SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Fluid Mechanics and Thermal Sciences

Journal homepage:
https://semarakilmu.com.my/journals/index.php/fluid_mechanics_thermal_sciences/index
ISSN: 2289-7879

Application of Machine Learning to Forecast Solar Photovoltaic Output Power

Muhammad Paend Bakht[1,2], Mohd Norzali Haji Mohd[1,*], Usman Ullah Sheikh[3], Nuzhat Khan[3], Mehr Gul[2]

[1] Faculty of Electrical and Electronics Engineering (FKEE), Universiti Tun Hussain Onn Malaysia, 86400 Parit Raja, Malaysia
[2] Department of Electrical Engineering, Balochistan University of IT, Engineering and Management Sciences, 87300 Quetta, Pakistan
[3] School of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia

ARTICLE INFO

Article history:
Received 3 April 2024
Received in revised form 5 July 2024
Accepted 16 July 2024
Available online 15 August 2024

Keywords:
Machine learning; random forest; prediction; support vector machine; solar photovoltaic; renewable energy; artificial intelligence

ABSTRACT

Due to the intermittent behaviour of the sun, accurate prediction of solar photovoltaic (PV) power is crucial for efficient and reliable operation of solar power plants. This paper presents state of the art approach for PV panels power prediction using machine learning (ML) method. Two ML models, namely Random Forest (RF) and Support Vector Machine (SVM) are trained and tested using input data of solar irradiance, ambient temperature, wind speed, humidity, precipitation and PV output power. The case study is presented for the grid tied PV system installed at University Tun Hussein Onn campus Batu Pahat Malaysia. The results indicated regression predictions reasonably fit the actual data, proving good potential of ML for PV power prediction. Besides, the predictive performance of RF and SVM was compared based on three evaluation metrics: coefficient of determination (R^2), root mean squared error (RMSE) and mean absolute error (MAE). Both ML models showed comparable predictive power with RF performing slightly better than SVM. The R^2 value for RF was 0.850 compared to 0.832 for SVM, indicating that RF was able to explain more of the variability in the data. Additionally, RF had lower values for both RMSE and MAE, indicating that it was better able to predict values of the solar PV power output. The conclusion from this study imparts the importance of ML methods to predict PV power which could be useful for optimizing the efficiency and reliability of solar energy systems.

1. Introduction

Energy is of paramount importance to any country's progress. It is a critical resource required for economic growth, social development, and improvement in the quality of life of citizens [1]. With advancements in technology and industrial growth, this energy dependence is swelling [2]. Today, most electricity is generated from burning fossil fuels, which is not only harmful to the environment, it is also not sustainable due to limited availability of coal, oil, and natural gas [3,4]. Considering these factors, it is necessary to take precautionary measures to address the potential energy deficit in the

* *Corresponding author.*
E-mail address: norzali@uthm.edu.my

https://doi.org/10.37934/arfmts.120.1.112

upcoming years and mitigate the problems that contribute to the global warming. Public awareness programs have expedited the search for alternative energy solutions, and the adoption of renewable energy sources (RES) has increased in the last two decades.

Among RES, solar photovoltaic (PV) systems are the most commonly used because of the ample availability of solar irradiance [5]. It is believed that the solar energy that reaches the Earth's surface in a single hour is greater than the total amount of global energy consumption over the entire year. Moreover, the amount of solar energy available on the planet is significantly greater than the current reserves of oil and coal. As mentioned by Zazoum [6], the estimated solar energy potential is approximately 157 times greater than the current coal reserves, and 516 times greater than the current oil reserves. These factors highlight the immense potential of solar energy as a sustainable and abundant source of renewable energy for future use. Furthermore, PV systems are easy to install, have low maintenance requirements, and do not require fuel for operation. Despite these benefits, PV systems are not considered dependable generators, owing to the sporadic behaviour of the sun [7]. The intermittency and variability of PV power output may affect the real-time control performance and economic viability of the system. In this situation, predicting the power output of a PV system (using an appropriate prediction model) is crucial for optimizing the operation and maximizing the economic benefits of the system. It is important to note that accuracy of a predictive model relies on the relevance of the selected features (input variables) used for making predictions. For PV power prediction, solar irradiance is considered a key factor, while other climatological parameters, such as temperature, wind speed, humidity, dust accumulation, and precipitation, can also influence PV efficiency. Their association may increase further in humid tropical environments such as Malaysia.

Recent literature on solar energy prediction shows a high research trend based on the Machine Learning (ML) paradigm. The use of ML models has resulted in improved reliability and safety in the decision-making processes. One practical application of predictive models is the identification of small power oscillations caused by external factors in photovoltaic (PV) systems. This contributes to improving the predictability of electricity generation and transmission [8]. In the literature, various ML models are employed based on their appropriateness for the dataset. Commonly utilized models for PV power prediction include Artificial Neural Networks (ANN), Support Vector Machines (SVM), and the Random Forest (RF) [9-14]. Each of these models exhibits distinct characteristics for different datasets. ANN are favored for their ability to capture complex relationships within datasets, yet their complicated structure with multiple layers and nodes demands substantial training data. SVM in contrast, feature fewer parameters and are often preferred for short-term PV output power prediction due to their efficiency and requirement of less data. However, the performance of SVM can be sensitive to internal parameters, leading to potential issues of overfitting or underfitting as demonstrated by Pan *et al.,* [15]. The RF algorithm has gained attention as a promising alternative. In studies such as that conducted by Benali *et al.,* [16], a comparative analysis between ANN and RF revealed RF superiority to forecast solar radiation and PV power generation. Similarly, Fouilloy *et al.,* [14], compared multiple statistical learning and machine learning tools, finding that the RF method demonstrated higher forecasting accuracy, particularly when dealing with data exhibiting a higher degree of variation. While all models exhibit favorable performance on different datasets, the superiority of the RF model becomes apparent when comparing their performance on a specific dataset. Motivated by the superior performance of RF demonstrated in prior studies by Fouilloy *et al.,* [14] and Benali *et al.,* [16], we opted to incorporate the RF model in our research and and contrasted the outcomes with those of SVM, acknowledged as the best-performing model in previous studies [12,17,18]. This decision is rooted in the algorithms' ability to provide accurate and reliable predictions for solar energy output, making it a suitable choice for our specific forecasting needs.

While various studies have employed these models for solar forecasting, their application in the context of Malaysia remains unexplored.

The main contribution of this paper is to utilize machine leaning potential to predict solar photovoltaic power in tropical regions. The prediction of solar PV output power is useful for grid operators to enhance usage of solar energy and to make decisions of grid operations. This work considers actual data profile of university campus in Malaysia and compares the performance of applied models (RF and SVM) on the basis of R^2, RMSE and MAE values. The rest of the paper is organised as follows: Section 2 outlines the materials and methods used in this study. Section 3 provides the results and corresponding discussions. Finally, Section 4 concludes this work and highlights potential areas for future research.

2. Materials and Methods

2.1 Study Area

The simulation and the case study of this work are performed for Universiti Tun Hussein Onn (UTHM) campus situated in Batu Pahat district Malaysia. The UTHM benefits from 6.9MWp solar PV grid tied system. The huge research plant consists of 15000 TRINA TSM-DE17M PV panels located physically across 26 campus buildings. The studied location is at latitude 1.532° N and longitude 103.0821° E. This region experiences a tropical climate, with an average yearly rainfall of 2494 mm and a relative humidity level of 79%. The average annual temperature is 26.4°C with average solar radiation from 4.07 to 5.22 kWh/m²/ day.

2.2 Data Collection and Analysis

In this study, data measured by the “NASA Langley Research Center (LaRC), USA)” accessed through an online platform is utilized [19]. This platform provides free access to solar and meteorological datasets to support renewable energy projects. For ML algorithms to be trained effectively and validated reliably, long term data is preferred [12]. The climatological data applied in this research includes temperature, solar irradiance, wind speed, relative humidity and precipitation of the considered location as shown in Figure 1. The daily data are applied and logged from November 2021 to May 2023, containing a total dataset of 553 samples.

The observed PV output power, which is used as the predictand (*i.e.*, the output variable) is collected from the Pejabat Pembangunan dan Penyenggaraan department of UTHM. The daily PV data is collected from 01/11/2021 to 08/05/2023. A detailed description of obtained data containing input variables, their statistical characteristics, and the average rate of change is provided in Table 1.

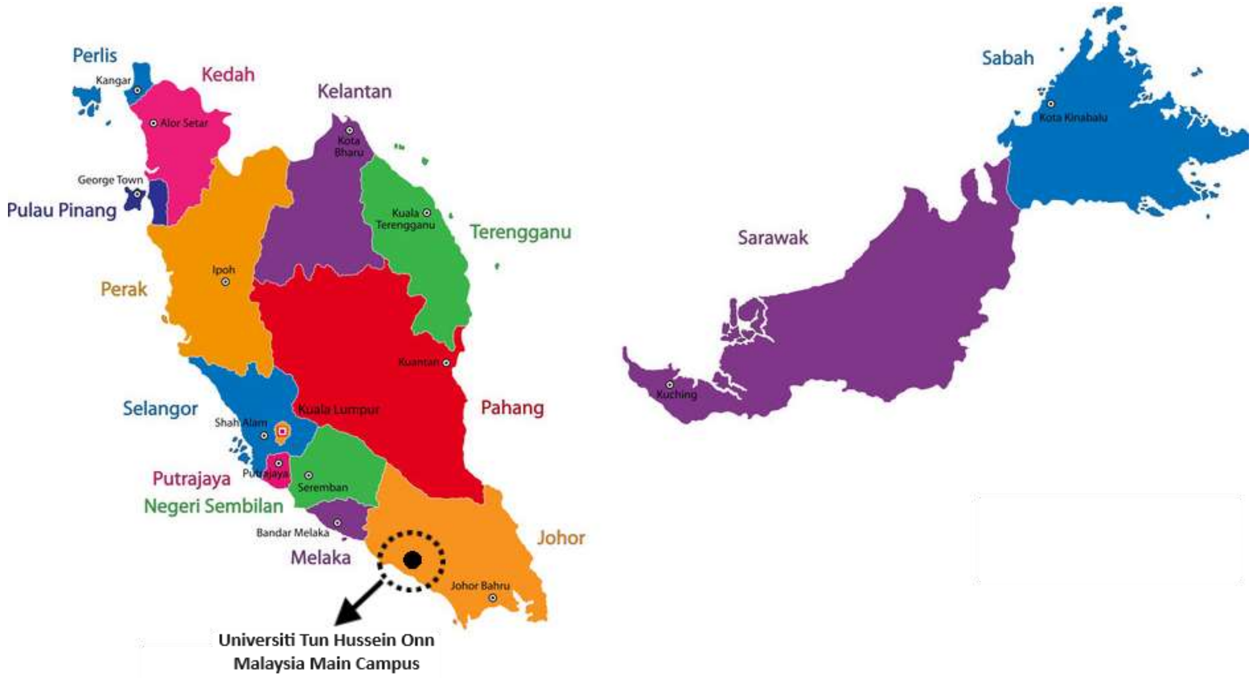


Fig. 1. The case studied location in Malaysia map

Table 1
Some important statistics of input data

Variable	Mean	Standard Deviation	Minimum	Maximum	Rate of change
Relative humidity	85.40	3.62	74.50	93.31	0.01
Precipitation	10.31	14.34	0.03	118.63	1.03
Temperature	27.05	0.88	24.90	29.18	0.01
Wind speed	1.90	0.66	0.53	3.94	0.05
Irradiance	6.56	0.252	5.47	7.16	0.71
Measured power	16.71	17.29	12.24	30.08	0.26

In machine learning, measuring the relationship between data variables is crucial for understanding how they interact and influence each other. The Pearson Correlation Coefficient (PCC) is commonly used to assess linear relationships between two continuous variables. It quantifies the strength and direction of relationships between variables, ranging its value from –1 to 1. The 1 indicates completely positive correlation between two random variables whereas -1 shows completely negative correlation between them. Likewise, 0 value of PCC indicates no obvious correlation between the considered variables. For this work, a basic correlation matrix using PCC is generated to evaluate the impact of different parameters on solar power generation. The correlation analysis between the considered input data and the observed power output can be visualized in Figure 2.

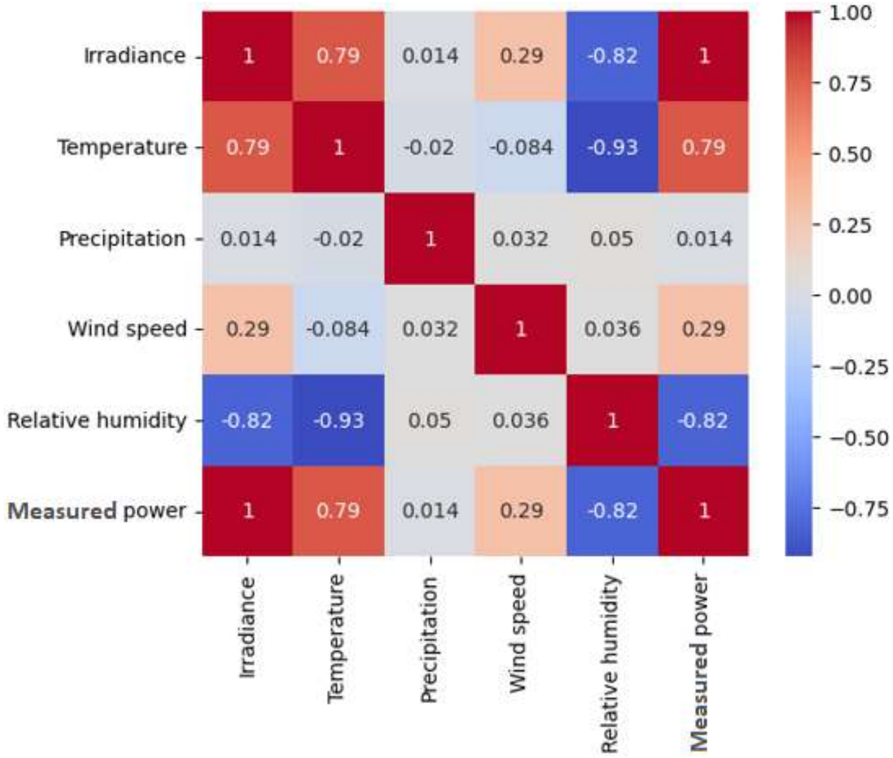


Fig. 2. Pearson correlation of the solar energy parameters

Among the variables, irradiance, temperature, precipitation and wind speed show positive correlation while humidity shows negative correlation with PV output power. The PCC values of irradiance (1) and temperature (0.79) indicate strong correlation while precipitation (0.014) does not contribute much to PV power. Besides, wind speed shows a moderate correlation (0.29) with PV power. Most of these findings are aligned with previous studies [20,21].

2.3 Machine Learning

ML is a highly effective tool for engineering applications, specifically in cases where physical modeling approaches fail due to the complex and unknown nature of the phenomena being described [22]. For PV plants, physical modeling necessitates knowledge of plant design data and multiple modeling steps, ML provides an attractive alternative for predicting PV power. Supervised learning is widely used branch of ML, where the training data includes both predictor and predictand values. It is preferred over conventional statistical methods due to its capability for long term predictions. The objective of supervised learning is to establish the best possible relationship between input and output variables. ML problems are typically classified into two main categories: classification, where outputs are categorical variables with discrete values, and regression, where outputs are continuous. In the case of PV power forecasting, typically the task is a regression problem as the objective is to predict a continuous variable.

Figure 3 shows proposed workflow for accomplishing PV prediction problem mainly comprising of three stages. In the first stage, the acquired data is pre-processed and transformed using minmax scaling normalization technique as proposed by Ahmed *et al.,* [23]. This technique involves scaling the values of a feature to a range from 0 to 1, with 0 and 1 being the minimum and the maximum values of the features, respectively. The minmax normalization technique is broadly used in literature to enhance ML model accuracy. Afterward, the prepared data is divided into training (80%) and testing (20%) set. In the next stage, the training data containing normalized input features is used for training of the selected models (RF, SVM). The training process led to the creation of a prediction model that can achieve long term PV power forecast. In the final stage, predictions based on testing data are compared using various evaluation metrics. Different parameters settings to perform this experiment are provided in Table 2.

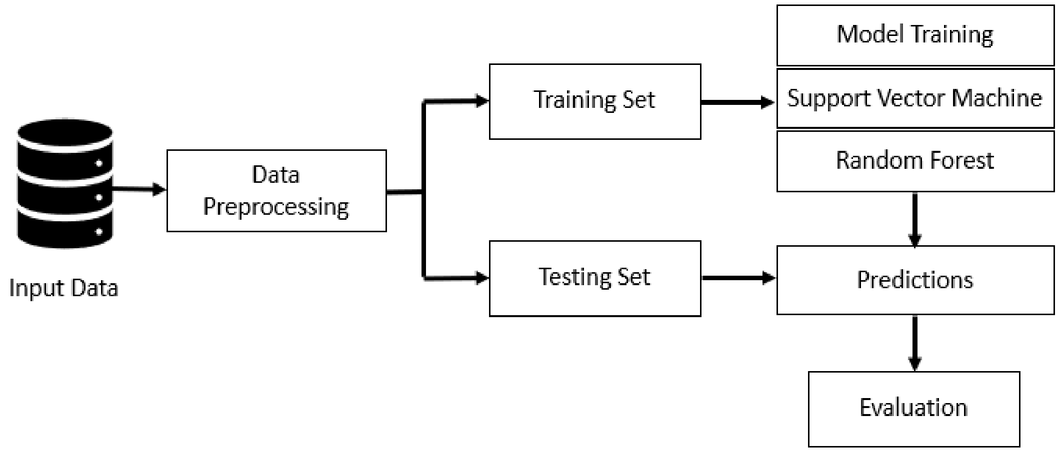


Fig. 3. Proposed workflow for PV panel power prediction

Table 2
Parameters setting

Parameter	Value
Training data	80 %
Testing data	20%
Number of Folds	10
Number of data features	6

2.3.1 Random forest

RF is popular algorithm for performing classification, regression, and feature selection tasks. It is based on ensemble learning technique that uses several decision trees instead of depending on single decision tree [24]. The trees are constructed using a random subset of input features and training data, which reduces overfitting problem and improves generalization performance. RF provides the output by averaging the prediction of the individual trees. Regression function for RF is stated in Eq. (1).

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} T_i(\mathrm{x}) \quad (1)$$

where x represents the vectored input variable, n represents the total number of trees, and $T_i(x)$ refers to a single regression tree built on a subset of input variables and the bootstrapped samples. The prediction process in RF is enhanced using voting method. Illustration of decision trees and the RF final output is shown in Figure 4. Major characteristics of RF are number of trees and their depth which has proven to be highly effective in real-world applications. Based on this, RF can be effectively applied for PV power prediction.

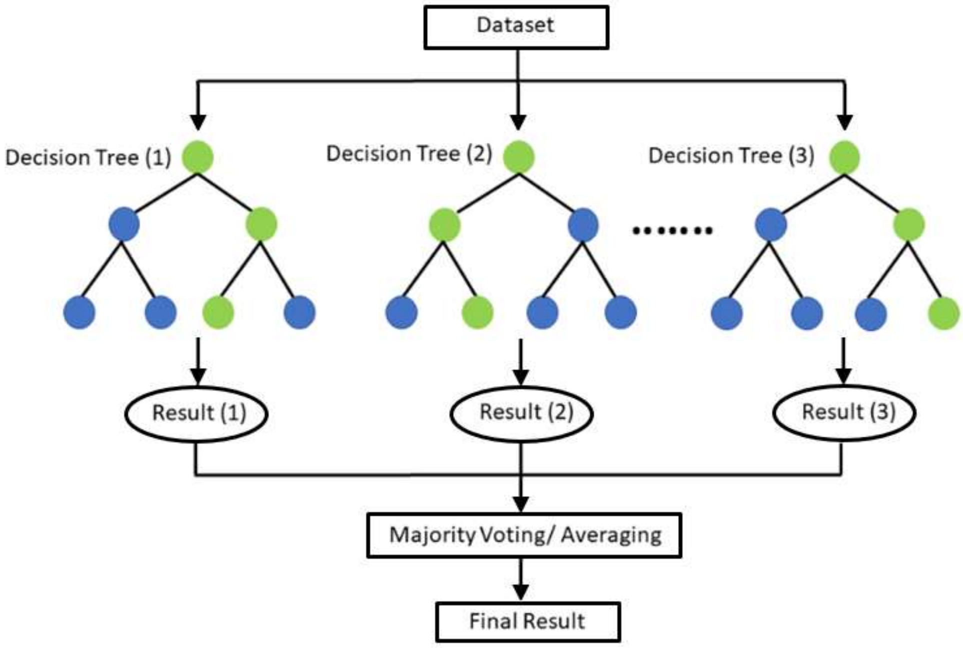


Fig. 4. Random forest prediction technique

2.3.2 Support vector machine

SVM is a popular ML algorithm used for classification and regression tasks as shown in Figure 5. It is well suited for complex and imbalanced dataset of small or medium-sized. The SVM creates a hyperplane within an *n*-dimensional vector space that helps to predict effectively. It utilizes different kernel functions depending on the type of dataset [25]. The function involves mapping the original data points into a higher-dimensional space where optimal hyperplane predicts continuous output values.

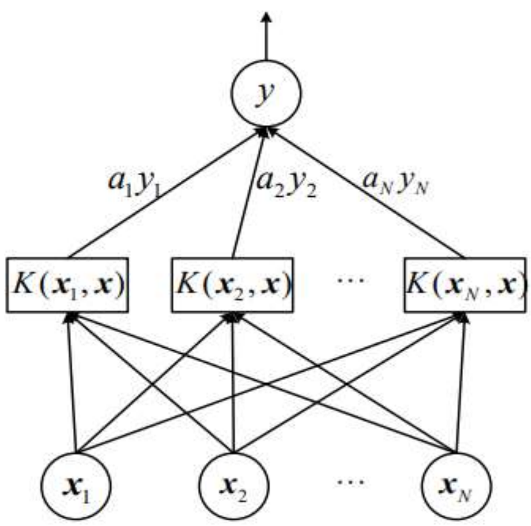


Fig. 5. Structure of Support Vector Machine Algorithm [26]

Considering the training dataset as D= $\{(x_1,y_1), (x_2,y_2),\ldots,(x_n,y_n)\}$, where $x_i \in R^d$ is the input vector and $y_i \in$R is the target value. The regression function is stated as mentioned in Eq. (2).

$$f(x) = w \,.\, \varphi\,(\mathrm{x}) + b \tag{2}$$

Here, $\varphi\,(\mathrm{x})$ represents the feature vector of the inputs *x*, the weight vector is represented by *w* and *b* shows the bias term. In this work, default Radial Base Function and default parameters (*C* =1 and Gamma = 'scale') are used according to Jebli *et al.,* [27].

2.4 Performance Evaluation

It is important to measure the performance of the forecasting model by comparing its predicted values to the actual observations. In literature several metrics are used depending on the type of problem and the nature of the data. Commonly utilized techniques are calculating the coefficient of determination (R^2), mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), root mean square error (RMSE) and max error (ME) etc. In this work, R^2 is used as the main performance evaluation metric, whereas RMSE and MAE are also included to improve the quality of forecasts. The formulas of these statistical functions are given as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(P_t - \hat{P}_t)^2}{\sum_{i=1}^{n}(P_t - P_{mt})^2} \tag{3}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |\, P_t - \hat{P}_t \,| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_t - \hat{P}_t)^2} \tag{5}$$

Here, P_t, $\hat{P}_t$ and $\hat{P}_{mt}$ represent the observed power, predicted power and the mean of the observed PV power and *n* shows the number of observations. Eq. (3) represents R^2 that estimates the proportion of variance in the in the target variable *i.e*., predicted power in this case. The value of R^2 ranges from 0 to 1, with higher values indicating a better fit of the model to the data. This is against MAE and RMSE where lower values are preferred as the small errors indicate better performance of the model in predicting the target variable. Eq. (4) calculates MAE as the average magnitude of the errors between the predicted and actual values, without considering the direction of the error. In other words, MAE is not sensitive to the direction of the errors (whether they are positive or negative), and provides a simple, interpretable measure of the model's performance. Accordingly, Eq. (5) calculates the RMSE that is the square root of the average of the squared differences between the predicted and actual values of the target variable. The RMSE is preferred over MAE in some cases because it gives more weight to larger errors, which can be useful in scenarios where large errors are particularly problematic or costly.

3. Results and Discussions

The evaluation metrics for performance of trained ML models on test data is listed in Table 3. The supervised ML models RF and SVM are used to predict solar PV power based on the input variables mentioned before. Out of 553 datapoints, 80% (443) are dedicated to train the model while 20% (111) are held out for validating models' performance. Using this method, long-term predictions of 111 days is achieved, which represents a substantial span of more than three months. The results underscore the exceptional performance of the ML models in learning linear and nonlinear relationships between dependent and independent variables. In terms of R^2, models accuracy

Journal of Advanced Research in Fluid Mechanics and Thermal Sciences
Volume 120, Issue 1 (2024) 1-12

consistently falls within the range of 0.832 to 0.850 for SVM and RF, respectively. This implies that the models can explain variability with an accuracy ranging from 83.2% to 85% based on the provided input features. Besides, the lower values of RMSE and MAE indicate that both models' predictions are close to the actual values. These statistical findings are aligned with other relevant studies on PV power prediction employing ML techniques [13,15]. Notably, in this study, RF outperforms SVM in all three-evaluation metrics. Thus, it can be concluded that RF model yields superior results for PV power prediction considering the input variables of temperature, irradiance, precipitation, humidity, wind speed and PV power. This research is based on the realistic performance of supervised ML models on real-world data. However, it is important to note that further accuracy enhancements can be explored by implementing and testing various data processing methods. Figure 6 shows the predicted and actual values of PV output power for RF and SVM model on the basis of R2 against the considered location.

Table 3
Performance comparison between the employed regression models (testing phase)

Model	R^2 (%)	RMSE (kW)	MAE (kW)
Random Forest	85.0	2.1637	1.3695
Support Vector Machine	83.2	2.3431	1.5074

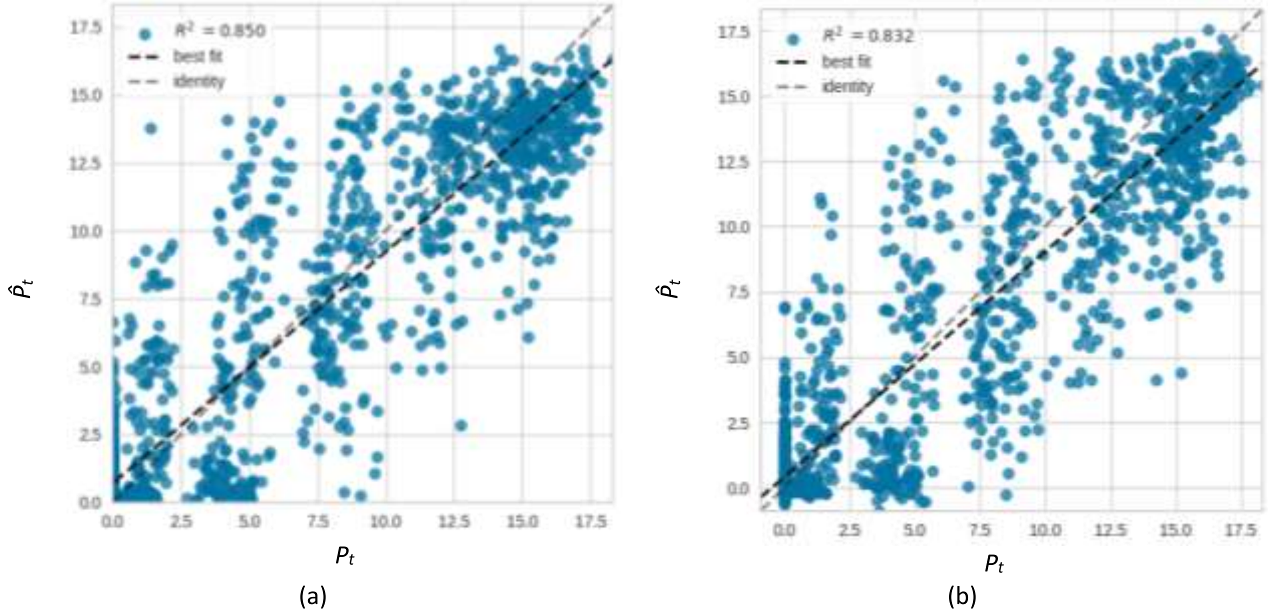
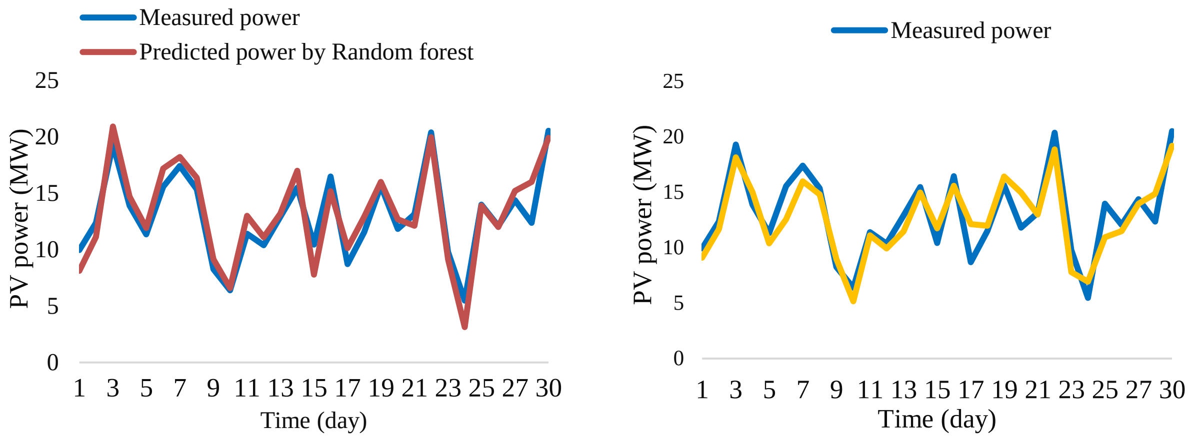


Fig. 6. Scatterplots of the predicted and actual PV panel output power values against UTHM campus, Malaysia (a) Random forest model, (b) Support vector machine model (Note: the gray identity lines represent 1:1 lines and black lines are the best fitted lines of models)

The time series variations of predicted and observed PV output power by the employed models are depicted in Figure 7. The plots are shown for 15000 PV panels of 450W each over one month from 01/11/2021 to 30/11/2021. It should be noted that the size of PV panels is considered to maintain consistency with the actual PV data. As can be observed, the predicted power values exhibit a pattern of fluctuation that follows the actual data. The results prove both ML models' capability to successfully capture most of the underlying patterns and relationships in the data. A few points where predicted patterns do not match the actual data are the result of models' impotence to learn the complexities of the data. Potential improvements such as larger datasets with better quality and close relevance of the additional data can play a vital role in improving the accuracy of the ML models.



(a) (b)

Fig. 7. Forecasting PV output power using (a) Random Forest (b) Support Vector Machine

4. Conclusions

This study demonstrated the potential of machine learning (ML) models for predicting solar photovoltaic (PV) power output. Two ML models, random forest (RF) and support vector machine (SVM) are applied considering the input data of solar irradiance, ambient temperature, humidity, wind speed, precipitation and PV output power. The models' performance was assessed using three metrics: coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE). The results showed strong potential of both ML models for predicting solar PV power output, indicating that ML methods are a promising approach to improve the efficiency and reliability of solar energy systems. The results of this study provide useful insights for researchers working in the field of renewable energy and machine learning. Future work can incorporate more features related to weather patterns and other environmental factors to further improve the accuracy of the models.

Acknowledgement

This research was supported in part by the Ministry of Higher Education, Malaysia, under Fundamental Research Grant K187/FRGS/1/2019/ICT04/UTHM/02/2; and in part by Universiti Tun Hussein Onn Malaysia (UTHM).

References

[1] Bakht, Muhammad Paend, Zainal Salam, Mehr Gul, Waqas Anjum, Mohamad Anuar Kamaruddin, Nuzhat Khan, and Abba Lawan Bukar. "The potential role of hybrid renewable energy system for grid intermittency problem: a techno-economic optimisation and comparative analysis." *Sustainability* 14, no. 21 (2022): 14045. https://doi.org/10.3390/su142114045

[2] Albarhami, Bahaa Abdulhur Hatem, and Mohammed Najeh Nemah. "Modelling and Simulation of Grid Connected Photovoltaic System using Matlab\Simulink Program." *CFD Letters* 15, no. 8 (2023): 19-30. https://doi.org/10.37934/cfdl.15.8.1930

[3] Zakaria, Siti Umaira, Sahriah Basri, Siti Kartom Kamarudin, and N. A. A. Majid. "Public awareness analysis on renewable energy in Malaysia." In *IOP Conference Series: Earth and Environmental Science*, vol. 268, no. 1, p. 012105. IOP Publishing, 2019. https://doi.org/10.1088/1755-1315/268/1/012105

[4] Lubis, Hamzah. "Renewable energy of rice husk for reducing fossil energy in Indonesia." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 11, no. 1 (2018): 17-22.

[5] Mary, S. Merlin Joys, S. Rajesh Babu, and D. Prince Winston. "Fuzzy logic based control of a grid connected hybrid renewable energy sources." *International Journal of Scientific & Engineering Research* 5, no. 4 (2014): 1043-1048.

[6] Zazoum, Bouchaib. "Solar photovoltaic power prediction using different machine learning methods." *Energy Reports* 8 (2022): 19-25. https://doi.org/10.1016/j.egyr.2021.11.183

[7] Bakht, Muhammad Paend, Zainal Salam, Abdul Rauf Bhatti, Waqas Anjum, Saifulnizam A. Khalid, and Nuzhat Khan. "Stateflow-based energy management strategy for hybrid energy system to mitigate load shedding." *Applied Sciences* 11, no. 10 (2021): 4601. https://doi.org/10.3390/app11104601

[8] David, Tudur Wyn, Helder Anizelli, T. Jesper Jacobsson, Cameron Gray, William Teahan, and Jeff Kettle. "Enhancing the stability of organic photovoltaics through machine learning." *Nano Energy* 78 (2020): 105342. https://doi.org/10.1016/j.nanoen.2020.105342

[9] Voyant, Cyril, Gilles Notton, Soteris Kalogirou, Marie-Laure Nivet, Christophe Paoli, Fabrice Motte, and Alexis Fouilloy. "Machine learning methods for solar radiation forecasting: A review." *Renewable Energy* 105 (2017): 569-582. https://doi.org/10.1016/j.renene.2016.12.095

[10] Khandakar, Amith, Muhammad E. H. Chowdhury, Monzure Khoda Kazi, Kamel Benhmed, Farid Touati, Mohammed Al-Hitmi, and Antonio Jr S. P. Gonzales. "Machine learning based photovoltaics (PV) power prediction using different environmental parameters of Qatar." *Energies* 12, no. 14 (2019): 2782. https://doi.org/10.3390/en12142782

[11] Almonacid, Florencia, P. J. Pérez-Higueras, Eduardo F. Fernández, and L. Hontoria. "A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator." *Energy Conversion and Management* 85 (2014): 389-398. https://doi.org/10.1016/j.enconman.2014.05.090

[12] Eseye, Abinet Tesfaye, Jianhua Zhang, and Dehua Zheng. "Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information." *Renewable Energy* 118 (2018): 357-367. https://doi.org/10.1016/j.renene.2017.11.011

[13] Fan, Junliang, Xiukang Wang, Lifeng Wu, Hanmi Zhou, Fucang Zhang, Xiang Yu, Xianghui Lu, and Youzhen Xiang. "Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China." *Energy Conversion and Management* 164 (2018): 102-111. https://doi.org/10.1016/j.enconman.2018.02.087

[14] Fouilloy, Alexis, Cyril Voyant, Gilles Notton, Fabrice Motte, Christophe Paoli, Marie-Laure Nivet, Emmanuel Guillot, and Jean-Laurent Duchaud. "Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability." *Energy* 165 (2018): 620-629. https://doi.org/10.1016/j.energy.2018.09.116

[15] Pan, Mingzhang, Chao Li, Ran Gao, Yuting Huang, Hui You, Tangsheng Gu, and Fengren Qin. "Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization." *Journal of Cleaner Production* 277 (2020): 123948. https://doi.org/10.1016/j.jclepro.2020.123948

[16] Benali, Lamara, Gilles Notton, A. Fouilloy, Cyril Voyant, and Rabah Dizene. "Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components." *Renewable Energy* 132 (2019): 871-884. https://doi.org/10.1016/j.renene.2018.08.044

[17] Khosravi, A., R. N. N. Koury, L. Machado, and J. J. G. Pabon. "Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms." *Journal of Cleaner Production* 176 (2018): 63-75. https://doi.org/10.1016/j.jclepro.2017.12.065

[18] Winston, David Prince, Madhu Shobini Murugan, Rajvikram Madurai Elavarasan, Rishi Pugazhendhi, O. Jeba Singh, Pravin Murugesan, M. Gurudhachanamoorthy, and Eklas Hossain. "Solar PV's micro crack and hotspots detection technique using NN and SVM." *IEEE Access* 9 (2021): 127259-127269. https://doi.org/10.1109/ACCESS.2021.3111904

[19] NASA. "Power - Data Access Viewer | Prediction Of Worldwide Energy Resources." *NASA*, 2022. https://power.larc.nasa.gov/data-access-viewer/.

[20] Markovics, Dávid, and Martin János Mayer. "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction." *Renewable and Sustainable Energy Reviews* 161 (2022): 112364. https://doi.org/10.1016/j.rser.2022.112364

[21] Mayer, Martin János, and Gyula Gróf. "Extensive comparison of physical models for photovoltaic power forecasting." *Applied Energy* 283 (2021): 116239. https://doi.org/10.1016/j.apenergy.2020.116239

[22] Khan, Nuzhat, Mohamad Anuar Kamaruddin, Usman Ullah Sheikh, Mohd Hafiz Zawawi, Yusri Yusup, Muhammed Paend Bakht, and Norazian Mohamed Noor. "Prediction of oil palm yield using machine learning in the perspective of fluctuating weather and soil moisture conditions: Evaluation of a generic workflow." *Plants* 11, no. 13 (2022): 1697. https://doi.org/10.3390/plants11131697

[23] Ahmed, Razin, Victor Sreeram, Yateendra Mishra, and M. D. Arif. "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization." *Renewable and Sustainable Energy Reviews* 124 (2020): 109792. https://doi.org/10.1016/j.rser.2020.109792

[24] Khan, Muhammad Yaseen, Abdul Qayoom, Muhammad Suffian Nizami, Muhammad Shoaib Siddiqui, Shaukat Wasi, and Syed Muhammad Khaliq-ur-Rahman Raazi. "Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep

Learning Techniques." *Complexity* 2021, no. 1 (2021): 2553199. https://doi.org/10.1155/2021/2553199
[25] Ayodele, T. R., A. S. O. Ogunjuyigbe, A. Amedu, and J. L. Munda. "Prediction of global solar irradiation using hybridized k-means and support vector regression algorithms." *Renewable Energy Focus* 29 (2019): 78-93. https://doi.org/10.1016/j.ref.2019.03.003
[26] Zhang, Zhengxie, Shuguo Pan, Chengfa Gao, Tao Zhao, and Wang Gao. "Support vector machine for regional ionospheric delay modeling." *Sensors* 19, no. 13 (2019): 2947. https://doi.org/10.3390/s19132947
[27] Jebli, Imane, Fatima-Zahra Belouadha, Mohammed Issam Kabbaj, and Amine Tilioua. "Prediction of solar energy guided by pearson correlation using machine learning." *Energy* 224 (2021): 120109. https://doi.org/10.1016/j.energy.2021.120109