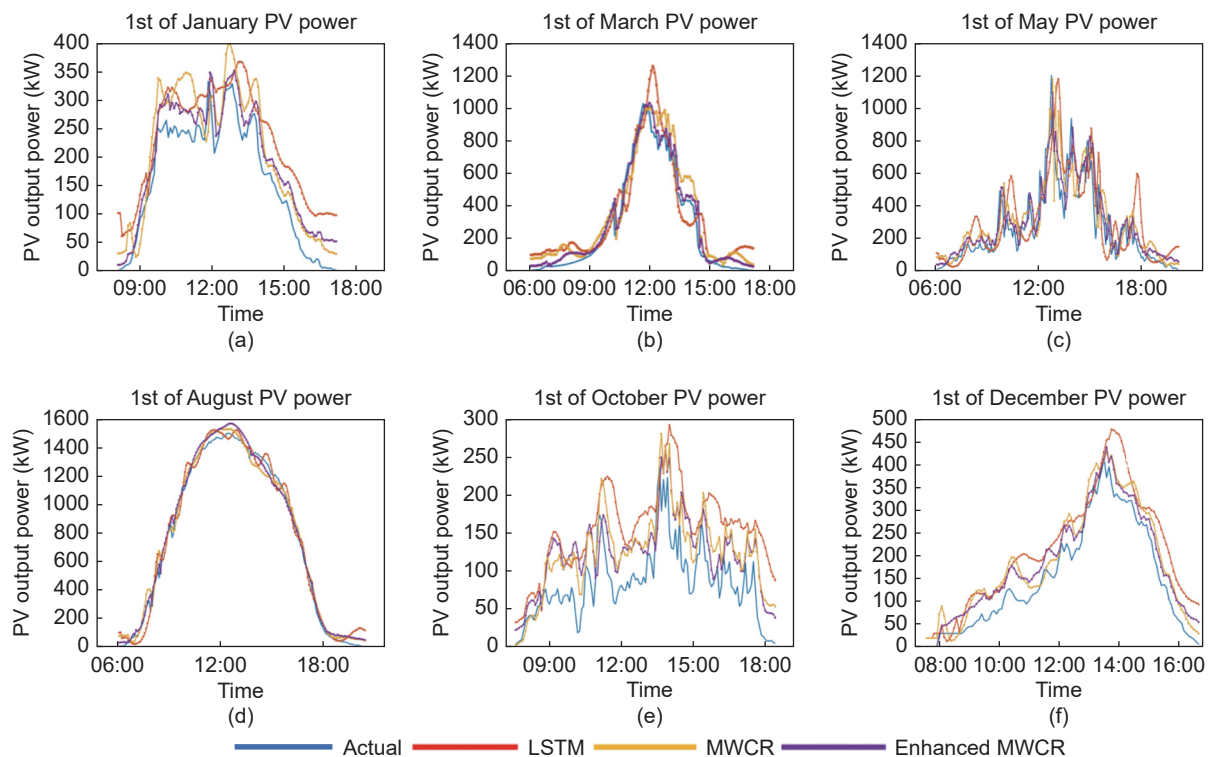


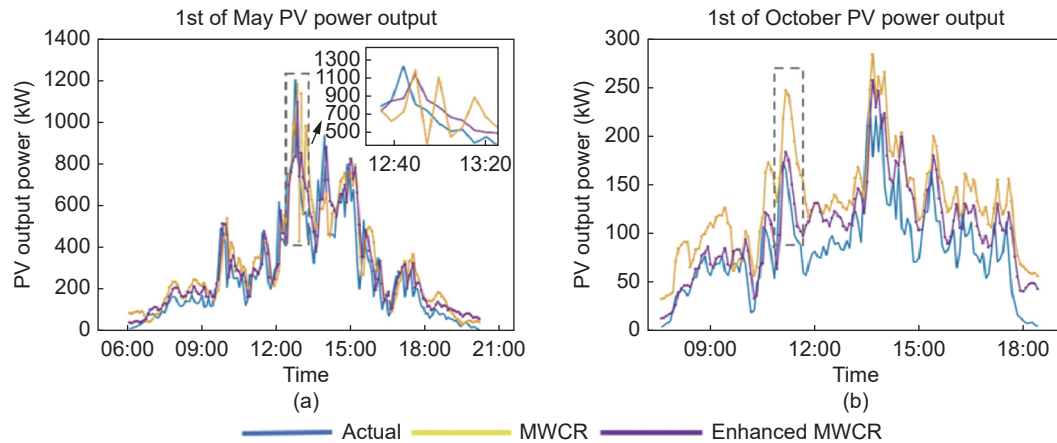
Fig. 10 Actual and predicted power in the first day of (a) January, (b) March, (c) May, (d) August, (e) October, and (f) December in the year of 2021 using LSTM and MWCR at the 5-min actual weather data.

Table 6 Performance comparison between single regression, CNN, GRU, LSTM, MWCR, and E-MWCR using the 5-min actual weather data

Model	r (%)		MAE (kW)		MAPE (%)		RMSE (kW)	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Single regression	79.40	81.12	123.60	115.82	198.41	51.10	157.23	212.20
CNN	80.30	82.90	120.50	116.40	177.42	48.98	149.26	195.59
GRU	80.50	83.76	117.50	110.22	160.54	46.21	136.49	154.67
LSTM	80.91	86.74	112.40	93.41	111.90	35.40	127.70	168.23
MWCR	86.67	90.11	82.15	51.94	40.23	24.67	88.95	98.87

Table 7 Performance comparison between LSTM, MWCR, and E-MWCR on different days using the 5-min actual weather data

1st day	r (%)			MAE (kW)			MAPE (%)			RMSE (kW)		
	LSTM	MWCR	E-MWCR	LSTM	MWCR	E-MWCR	LSTM	MWCR	E-MWCR	LSTM	MWCR	E-MWCR
January	70.79	95.91	97.95	66.28	48.18	34.61	97.92	49.04	32.158	72.60	55.36	39.03
March	92.58	97.11	98.54	102.91	83.35	43.87	89.76	33.40	29.45	127.50	95.59	61.28
May	76.52	86.49	90.66	128.66	93.07	75.03	117.06	68.73	63.52	174.32	131.46	105.97
August	95.57	99.67	99.73	52.303	39.86	38.45	21.96	17.68	13.57	63.92	49.54	45.95
October	73.89	79.67	86.54	72.15	43.26	42.55	176.50	93.87	67.31	80.38	49.18	48.47
December	72.50	98.08	98.77	59.35	38.97	34.01	86.93	67.31	48.43	68.44	44.18	37.95

**Fig. 11** Actual and predicted power in the first day of (a) May and (b) October in the year of 2021 using MWCR and E-MWCR at the 5-min actual weather data.

winter days with variable weather conditions, as in the first days of January and December, as shown in Table 6, LSTM records less accuracy: r is 70.79% and 72.5%; The MWCR model improves the accuracy significantly: r increases to 95.91% and 98.08%. In days with seasonal changes such as in the first days of May and October when there are more fluctuations in the weather conditions, MWCR enhanced the prediction accuracy over the LSTM as shown in Table 7: r increases from 76.52% to 86.49% and from 73.89% to 79.67%. These days show rapid fluctuations in the PV output power as shown in Figs. 10(c) and 10(e). From 17:00 to 18:20 of 1st of May, LSTM power shows some spikes that are much higher than actual power by more than 200 W (210%).

The predicted power obtained by the MWCR can be further improved by the enhanced model as to be shown next. As shown in Table 6, the E-MWCR model improved the prediction accuracy for all days. There is a significant improvement during days with seasonal changes such as the first days of May and October when more PV power fluctuations are recorded. During the first day of May, the enhanced model shows better performance as compared to MWCR as shown in Table 7: r increases from 86.49% to 90.66%. In addition, during the first day of October, r increases from 79.67% to 86.54%. As shown in Fig. 11(a), from 12:20 to 13:20 the MWCR predicted power shows some spikes that are far away from the actual power, while E-MWCR predicted power is close to the actual power. In addition, as shown in Fig. 11(b), from 11:25 am to 11:55 am the MWCR predicted power is much higher than actual power, e.g., the predicted power at 11:40 am is higher than the actual by 150 kW (106%). The E-MWCR predicted power is closer to the actual power with a difference of less than 10 kW.

4.5.3 Prediction results using the 5-min actual data with noise

Testing using the actual weather data is impractical in real life since the forecasting error in the input data will affect the

prediction model performance. Unfortunately, the historical forecasted weather data are not available at the 5-min time step. The future forecasted weather data are available at the 15-min timestep form Ref. [42] collected manually for June 2024. However, the PV power data are available only starting from the middle of 2018 till the end of 2021 every 5 min, so the historical forecasted 15-min June data of the year 2024 cannot be used for testing. Testing requires a continuous three-month PV dataset as input for the LSTM layer. To test the model performance within minutes of PV power fluctuations, a noise is added to the 5-min actual weather data of June 2021 to reflect the forecasting error. To avoid adding random noise, the error is added to the 5-min data in a way that matches the forecasting error that occurred in the 15-min weather data of June 2024. The actual and predicted weather conditions are manually collected for June 2024 every 15 min. The 15-min data are used to generate noise in the 5-min data for days with close PV deviations. The model is trained using the same actual 5-min dataset used in Section 4.5.3 and tested only for June 2021 month using the actual 5-min weather dataset with the added noise from June 2024 data. The same LSTM model configuration and MWCR model inputs are used as in Section 4.5.2. The error metrics are calculated for LSTM, MWCR, and E-MWCR and presented in Table 8.

As shown in Table 8, E-MWCR improved prediction accuracy and increased r by 6% compared to LSTM and 4% compared to MWCR. The model with the noised 5-min data is compared to

Table 8 Performance comparison between LSTM, MWCR, and E-MWCR on different days using the 5-min actual data with noise

Model	r (%)	MAE (kW)	MAPE (%)	RMSE (kW)
LSTM	90.37	53.23	30.12	66.29
MWCR	92.38	41.47	23.56	54.28
E-MWCR	96.38	36.18	21.45	49.27

the E-MWCR model with the actual 5-min weather data during sunny, rainy, and a day with variable weather. Results are presented in Table 9.

The E-MWCR model is still efficient in predicting PV power in variable weather conditions using noised weather data. On a sunny day, prediction using noised or actual 5-min weather data did not show much difference since during these days the forecasting error of the weather data are expected to be low, and thus a close weather data input is used in both models. On a rainy day, a 6% difference in r between the noised and actual weather data are recorded since more forecasting errors and noise are added to the 5-min actual weather data that results in higher PV prediction error. The most significant difference occurs during a day with variable weather conditions. A r of 80.67% is recorded using noised weather data compared to 87.47% using the actual weather data as an input.

During days with highly variable weather conditions, more forecasting error is expected in the weather data and hence more prediction error in the PV power.

5 Conclusions

Forecasting the PV output power is challenging due to the intermittent nature of weather conditions while existing models cannot provide accurate predictions during seasonal changes. To address this, a multi-layer classification-based regression model (MWCR) is developed, where for each weather class a more accurate regression model is developed. The model consists of four layers: the physical layer, the power-to-power prediction layer using LSTM, the KNN weather classifier layer, and the GRNN regression layer. To further improve the prediction accuracy in days with erratic weather, an enhanced model (E-MWCR) is developed by adding LSTM power to the regression model. With the actual PV power collected from PV panels at Rochester Institute of Technology-New York for 3.5 years, the single regression, LSTM, MWCR, and E-MWCR models are compared. Models are tested using three different datasets at 1-h and 5-min timesteps. E-MWCR shows the best performance on all testing days, especially on days with rapid weather condition fluctuations. The future work will focus on grid optimization operation with the forecasted PV power by using the E-MWCR model to reduce operation costs and greenhouse gas emissions.

Acknowledgements

Micatu Inc. donated its groundbreaking grid view optical sensors to the Rochester Institute of Technology to help in data collection for microgrid operation optimization future work. This work is based upon work supported by the National Science Foundation (NSF) under Award ESSC-2340095. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Table 9 Prediction accuracy on days with different weather conditions

Weather	Actual	Noised	Actual	Noised
Sunny	99.67	95.77	93.65	108.59
Rainy	93.46	87.49	78.54	95.18
Variable	87.47	80.67	134.32	185.46

References

- [1] Arnell, N. W. Climate change and global water resources. *Global Environmental Change* Vol. 9, S31–S49, 1999.
- [2] Arnell, N. W.; Cannell, M. G. R.; Hulme, M.; Kovats, R. S.; Mitchell, J. F. B.; Nicholls, R. J.; Parry, M. L.; Livermore, M. T. J.; White, A. The consequences of CO₂ stabilisation for the impacts of climate change. *Climatic Change* Vol. 53, No. 4, 413–446, 2002.
- [3] Feldman, D.; Zuboy, J.; Dummit, K.; Stright, D.; Heine, M.; Mirlletz, H.; Margolis, R. Winter 2024 Solar Industry Update. 2024. Available at <https://www.nrel.gov/docs/fy24osti/88780.pdf>
- [4] Adam, V. Renewables made up half of the net electricity capacity added last year. *The Guardian* 2017.
- [5] Hosseini, S. E. An outlook on the global development of renewable and sustainable energy at the time of COVID-19. *Energy Research & Social Science* Vol. 68, Article No. 101633, 2020.
- [6] McEvoy, A. J.; Markvart, T.; Castañer, L. *Practical Handbook of Photovoltaics: Fundamentals and Applications*. Amsterdam: Academic Press, 2012.
- [7] Raya-Armenta, J. M.; Ortega, P. R.; Bazmohammadi, N.; Spataru, S. V.; Vasquez, J. C.; Guerrero, J. M. An accurate physical model for PV modules with improved approximations of series-shunt resistances. *IEEE Journal of Photovoltaics* Vol. 11, No. 3, 699–707, 2021.
- [8] Ammar, R. B.; Ammar, M. B.; Oualha, A. Comparative study among physical models and artificial intelligence methods for PV power forecasting. In: Proceedings of the 18th International Multi-Conference on Systems, Signals & Devices, 1038–1046, 2021.
- [9] Kalogirou, S.; Senc, A. Artificial intelligence techniques in solar energy applications. In: *Solar Collectors and Panels, Theory and Applications*. Manyala, R. Ed. Rijeka, Croatia: Sciyo, 2010.
- [10] Sayegh, A. S.; Munir, S.; Habeebullah, T. M. Comparing the performance of statistical models for predicting PM₁₀ concentrations. *Aerosol and Air Quality Research* Vol. 14, No. 3, 653–665, 2014.
- [11] Hossain, M. S.; Mahmood, H. Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. *IEEE Access* Vol. 8, 172524–172533, 2020.
- [12] Sansa, I.; Boussaada, Z.; Mazigh, M.; Bellaaj, N. M. Solar radiation prediction for a winter day using ARMA model. In: Proceedings of the 6th IEEE International Energy Conference, 326–330, 2020.
- [13] Zhang, N.; Tian, S. Y. The institution of the hourly solar radiation model. *Acta Energetica Sinica* Vol. 18, No. 3, 273–277, 1997.
- [14] Ramaswamy, S.; Sadhu, P. K. Forecasting PV power from solar irradiance and temperature using neural networks. In: Proceedings of the International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions), 244–248, 2018.
- [15] Sauer, K. J.; Roessler, T.; Hansen, C. W. Modeling the irradiance and temperature dependence of photovoltaic modules in PVsyst. *IEEE Journal of Photovoltaics* Vol. 5, No. 1, 152–158, 2015.
- [16] Zakzouk, A. K. M.; Alamoud, A. R. M.; Khoshaim, B. H. Factors affecting the performance of a photovoltaic power system (PVPS). *International Journal of Solar Energy* Vol. 5, No. 2, 67–81, 1987.
- [17] Chaichan, M.; Kazem, H. A. Experimental analysis of solar intensity on photovoltaic in hot and humid weather conditions. *International Journal of Scientific and Engineering Research* Vol. 7, 91–96, 2016.
- [18] Park, N. C.; Oh, W. W.; Kim, D. H. Effect of temperature and humidity on the degradation rate of multicrystalline silicon photovoltaic module. *International Journal of Photoenergy* Vol. 2013, Article No. 925280, 2013.
- [19] Jena, S.; Naik, A. K.; Neha; Panigrahi, C. K.; Sahu, P. K. Impact of environmental factors on the performance of solar PV cells: An experimental study. In: Proceedings of the 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology, 1–5, 2021.
- [20] Hosenuzzaman, M.; Hasanuzzaman, M.; Rahim, N. A.; Selvaraj, J. Factors affecting the PV based power generation. In: Proceedings of the 3rd IET International Conference on Clean Energy and Technology, 20, 2014.
- [21] Samal, K. B.; Bisoyi, A. Investigation of environmental effects on the performance of solar PV modules. In: Proceedings of the

- International Conference on Emerging Frontiers in Electrical and Electronic Technologies, 1–5, 2020.
- [22] Mathur, A.; Singh, N.; Dhaked, D.; Gupta, H.; Chauhan, R. Comparative study of regression models for PV power prediction. In: Proceedings of the XIX International Scientific Technical Conference Alternating Current Electric Drives, 1–5, 2023.
- [23] Alfadda, A.; Adhikari, R.; Kuzlu, M.; Rahman, S. Hour-ahead solar PV power forecasting using SVR based approach. In: Proceedings of the IEEE Power & Energy Society Innovative Smart Grid Technologies Conference, 1–5, 2017.
- [24] Shaban, N. A.; Abdelhafez, E.; Hamdan, M.; Saber, M.; Aboushi, A. A. Estimation of power produced by PV generator using weather data. In: Proceedings of the 12th International Renewable Engineering Conference, 1–5, 2021.
- [25] Fentis, A.; Bahatti, L.; Mestari, M. Network (GRNN) for solar panel power estimation. In: Proceedings of the International Conference on Information and Communication Technology Convergence, 294–299, 2020.
- [26] Serfa Juan, R. O.; Kim, J. Implementation of generalized regression neural network (GRNN) for solar panel power estimation. In: Proceedings of the International Conference on Information and Communication Technology Convergence, 294–299, 2020.
- [27] Gupta, R.; Yadav, A. K.; Jha, S.; Pathak, P. K. Time series forecasting of solar power generation using Facebook Prophet and XGBoost. In: Proceedings of the IEEE Delhi Section Conference, 1–5, 2022.
- [28] Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Computation* Vol. 9, No. 8, 1735–1780, 1997.
- [29] Mellit, A.; Pavan, A. M.; Lughi, V. Deep learning neural networks for short-term photovoltaic power forecasting. *Renewable Energy* Vol. 172, 276–288, 2021.
- [30] Raza, M. Q.; Nadarajah, M.; Ekanayake, C. On recent advances in PV output power forecast. *Solar Energy* Vol. 136, 125–144, 2016.
- [31] Kardakos, E. G.; Alexiadis, M. C.; Vagropoulos, S. I.; Simoglou, C. K.; Biskas, P. N.; Bakirtzis, A. G. Application of time series and artificial neural network models in short-term forecasting of PV power generation. In: Proceedings of the 48th International Universities' Power Engineering Conference, 1–6, 2014.
- [32] Zhao, X.; Wu, C. Weather classification based on convolutional neural networks. In: Proceedings of the International Conference on Wireless Communications and Smart Grid, 293–296, 2021.
- [33] Mali, Y.; Kurlekar, A.; Lalsinge, S. Prediction and classification of weather using machine learning. *International Journal for Research in Applied Science and Engineering Technology* Vol. 10, No. 12, 809–814, 2022.
- [34] Naik, N. S.; Pathan, R. K. Weather classification and forecasting using feed-forward artificial neural network. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 3389–3394, 2012.
- [35] Shi, J.; Lee, W. J.; Liu, Y.; Yang, Y.; Wang, P. Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Transactions on Industry Applications* Vol. 48, No. 3, 1064–1069, 2012.
- [36] Agarwal, V.; Singh, V.; Gaur, P.; Agarwal, R. PV output forecasting based on weather classification, SVM and ANN. *Indian Journal of Engineering and Materials Sciences* Vol. 29, No. 2, 211, 2022.
- [37] Wang, X.; Sun, Y.; Luo, D.; Peng, J. Comparative study of machine learning approaches for predicting short-term photovoltaic power output based on weather type classification. *Energy* Vol. 240, Article No. 122733, 2022.
- [38] Tiwari, G. N.; Dubey, S. *Fundamentals of Photovoltaic Modules and Their Applications*. Cambridge, UK: The Royal Society of Chemistry, 2009.
- [39] Jenkins, A. The Sun's position in the sky. *European Journal of Physics* Vol. 34, No. 3, 633–652, 2013.
- [40] Specht, D. F. A general regression neural network. *IEEE Transactions on Neural Networks* Vol. 2, No. 6, 568–576, 1991.
- [41] Kiley, R. Solar energy farm to begin operation in April: 2-megawatt solar array among the largest of any college in New York state. 2015. Available at <https://www.rit.edu/news/solar-energy-farm-begin-operation-april>
- [42] National Centers for Environmental Information Climate data online (CDO). 2024. Available at <https://www.ncei.noaa.gov/cdo-web/>
- [43] Accuweather Minutecast Rochester NY. 2024. Available at <https://www.accuweather.com/en/us/rochester/14614/minute-weather-forecast/329674>



operation optimization.



Department of Electrical and Microelectronic Engineering, Rochester Institute of Technology, Rochester, NY, USA. Her research interests include power and energy system operation, manufacturing system scheduling, and mathematical optimization.



Katie McConky is a professor in the Department of Industrial and Systems Engineering at Rochester Institute of Technology. She earned her Ph.D. degree in industrial engineering in 2013 from the SUNY Buffalo. Her long-term research goal is to integrate operations research and machine learning to solve complex problems in energy efficiency and healthcare.