Short Term Performance Investigation of Solar PV Module: A Machine Learning Based Approach

Sabbir Ahmed
Department of EEE
Brac University
66 Mohakhali, Dhaka-1212, Bangladesh
sabbir.ahmed2@g.bracu.ac.bd

Md. Kamrul Islam
Department of EEE
Brac University
66 Mohakhali, Dhaka-1212, Bangladesh
md.kamrul.islam@g.bracu.ac.bd

Mohaimenul Islam
R&D Laboratory,Department of EEE
Brac University
66 Mohakhali, Dhaka-1212, Bangladesh
mohaimenul.islam@bracu.ac.bd

Md. Mosaddequr Rahman
Department of EEE
Brac University
66 Mohakhali, Dhaka-1212, Bangladesh
mosaddeq@bracu.ac.bd

2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC) | 978-1-7281-1110-0/20/$31.00 ©2020 IEEE | DOI: 10.1109/R10-HTC49770.2020.9357027

***Abstract*—This study presents a short-term performance analysis of the photovoltaic (PV) module considering weather impact in the context of Bangladesh by using machine learning. A Multilayer perceptron model is used to analyze the data and to predict the output. To collect the weather data and the output data, a weather station has been developed and deployed on the rooftop of a 7-story building in Gabtoli, Dhaka, Bangladesh. All the sensor data can be accessed remotely. In this study data from 1st November 2019 to 28th February 2020 are used in four separate data set for training purpose. It is observed that the output energy prediction improves with the increase in training data. The result shows that the temperature has the highest linear correlation with the module short circuit current among all the weather parameters i.e. humidity, wind speed, and air pressure.**

Keywords—short circuit current, irradiation, artificial neural network, multilayer perceptron, raspberry pi, energy.

I. Introduction

The world is developing every second in every sector and even within a blink of an eye. This overall improvement requires huge energy to meet the demands. The hike in energy demand has led to the crisis of conventional energy resources such as oil, gas and many more. Moreover, these conventional resources would take more than a human lifetime to reform naturally [1]. In light of this notion, experts recommend shifting towards the sustainable energy source, where solar energy is the most reliable source to meet the escalating energy demand.

The dependency on solar energy is increasing at a rapid pace around the world. For example, in 2019 China added new installations of 30.1 GW, US with 13.3 GW, India with 8.8 GW, Japan with 7 GW and Vietnam with 6.5 GW [2]. Despite its tremendous positive features, the energy produced from solar energy lags behind due to the constraint of converting solar energy into electrical energy in a productive way. The uncontrollable solar irradiance and weather parameters are the main key features which add up the difficulty to produce and distribute solar energy on a large-scale. However, to increase the reliability of the renewable energy-based power system, escalating the revenue and customer satisfaction different forecasting models are applied to predict the energy output [2,3].

The application of Artificial Neural Network (ANN) in the field of solar forecasting have been widely used due to its ability in dealing with non-linear problems [4]-[6]. Study shows that the most used short-term forecasting algorithm includes neural network, non-linear regression algorithms, time series algorithms, wavelet analysis, and random forest. Among them, neural networks and non-linear regression algorithm are the most often used models in PV power prediction [7].

The paper focuses on the short-term forecasting of solar PV output in the context of Dhaka, Bangladesh. This work uses the Multi-Layer Perceptron (MLP) model which is a class of feed-forward ANN. It has been implied based on the data collected over the four months from November 2019 to February 2020, to predict and analyze the daily performance of the system installed in Gabtoli, Mirpur, Dhaka, Bangladesh. The set-up is depicted in Fig. 1. It considers the weather factor i.e. humidity, wind speed, air pressure and surface temperature of the PV module.

The rest of the paper is organized as follows: section II, provides the theoretical background, section III presents the experimental setup and method, results and analysis is presented in section IV. Finally, section V makes concluding remarks.

(a) (b)

Fig. 1. (a) Location coordinate and (b) real image of the weather station deployed on the rooftop of a 7-story building in Gabtoli, Mirpur, Dhaka, Bangladesh.

II. Theoretical Background

A. Irradiation Calculation

For calculating the solar irradiance (W/m^2) of a particular day, the empirical method is used based on the theoretically found irradiance and experimental short-circuit current of that particular day. As it is known that, module short circuit current proportionally changes with the irradiance. By integrating this solar irradiance over a given time period results in cumulative incident energy (Wh/m^2). This study is dependent on fixed axis photovoltaic module.

Irradiance for fixed axis tracking can be calculated by (1) [8],

$$I = I_o * \cos\delta * cos\,\theta_z \quad (1)$$

Here, δ is the declination angle and θ_z is the Zenith Angle.

The declination angle is the angle of deviation of the sun from directly above the equator, i.e., zero degree latitude. To find declination angle (2) is used [8].

$$\delta = 23.45°\sin\left\{\frac{360}{365}(n+284)\right\} \quad (2)$$

Here, n= nth day of the year (i.e. January 1st means n=1)

The angle between the zenith and the center of the Sun's disc is known as the Zenith Angle(θ_z). The direction of reference for measuring the zenith angle (θ_z) is known as the Zenith. The Zenith angle (θ_z) can be expressed as

$$\theta_z = 90^{\circ} - \alpha \quad (3)$$

Irradiance at any given day of the year is then found by (4) and it is considered as direct normal irradiance (DNI). This is dependent on air mass [8].

$$I_o = 1367 * 0.7^{AM^{0.678}} \quad (4)$$

1367 W/m^2 is solar irradiance in space. AM is air mass given by (5).

$$AM = sec\theta_z = \csc\alpha \quad (5)$$

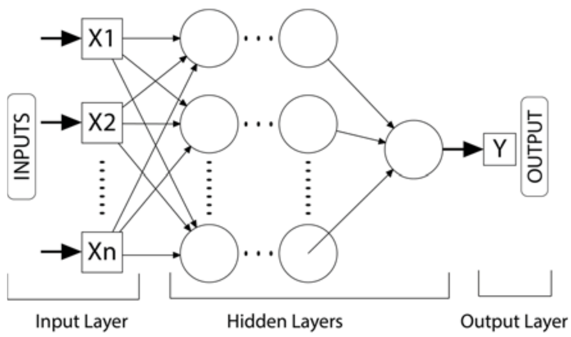

Fig. 2. Multilayer perceptron model architecture.

Solar altitude angle, α can be found from (6) [8],

$$\alpha = \sin^{-1}(\sin\delta\sin\gamma + \cos\delta\cos\gamma\cos\omega) \quad (6)$$

The cumulative incident energy (E) can be obtained by integrating the incident solar irradiance over the total sunlight hours (T) [8].

B. Machine Learning Model

Artificial Neural Networks (ANN) is an applied phenomenon of the human central nervous system in the field of data science which creates an environment for the machine learning model to learn and further process the data. The application of ANN in the field of solar PV power forecasting is very high. The feed-forward neural network manipulates the information in two phases. First, the training process is done to adjust the ANN weights. The network is then generalized to match the training data for the conclusion of the unknown samples. The training of data repeats the learning process and updates the weights until the desired output is achieved [9].

Multilayer perceptron (MLP) is a type of ANN as shown in Fig. 2 with input and output layer, and several hidden

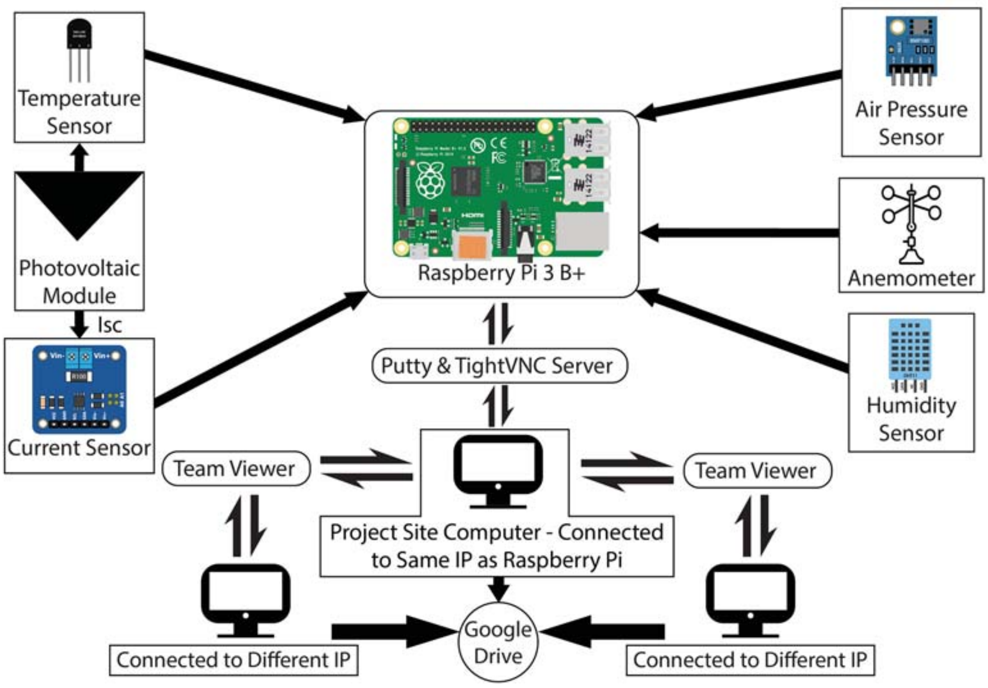

Fig. 3. Block diagram of the hardware set-up along with the data collection process.

layers. The input layer distributes the input values to the hidden layers. The output of the first layer, which is the sum of the inputs multiplied by its weights can be determined by (7) [9].

$$Y_i = f\ (\Sigma w_i\ x_i) \tag{7}$$

Where x_i is an input signal, w_i is the strengths weighting and Y_i is the output as a function of the sum.

III. Experimental Setup And Methodology

The Experimental setup of this study is up and running since November 2019. It was installed on a 7-story building around Gabtoli, Mirpur, Dhaka, Bangladesh. The setup was designed to take the data of different weather parameters and short circuit current data of Photovoltaic Module. Surface temperature of PV module, wind speed, humidity and air pressure are the weather data that were collected by this setup. For training the model, training datasets and testing datasets were made out of the data that were collected by each sensor. In this study, 29th February 2020 is selected for the testing data set. The training data set are as follows:

- Training Dataset 1: 1st Nov 2019 to 30th Nov 2019
- Training Dataset 2: 1st Nov 2019 to 31st Dec 2019
- Training Dataset 3: 1st Nov to 31st Jan 2020
- Training Dataset 4: 1st Nov to 28th February 2020

A. Hardware Setup

For setting up the hardware, a current sensor, a temperature sensor, an anemometer, a humidity sensor and a barometric pressure sensor in total five sensors were used as in Fig. 3. The data collected by each of the sensors get stored in different data loggers with the help of the built-in real-time clock in the Raspberry Pi 3 B+. This way, the data were preserved according to the date and time. Each of the sensors communicates with the Raspberry Pi 3 B+ through a different communication protocol. The communication protocol between Raspberry Pi 3 B+ and temperature sensor, humidity sensor, and an anemometer is 1-wire bus communication protocol. Whereas, the communication protocol between Raspberry Pi 3 B+ and barometric pressure sensor, and current sensor is I2C Communication protocol.

B. Software Setup

After the Raspberry Pi 3 B+ collects each sensor's data and stores the data of each sensor in different data loggers, the data extraction process begins. The process is shown in Fig. 3. As Raspberry Pi has internet connectivity, the setup is connected through Ethernet. Through Ethernet, multiple computers can be connected together in a local area network or LAN. A dedicated computer is stationed in the setup location and also connected to this same network as the Raspberry Pi. On this laptop, two software known as "Putty" & "TightVNC" are installed and used to access the Raspberry Pi through Secure Shell (SSH) Network protocol. Also, Team Viewer is being used to make a connection with the computer dedicated to the setup for collecting data remotely. It helps to access from any other computer with internet connectivity and upload the data on Google Drive directly from the Raspberry Pi.

C. Empirical irradiance calculation method

The Empirical method is used to calculate the irradiance. By following the theory discussed in the irradiance calculation of the theory section of this study, the theoretical irradiance of each day of a year is calculated. For a specific sunny day, from

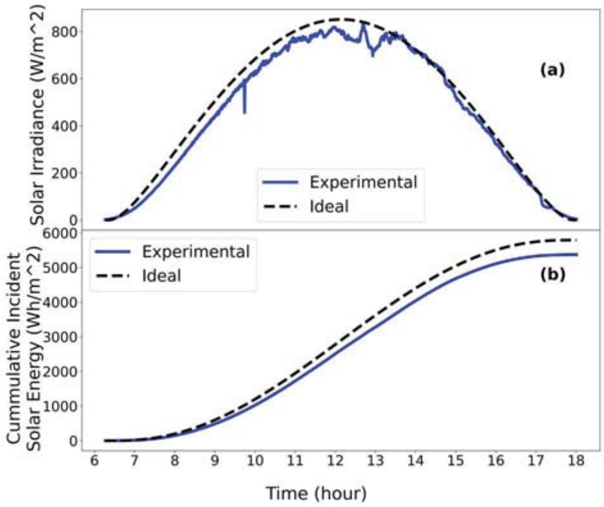

Fig. 4. Plots of (a) solar irradiance and (b) cumulative incident solar energy, calculated theoretically (dashed line) and those estimated using experimentally measured short circuit current (solid line), recorded on 29th February,2020.

the short-circuit current curve, the highest generated short-circuit current needs to be marked. Using this marked data and its corresponding time, from the theoretical irradiance curve of the same day at the same time the ideal irradiance needs to be picked. Now depending on this marked short-circuit current and picked ideal irradiance, empirically calculated irradiance can be found by unitary method. Cumulative Incident Energy (E) can be obtained by integrating the total irradiance over the total sunlight hours. Fig. 4 shows the ideal and empirically converted irradiation plots. In Fig. 4(a) theoretical or ideal irradiance (dashed line) is shown along with the experimental irradiance converted from short circuit current by taking the theoretical irradiance as reference. Fig. 4(b) shows the cumulative incident solar energy outcome.

D. Multi Layer Perceptron Model of ANN

The MLP is used because of its ability to approximate the function of the complex relationship between the input and the target or output parameters. The most important case is to get a meaningful output that depends on the model architecture. The model architecture is designed with three basic layers input layer, hidden layer and output layer. Following the convention mentioned in the book "Neural Smithing" it can be shown using the notation: 5/8/1. The size is 802 which shows the number of nodes in the model. The width is 100 that is the number of nodes in each layer and the depth of the neural network is 8. Rectified Linear Unit (ReLU) is used as the activation function which is able to solve non-linear transfer functions easily. The optimizer used in model architecture is called adamax. Optimizer is the key function of the models learning rate. Initially, the algorithm is fed with training dataset consisting weather parameters humidity, wind speed, air pressure and surface temperature of PV module. Moreover, the relative I_{SC} (mA) of the module is also fed as the target parameter and to establish a transfer function between the weather parameters by the learning model. The epoch was set at 250; it indicates the number of times the training dataset is passed through the neural network. To refrain over fitting, the method Early Stoppage is used with assigning a value of 10 as the patience parameter. Accomplishing the model training, the test dataset is given which includes a probable set of values of all the weather parameters relative to the day's I_{SC} (mA) to be

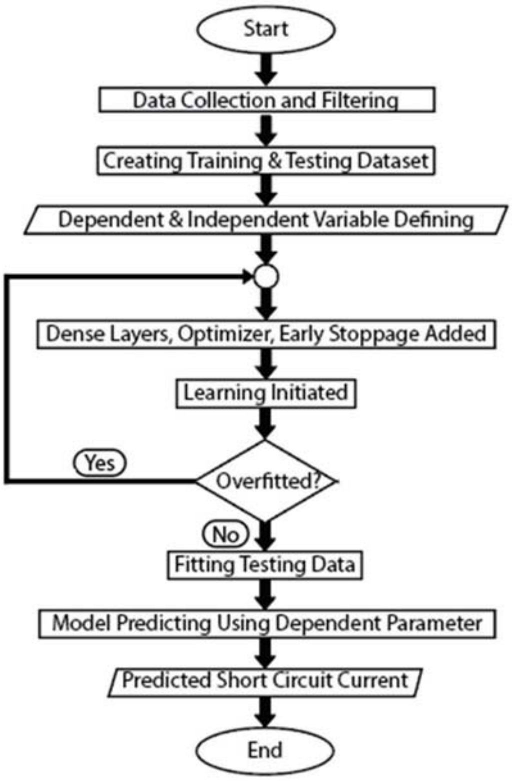

Fig. 5. Flow chart of the MLP model used in this study.

predicted by the algorithm. Lastly, the prediction results of I_{SC} (mA) is compared with the actual experimental data of the specified day. The procedure is repeated to all the four distinct training datasets consisting of data of a distinguished period of months. The overall process is shown in Fig. 5.

IV. Results Analysis And Discussion

This section discusses the results of this study. First, short circuit current data are trained and predicted for the months of November 2019 to February 2020 considering all the environmental parameters mentioned in the experimental setup part. Then, a similar study has been shown considering the temperature only. The data extracted over the period of months from 1st November 2019 to 28th February 2020 are used to predict the probable short circuit current using four training datasets of the distinct time period. Empirical relation is used to calculate the solar irradiance and cumulative incident solar energy mentioned in section III. For easy understanding, only 29th February is considered for test data throughout the analysis.

A. Prediction with all-weather parameters

Fig. 6 shows the plots of experimental and predicted short circuit current, I_{SC} (mA) of the PV module recorded and predicted for 29th February. All the data estimated by using four training datasets considering all weather parameters i.e. temperature, humidity, wind speed and air pressure. As the training data set is increased gradually from the Training data set 1 to Training data set 4, the predicted short circuit current (blue dashed line) shows variation from Fig. 6(a) to Fig. 6(d).

The solar irradiance (W/m^2) is calculated from short circuit current using the empirical method. Furthermore, the cumulative incident solar energy (Wh/m^2) is calculated by integrating the solar irradiance for each of the four distinguished training datasets. The data of the predicted

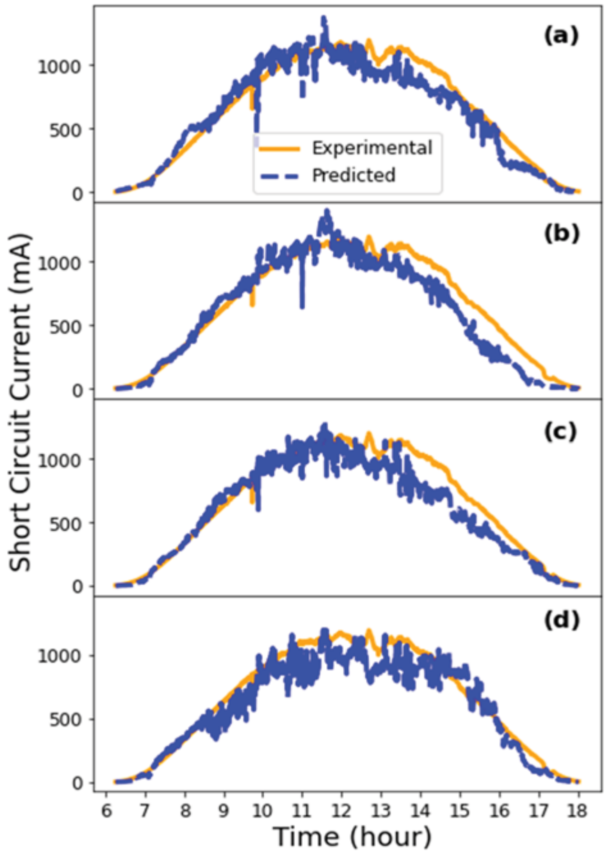

Fig. 6. Plots of experimental and predicted short circuit current of PV module on 29th February, estimated using (a) Training dataset 1, (b) Training dataset 2, (c) Training dataset 3, and (d) Training dataset 4; considering all weather parameters.

energy is compared with the experimental data (empirically calculated cumulative incident solar energy) of 29th February. Fig. 7(a) presents the energy differences and 7(b) shows the

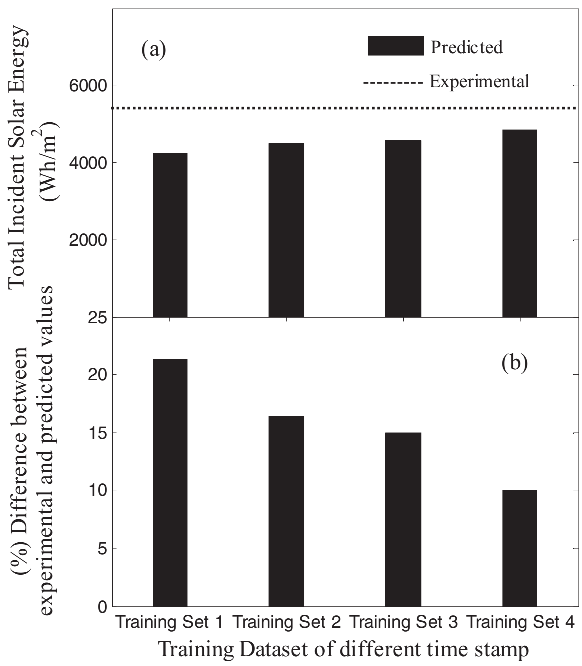

Fig. 7. (a) Comparison of experimental and predicted total incident solar energy for 29th February, estimated using four training data sets, (b) percentage difference between the experimental and predicted values.

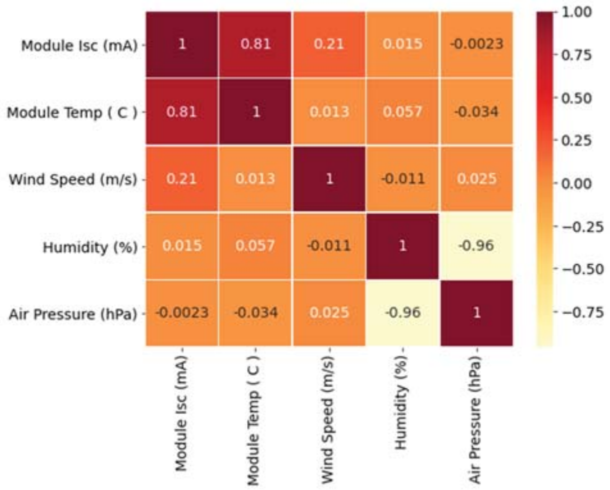

Fig. 8 Pearson Correlation between each weather parameter corresponding to short circuit current, *Isc* (mA) of the PV module.

percentage difference of the predicted energy with respect to the experimental energy data. As the training data is increased, the percentage of differences between the predicted energy and the experimental energy data has been decreased; which is depicted in Fig. 7(b). Fig. 7(b) illustrates that after including four months of data (1st November-28th February) as training data, the percentage difference of predicted data falls from 21.28% to 10.04%. It represents that predicted data gets closer to the experimentally found data after including training set 4.

Fig. 8 illustrates the Pearson correlation coefficient matrix of weather parameters corresponding to the module short circuit current. The weather parameters consider having a specific impact on the short circuit current, the output of PV module. The training dataset 4 is used in this case and the values of each feature corresponding to the other features is plotted using the heat map as shown in Fig. 8.

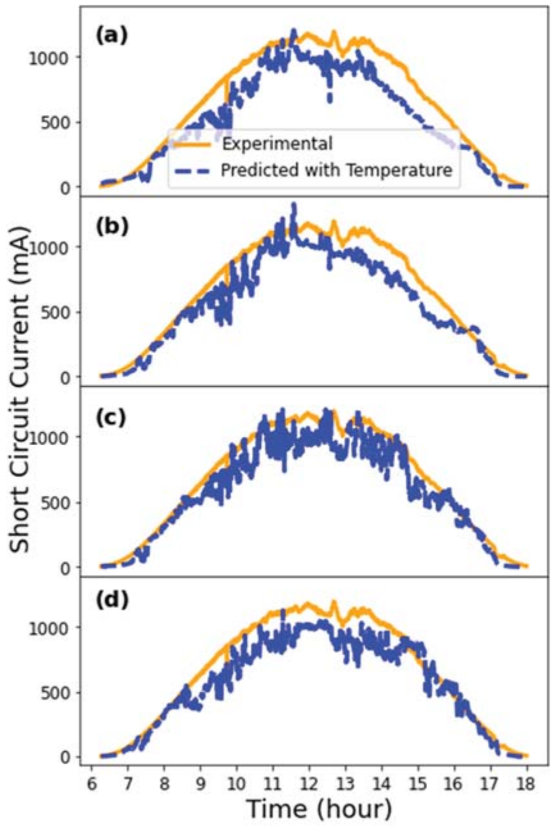

Fig. 9. Plots of experimental and predicted short circuit current of PV module for 29th February, predicted using (a) Training dataset 1, (b) Training dataset 2, (c) Training dataset 3, and (d) Training dataset 4; considering only temperature as the weather parameter.

The dark color represents a strong correlation while the lighter color depicts low correlation with each other. It measures only the linear relationship between two variables and ranges from -1 to +1. As an example, the short circuit current of the PV module shows the coefficient value 0.81 with temperature. That means I_{SC} is highly correlated with the temperature for the PV module. The second most correlated parameter is wind speed, which shows the value 0.21. This matrix confirms that the weather parameter Temperature (□) has the strongest correlation with the I_{SC} (mA) of the photovoltaic module.

In the next part of the analysis, only the temperature is considered to predict the output.

B. Prediction only with Temperature

Fig. 9 presents the plots of experimental and predicted short circuit current, I_{SC} (mA) of the PV module recorded and predicted for 29th February. All the data estimated by using four training datasets considering only the temperature data. As the temperature is highly correlated with the current. It is seen from Fig. 9(a) to Fig. 9(d) that the predicted plots do not closely follow the experimentally found data. The reason behind this, wind speed, humidity and air pressure impact are not considered here.

Fig. 10 (a) presents the predicted cumulative incident solar energy differences considering only temperature and all the four environment parameters in a bar chart. Fig. 10 (b) shows the percentage differences among the predicted energy. There is no significant differences (below 5%) between the energy predictions, considering all four parameters and considering

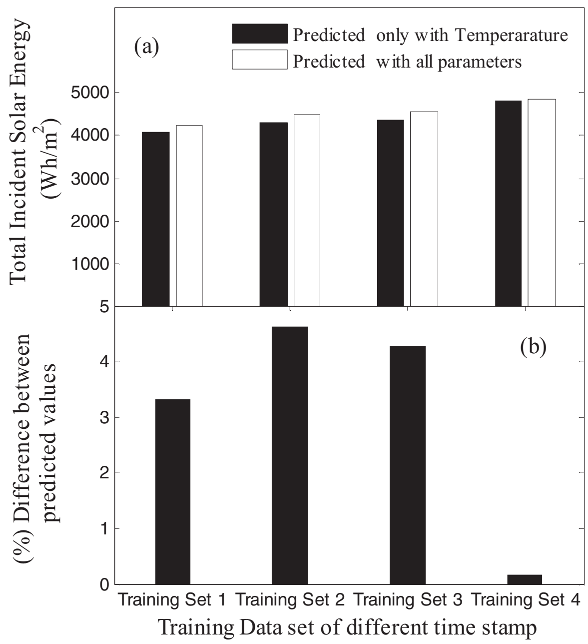

Fig. 10. (a) Comparison of total solar energies predicted considering only temperature with those predicted considering all weather parameters, estimated using four training data sets for 29th February and (b) percentage difference between the two.

temperature only for all the data set. Moreover, after using training data set 4, the result shows that the two predicted energy values are similar (difference below 0.5%) for the test day of 29th February. It emphasizes that for a short term data prediction, temperature shows the highest impact among all the other environmental parameters. In other words, it may say that ignoring other parameters will not create much impact in the short term energy prediction of PV module.

V. Conclusions

This study presents a short-term performance analysis of a PV module using machine learning. Data are taken from the month of November 2019 to February 2020. There are four training data set of the different timestamp. Collecting the short circuit current data, that are empirically converted to energy data. To observe the environmental impact, first, all the weather parameters are considered for the analysis. Then a similar analysis has been done only taking the temperature data.

The result shows that, if the training data set is increased, the difference between the experimental data and the predicted data is decreased. It shows that the percentage difference is decreased from 21% to 10% from training set 1 to training set 4. Another observation is that temperature data shows the highest linear correlation with the PV module output parameter; in this study short circuit current. That means other weather parameters have a negligible impact on module short circuit current. It is confirmed after taking the temperature into consideration as training data. The result depicts that the difference between energy predictions is below 0.5%.

In future work, yearly data will be analyzed to get a complete picture of the performance of this module. Also irradiation will be cross checked with the irradiation sensor to compare the empirical data. Besides, validation of the multilayer perceptron model will be analyzed.

Acknowledgment

Authors' acknowledged the technical support provided by the Research & Development (R&D) Laboratory, Department of EEE, Brac University.

References

[1] N. Sodsong, K. M. Yu and W. Ouyang, "Short-Term Solar PV Forecasting Using Gated Recurrent Unit with a Cascade Model," 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Okinawa, Japan, 2019, pp. 292-297.

[2] E. Du, N. Zhang, B. Hodge, Q. Wang, Z. Lu, and C. Kang, "Operation of a high renewable penetrated power system with csp plants: A lookahead stochastic unit commitment model." IEEE Transactions on Power Systems, vol. 34, pp. 140-151, Jan. 2019.

[3] R. Lahon, and C. P. Gupta, "Energy management of cooperative microgrids with high-penetration renewables." IET Renewable Power Generation, vol. 12, pp. 680-690, Apr. 2018.

[4] M. Simonov, M. Mussetta, F. Grimaccia, S. Leva, R. Zich, " Artificial intelligence forecast of PV plant production for integration in smart energy systems", International Review of Electrical Engineering, Vol. 7, No. 1, pp. 3454-3460, 2012.

[5] A. Gandelli, F. Grimaccia, S. Leva, M. Mussetta, E. Ogliari, "Hybrid model analysis and validation for PV energy production forecasting," 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, pp. 1957-1962, 2014.

[6] H. T. C. Pedro, and C. F. M. Coimbra, "Short-term irradiance forecastability for various solar micro-climates." Solar Energy, vol. 122, pp. 587-602, Dec. 2015.

[7] H. He, R. Hu, Y. Zhang, Y. Zhang and R. Jiao, "A Power Forecasting Approach for PV Plant based on Irradiance Index and LSTM," 2018 37th Chinese Control Conference (CCC), Wuhan, 2018, pp. 9404-9409.

[8] S. Ahmed, A. H. Zenan, N. Tasneem and M. Rahman, "Design of a solar powered LED street light: Effect of panel's mounting angle and traffic sensing," 2013 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (CSUDET), Selangor, 2013, pp. 74-79.

[9] Jabar H. Yousif, Hussein A. Kazem, Nebras N. Alattar, Imadeldin I. Elhassan, "A comparison study based on artificial neural network for assessing PV/T solar energy production", Case Studies in Thermal Engineering,vol. 13,2019