Indian Journal of Engineering & Materials Sciences
Vol. 29, April 2022, pp. 211-217

PV Output forecasting based on weather classification, SVM and ANN

Varun Agarwal[a], Vatsala Singh[a], Prerna Gaur[a], & Rashmi Agarwal[b*]
[a]Netaji Subhas Institute of Technology, University of Delhi, Delhi 110 078, India
[b]J C Bose University of Science & Technology, YMCA, Faridabad, Haryana 121 006, India

Received: 13 February, 2021 ; Accepted: 7 October, 2021

The expansion in solar power is expected to be dramatic soon. A number of solar parks with high capacities are being set up to harness the potential of this renewable resource. However, the variability of solar power remains an important issue for grid integration of solar PV power plants. Changing weather conditions have affected the PV output. Thus, developing methods for accurately forecasting solar PV output is essential for enabling large-scale PV deployment. This paper has proposed a model for forecasting PV output based on weather classification, using a solar PV plant in Maharashtra, India, as the sample system. The input data is first classified using RBF-SVM (Radial Basis Function Support Vector Machines) into three types based on weather conditions, namely, sunny, rainy and cloudy. Then, the neural network model corresponding to that weather type has been applied to forecast the solar PV output. The obtained results for the overall model is studied for its effectiveness and are compared with existing research.

Keywords: Photovoltaic systems, Solar radiation, Forecasting, Weather classification, Support vector machine, Neural network

1 Introduction

Solar energy is a freely available, clean form of energy that produces no toxic emissions, unlike fossil fuels. Photovoltaic (PV) power plants use PV modules that convert solar radiation directly into electricity It may or may not consist of a battery back-up [1].

By the end of 2019, the total installed solar power capacity across the world was 629 GW. The potential of solar power in the country has been assessed to be around 1250 GW. The Indian Government has undertaken a number of initiatives under the National Solar Mission to achieve its target of development and deployment of 20 GW by 2022. Different schemes have been launched and are in progress to increase the grid-connected solar PV projects [2]..

However, there are a number of challenges associated with grid-connected PV systems, which make their large-scale deployment difficult. One problem is the sudden decrease in solar PV output after sunset, causing a ramp in the deployment of conventional sources in order to compensate for this decline [3]. Another problem is the fluctuating solar PV output even during day-time, because of its dependence on meteorological factors. This intermittent nature of grid-connected PV causes problems associated with grid-reliability and ancillary generation.

*Corresponding author (E-mail: p19varun@iima.ac.in)

Accurately forecasting of solar PV output can help in bringing down the effect of solar output unreliability on the distribution network, thereby improving the system reliability, maintaining quality of power, and increasing penetration level of the solar PV systems. A variety of forecasting models have been proposed by different authors, taking different sets of parameters as input.

S. Netsanet et. al. [4]has used a RBFN (Radial Basis Function Network) forecasting model to generate 24-hour ahead PV system output which generated a forecasting MAPE (Mean Absolute Percentage Error) of 9.45% for the given plant.

Ramsami *et al*.[5] has used a stepwise regression feed forward neural network which resulted in a RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) of 2.74 and 2.09 respectively.

Bacher *et al*. [6]has used an autoregressive model for prediction of the average output power of rooftop PV systems which resulted in a RMSE improvement of around 35% over the previous models.

Shi *et al*. [7] has proposed an algorithm for forecasting the output power of PV systems based upon weather classification and support vector machine. In this, the weather conditions are first divided into four types, i.e. clear sky day, cloudy day, rainy day and foggy day. So its observed that the finest performance has shown by Sunny Model, in which the mean error value of 4.85%, accompanied

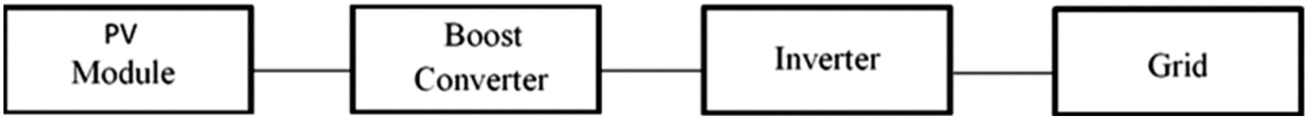

Fig. 1 — Block Diagram of Grid – Connected PV Syst[10].

by the other model such as Foggy Model, Rainy Model and Cloudy Model with the mean error value of 8.1%, 9.1% and 12.4% respectively[8].

Da Silva Fonseca *et al.* [9]has developed an SVM (Support Vector Machines) based technique to forecast the power production of a solar PV plant and obtained a RMSE and MAE of 9.48% and 5.8% respectively.

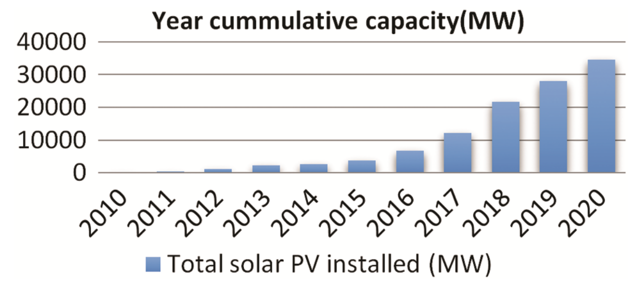

Fig. 2 — Total cumulative Grid Connected Solar Power Capacity in India till 31-12-2017[11].

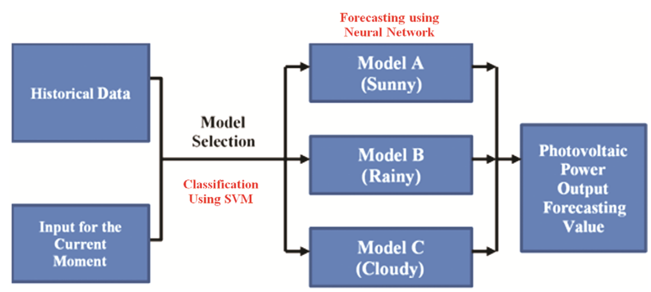

Fig. 3 — Flowchart for PV output forecasting.

2 Materials and Methods

In this paper, raw data from a solar power plant in Maharashtra, India, with a nameplate capacity of 125MW has been taken and pre – processed. Then, using this data, an output forecasting of PV, model has established based on different weather classification. The model is built using support vector machines and Neural Networks. . A grid-connected PV system consists of PV Module(s), DC-DC Boost convertor, invertor, and load, as its main elements has shown in Fig. 1. Figure 2 has shown the details of the cumulative installed solar capacity in different Indian states till December 2017. Hence, it can be clearly seen that there has been a rapid growth in the integrated PV system in the country, owing to its simple structure, together with preferential government policy. Figure 3 shows the flowchart of the PV output forecasting model.

The PV output power fluctuates according to the variations in solar radiation, which in turn is dependent on several meteorological factors. Additionally, the conversion efficiency of the PV cell is affected by its temperature. Thus, all these factors must be considered while developing a reliable PV output forecasting model.

2.1 Parameters Affecting Solar PV Output

The various input parameters used for training the PV Output Forecasting Model are described below:

- Irradiance: As the solar irradiance increases, the short circuit current and open circuit voltage increase and hence the maximum power point increases.
- Air Temperature: Higher temperature causes a reduction in the conversion efficiency. As temperature increases, the output current increases exponentially, thereby increasing heating effects. Therefore, the solar cells give their full performance on cold and sunny days, rather than hot and sunny days.
- Wind Speed: With increase in wind speed, the temperature of the PV Module decreases along with a reduction in its difference with the ambient temperature. Thus, with an increase in wind speed, there is a decrease in the module temperature and an increase in the module power[12].
- Relative Humidity: It has been observed that, as relative humidity decreases, the voltage, current and efficiency of PV module increases [13].
- Air Pressure: The output current and voltage of PV module rise with increase in air pressure. This happens because weight of air is gravitational, and this pull increases with decrease in altitude and exerts more downward pull on the (photons of) radiation from the sun. This causes the solar illuminance to increase, with simultaneous increase in the output current and voltage [14].

- Rainfall: Rain implies the presence of cloud cover which causes a decrease in PV module efficiency as clouds obstruct the amount of sunlight that can reach the PV module. However, occurrence of rainfall also helps to prevent soiling losses. Rainfall helps to clean the panels by washing away this accumulation, thereby increasing efficiency.
- Module Temperature: The temperature of the PV module is affected by all the above-mentioned factors, along with the tilt angle of the PV module. Normally, the inside temperature of the PV system is higher than the temperature of the surrounding environment.

The efficiency of the PV system at *t* time moment can be expressed as follows:

$$\eta = \eta_0[1 - \gamma(T_t - T_\gamma)] \quad ...(1)$$

where, T_t , is the temperature (at t time moment), reference temperature (298K) expressed as T_γ, $\eta_0$is efficiency of conversion at reference temperature and solar batteries temperature coefficient is expressed by γhere,whose value lies in between$(3*10^{-3})°C^{-1}$ and $(5*10^{-3})°C^{-1}$.

The PV system gives power output at time moment(t) is as follows:

$$P = I * A * \eta \quad ...(2)$$

where, Irepresents the radiation intensity of the PV inclined plane, Arepresents PV area(m^2), η is the rating conversion EFFICIENCY(kW/m^2).

The total power output, of PV array is given by, [15]:

$$P_{total} = n * P \quad ...(3)$$

where, n is number of PV cells working simultaneously at t time instant.

2.2 Forecasting of PV Power Output Model Based On Weather Classification

The output power and performance of a PV system are affected by a myriad of reasons and hence, it may become difficult to forecast the output using only a single pattern or model. Thus, based upon the unsteady and non-linear relationship between the solar PV output and the various affecting factors, we have developed a model in which the SVM (Support Vector Machines) classifier is first trained to classify the input data into one of three types among sunny, rainy and cloudy, to designate the corresponding weather condition to the type of model it belongs to. Each of the three weather types has its own neural network for forecasting the solar PV output. Once the classification is done, the corresponding neural network model works to forecast the power output. This can be clearly understood from Fig.3.

1.2.1 Data

For the purpose of research, the radiation angle and location of the PV Panel have been assumed to be fixed. The sample data for training and testing has been taken for a period of 3 months, starting from2017-05-01 to 2017-07-31.

The following parameters have been chosen as the inputs for training the classification model:

- Radiation
- Air Temperature
- Module Temperature
- Wind Speed
- Relative Humidity
- Air Pressure
- Rainfall

1.2.2 Weather Classification

In order to study the relation between the type of weather and the PV output, two plots are obtained.

In Fig. 4, PV output of the plant for two sunny days of each of the three months are plotted. It can be observed from the figure that graph pattern is the same for each of the 6 days. Thus, we infer from this figure, which has shown a high correlation in solar PV outputs of each of the sunny days, irrespective of the month of the year.

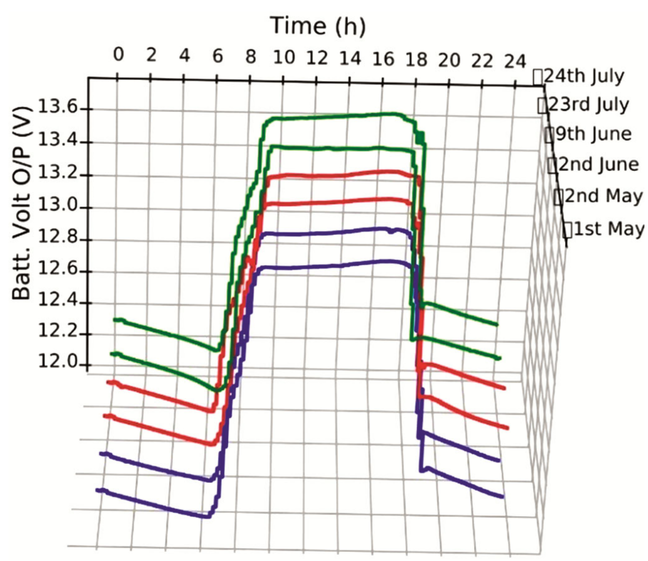

Fig. 4 — Daily PV system output for 6 sunny days.

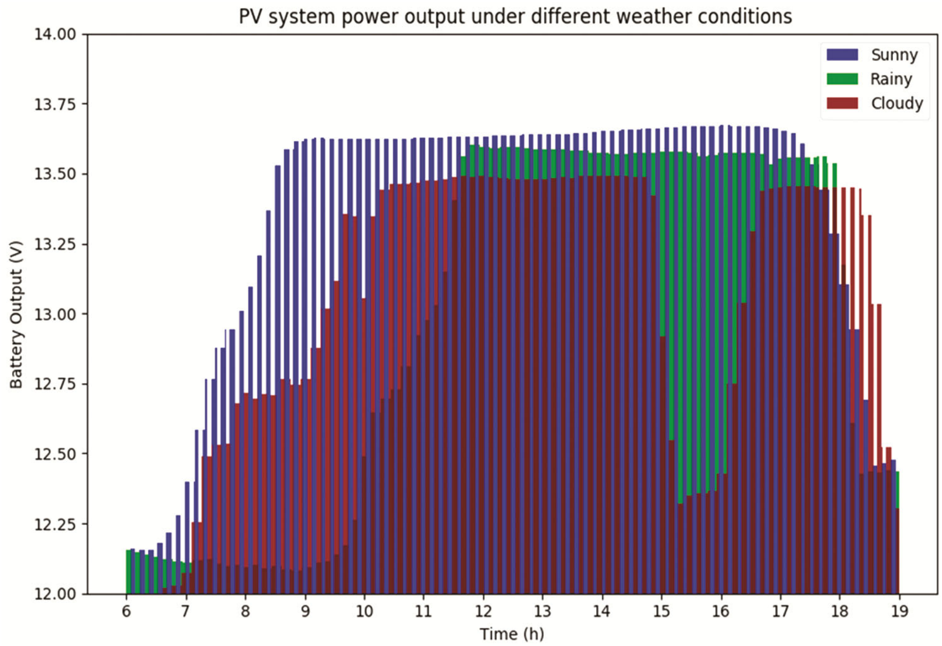

Fig. 5 — PV System output under 3 different weather conditions.

In Fig. 5, one day of each of the three weather types is chosen, and the PV output is plotted for these three days on a single graph. It can be observed from this figure that the solar PV output is maximum for a sunny day and that it is affected by changing weather conditions.

From Figs 4 and 5, it is clear that the solar PV output differ significantly under different weather conditions. This supports the proposal that the PV output model for forecasting should be in accordance with weather classification.

1.2.3 Data preprocessing

A classification model maps the non-linear input data into a higher space in order to make it linear and find a separating hyperplane that divides the dataset into various classes. Data that is sprayed across over a wide area will generate imprecise data fitting and thus will reduce the classification efficiency. Hence, the following data preprocessing steps are applied to get a more linearized dataset.

- Data Cleaning: In this step, the sample dataset is checked for missing and inconsistent values and the datasets outside the timeframe 6:00 AM to 7:00 PM are rejected because solar power output becomes zero during the night.
- Data Scaling: This forms an integral part of the data preprocessing scheme, as otherwise the classifier will depend much more on attributes whose scale is larger than others. To prevent this, we use the Z - Score Scaling method which gives an output between 0 and 1. It is expressed as follows:

$$Z = (X - \mu)/\sigma \quad \dots(4)$$

where, $X$is the value of the attribute, $\mu$is the mean and $\sigma$is the standard deviation.

1.2.4 Support Vector Machine Model

Support Vector Machine (SVM) is a machine learning algorithm that is based on the statistical learning theory. Several recent studies have reported that the SVM's can generally deliver higher performance in terms of classification-accuracy in comparison to other data-classification algorithms. They also have the advantage of having a good calculating speed and they do not suffer the limitations of data dimensionality and limited samples. The SVM model is set up using these steps as follows:

- To train the SVM model, historical data samples have been distributed into these 3 groups, in accordance with the given day's weather conditions such as Cloudy Day, Sunny Day, Rainy Day and two kernel functions are evaluated: Linear and RBF.
- The classifier performance is said to be directly proportional to the number of training data. The SVM model was tested for performance by dividing the data-set into training and test data into 80:20.

- Various tuning parameters for the SVM model are defined as C, γ and d. C represents the regularization parameter or the cost function. ϒ defines how far the influence of a single training example reaches, with low values meaning 'far' and high values meaning 'close' and d represents the width of the kernel. The values of the parameters are changed as per the selected model experience.
- After setting up the model by tuning the parameters and splitting the dataset appropriately, the SVM model is trained and the accuracy is obtained.

Table 1 depicts the accuracy obtained during the classification process using the SVM network with Linear and RBF(Radial Basis Function) kernels, with the Testing Dataset for each of the 3 months.

It can be observed that the RBF SVM network performs better than Linear SVM network for each of the 3 months. Thus, the RBF SVM model is finally selected for classification.

1.2.5 Neural Network Model

A neural network (NN)-based model has been used to predict the PV output power after the respective classification of the input data is done. A neural network is a parallel, distributed information processing structure consisting of processing elements, which can possess a local memory and can carry out localized information processing operations, interconnected together with unidirectional signal channels that are called connections. These structures are called neurons.Each neuron receives some signals from other neurons or from the input. Every neuron employs an activation function that fires when the total input is more than a given threshold.

In this paper, we focus on Multi Layered Perception (MLP) networks that are layered feed forward networks, typically trained through back propagation. A NN has 3 kinds of layers: an input layer, an output layer and a hidden layer[16-17].

The parameters for the Neural Network used are as follows:

- Number of Hidden Layers – 3
- Number of Hidden Neurons per Layer – 500
- Maximum number of Epochs – 100
- Activation Function – Relu
- Optimization Function – Adam Optimizer
- Performance Function – Mean Square Error

The computation graph obtained with these parameters is shown in Fig. 6.

The training constitutes the process of receiving back propagation errors from the output layer of the system towards its hidden layers. Back propagation is one of the main steps of the training process as the hidden units have no prior training target value that can be used, and hence they must be trained based on the errors from the previous layers. Output layer is the only layer which has a target value which can be compared, in order to obtain the error. The training process is continued until all the differences between the updated weights and the old weights calculated in the previous epoch, are below the set threshold value.

1.2.6 PV system forecasting Accuracy Evaluation

The data of PV output for the given power plant, which constitutes of samples at 1-min interval for a 90-day period, is used in this paper for forecasting the 1-day-ahead PV output. The RMSE and RME(Relative Mean Error) values have been used for evaluation of the forecasting accuracy.

$$RMSE(y, \hat{y}) = \left(\frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2\right)^{1/2} \quad ...(5)$$

where, $\hat{y}_i$ is the predicted value of the ith sample, and y_i is its corresponding true value.

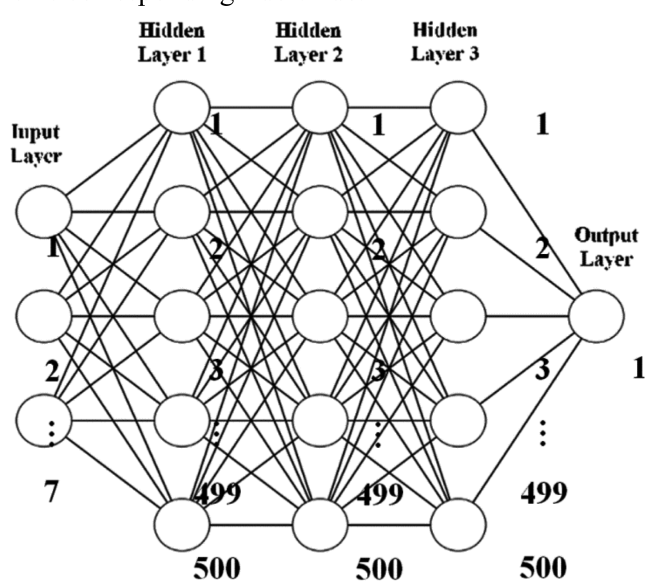

Fig. 6 — Computation Graph of the BPNN.

Table 1 — Accuracy obtained using SVM Model

Month	Linear SVM (Accuracy)	RBF SVM (Accuracy)
May	88	97
June	93	95
July	40	88
Average Accuracy	73.66%	93.33%

$$RME = \frac{1}{N}\sum_{i=1}^{n}(\mathcal{E}_i) \quad ...(6)$$

$$\mathcal{E}_i = y_{i,forecast} - y_{i,observed} \quad ...(7)$$

where, $y_{i,forecast}$ and $y_{i,observed}$ are the i^{th} forecasted and observed values, respectively, and $\mathcal{E}_i$ is the i^{th} error, with $i = 1, \ldots, N$ running through all forecast – observation pairs in the test dataset.

3 Results and Discussion

The forecasting models explained and trained in order to achieve the best possible forecasting results. In order to obtain 24-hours ahead forecast, the input parameters for the current moment fed to classifier, which were classified it into one of the three models. The corresponding neural network model applied for forecasting the PV output after 24 hours. The forecasting results of the three models can be seen in Fig. 7. It can be observed that the forecasted value of the PV output best followed the true value for the Sunny model, out of the three models. This can be further verified from Fig.8. in which it can be seen that the range of relative error is minimum for the Sunny model, i.e. between -3.5% to 2.5%, followed by that of the Cloudy model, which has the relative error range between -5.5% to 2.5%, finally followed by the maximum error range for the Rainy model, i.e. between -6.5% to 2.5%.

From Table II, three models can be observed, especially the Sunny Model, accomplish decently in forecasting. The mean values of RME and RMSE of the 3 models for 1 day ahead forecasting are 1.53% and 2.37% respectively.

In general terms, the Sunny Model gave the performance with a RME and RMSE of 0.85% and 1.26% respectively following which were the Cloudy Model with the values of 1.66% and 2.59%, and the Rainy Model with the values of 2.09% and 3.28%.

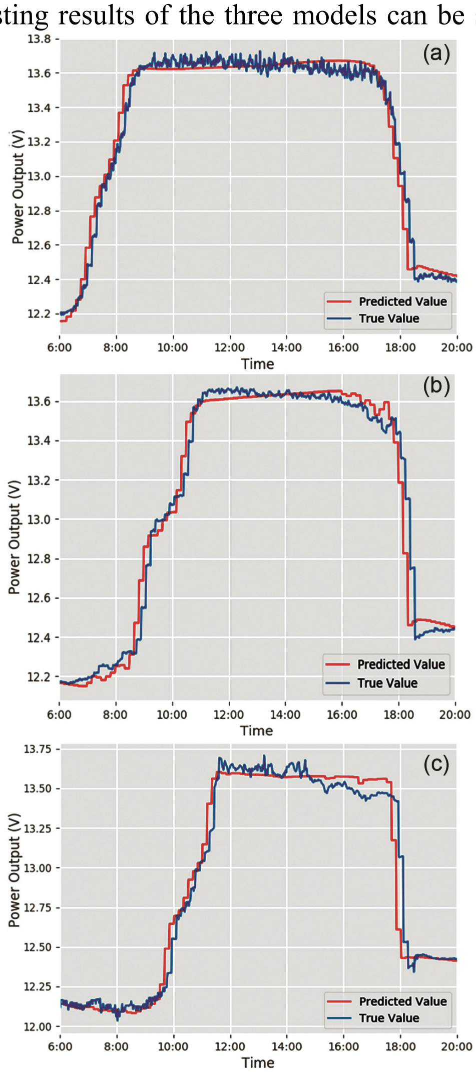

Fig. 7 — Predicted Value vs True Value Curve for (a) Sunny model (b) Cloudy model, and (c) Rainy model

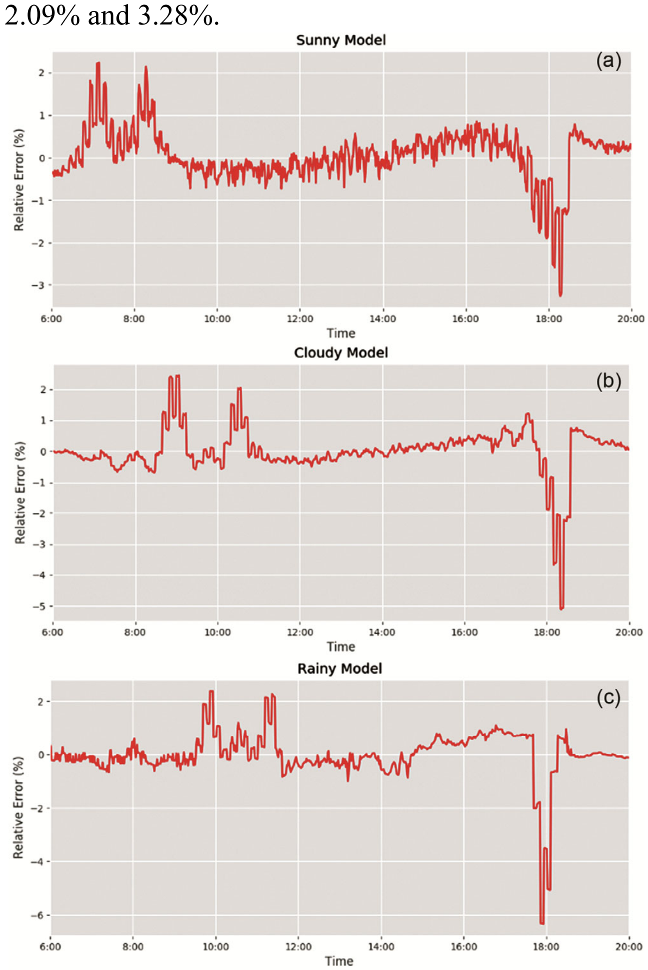

Fig. 8 — Relative error for (a) Sunny model (b) Cloudy model, and (c) Rainy model

Table 2 — RME's and RMSE's for the 3 models

Model Classification	RME (%)	RMSE(%)
Sunny Model (Model A)	0.85	1.26
Rainy Model (Model B)	2.09	3.28
Cloudy Model (Model C)	1.66	2.59
Average	1.53	2.37

4 Conclusion

The model has presented for 1-day-ahead PV system output forecasting using SVM classification and Neural Networks.

- The input data corresponding to the current moment has been first fed to the SVM classifier, which classifies it into one of the 3 weather types, namely, sunny, rainy and cloudy.
- Three Neural Network models corresponding to each weather type are set up to forecast the output.

The RME obtained for the overall model is 1.53%, which is better as compared to 8.64% obtained in previous research. Thus, the method has shown favourable results for the application of SVM-NN combination for forecasting solar PV output.

References

1 Panigrahi R, Mishra K S, Srivastava C S, Srivastava K A, & Schulz N N, *IEEE Trans Ind Application*, 56 (2020) 3178

2 International Energy Agency, "Renewables 2017 – Analysis and Forecasts 2022"

3 Naqvi Q B S, & Singh B, *Grid Connected Two Stage PV-Battery System with PV Intermittency Mitigation and Improved Power Quality, Internati Conf on Electric and Electronic Engineer (ICE3),*Gorakhpur, India, 2020

4 Jianhua N S, Dehua Zhang, & Ma Zheng, Hui, *Input parameters selection and accuracy enhancement techniques in PV forecasting using Artificial Neural Network*, IEEE International Conference on Power and Renewable Energy (ICPRE), Shangai, 2016.

5 Oozeki T, Ohtake H, Shimose K, Takashima T, & Ogimoto K, *Energy Procedia*, 57 (2014) 1337

6 Bacher P, Madsen H, &Nielsen HA, J *Solar Energy,*83(2009) 1772

7 Shi, Lee J, Liu J W,& Yang Y, *IEEE Industrial Applicat Society Ann Meeting*, (2011), 1.

8 Sun Y, Wang F, Mi Z, Sun H, Liu C, Wang B, Lu J, Zhen Z, LiK, *Short-term prediction model of module temperature for photovoltaic power forcasting based on support vector machine*, *International Conference on Renewable Power Generation (RPG),*Beijing, 2015.

9 Joseph A, *J Internati J Scientific Engineering and Applied Science,* 2 (2016) 139

10 Kumar A., Seema, Singh B. , &Jain R., *Double Stage Grid-Tied Solar PV System Using HC-LMS Control*, *IEEE 9th Power India International Conference* (PIICON), Sonepat, Haryana, 2020

11 Central Electricity Authority Report, 'Growth of electricity sector in India from 1947-2020'.

12 Sheng H, Xiao J, Cheng Y, Q. Ni, & Wang S, *IEEE Transactions on Industrial Electronics*, 65 (2018) 300

13 Arjyadhara Pradhan, Ali S.M, & Chitralekha Jena, *International J Engineering And Computer Science*, 2 (2013) 214

14 Rouholamini A, Hamed P, Roohollah F, Morteza A, *Int J Ambient Energy*, 1 (2014) 86

15 Wu T, Chang C, & Chang Y Lee K, *Power loss analysis of grid connection photovoltaic systems*,*2009 Internat. Conf. Power Electronics and Drive Systems (PEDS)*, Taipei, Taiwan, 2009.

16 Kazem H, Chaichan T M, Al-Shezawi M I, Hanan S A S, Hanan S A R, Khamis Al-Sinani J, & Al-Waeli H A A, *Asian Transactions on Engineering*, 2 (2012) 29

17 Arora Ankita, & Gaur Prerna, *IEEE INDICON*, 1 (2015) 157