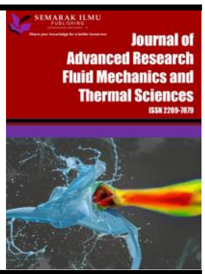
SEMARAK ILMU PUBLISHING
202103268166(003316878-P)

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



Long-Term Solar Power Generation Forecasting in the Eastern Coast Region of Malaysia using Artificial Neural Network (ANN) Method

Muhammad Aiqal Iskandar[1,2], Muhammad Azfar Shamil Abd Aziz[1], S. S. Sivaraju[3], Nurdiyana Borhan[4], Wan Abd Al-Qadr Imad Wan Mohtar[5], Nurfadzilah Ahmad[1,*]

1 Solar Research Institute (SRI), Universiti Teknologi MARA (UiTM) Shah Alam, Selangor, Malaysia
2 School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM) Shah Alam, Selangor, Malaysia
3 RVS College of Engineering and Technology, Coimbatore, Tamil Nadu, India
4 Davex (Malaysia) Sdn. Bhd., Subang Jaya, Selangor, Malaysia
5 Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 24 December 2023
Received in revised form 25 April 2024
Accepted 7 May 2024
Available online 30 May 2024

Keywords:
Correlation; power generation; forecasting; ANN

ABSTRACT

Accurate prediction of power demand and generation is crucial for modern energy systems to efficiently allocate resources and facilitate energy trading. The integration of artificial intelligence (AI) and machine learning techniques has significantly improved the precision of power forecasting. This study focuses on the application of Artificial Neural Networks (ANN) for forecasting power generation in the Eastern Coast region of Malaysia, with a specific emphasis on solar power. The research methodology involves collecting and analyzing historical power data, weather data, and relevant variables. ANN models are trained, validated, and tested on a selected power grid to assess their accuracy and predictive capabilities. The expected outcomes aim to include the development of a precise power generation forecasting model, providing valuable insights for decision-makers to optimize energy operations and seamlessly integrate renewable sources. Additionally, the study explores potential challenges, limitations, and best practices associated with ANN-based power forecasting. The dataset covers the period from 2020 to 2023, with variables such as average output power, ambient temperature, PV module temperature, global horizontal irradiance, and wind speed recorded at 30-minute intervals. The architecture of the ANN model, implemented using the Keras framework, is described as a Sequential model with layers utilizing the 'ReLU' activation function. Model evaluation employs metrics like root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE) on the test set, offering insights into the model's overall fit, average deviation, and sensitivity to outliers. Results reveal strong correlations between PV module temperature, irradiance, and AC power generated.

1. Introduction

The efficient and reliable forecasting of power demand and generation has become increasingly crucial in modern energy systems. Renewable energy has significant challenge is the limited

* *Corresponding author.*
E-mail address: nurfadzilah6344@uitm.edu.my

https://doi.org/10.37934/arfmts.117.2.6070

predictability of these factors in the near future, and the power output is contingent on atmospheric parameters [1]. These improvements in accurate power forecasting give advantages for electrical energy companies to have a better sight regarding resource allocation and energy trading. It assists in optimizing the functioning of alternative energy sources to supplement the sporadic characteristics of solar energy [2]. Over the past few years, the rapid development of artificial intelligence (AI) and machine learning techniques has revolutionized power forecasting by providing more accurate and adaptive prediction models. Solar power forecasting enables better management and utilization of renewable energy resources. It helps optimize the operation of other energy sources to complement the intermittent nature of solar energy [2]. While solar energy is environmentally friendly, its consistency over extended periods is limited [3].

This research project focuses on utilizing the use of Artificial Neural Networks (ANN) on long-term power forecasting at a large scale solar (LSS) farm in the Eastern Coast region of Malaysia. Artificial neural network (ANN) has been viewed as a convenient way to forecast solar radiation intensity and power output of PV systems, which can be trained to overcome the limitations of traditional methods to solve complex problems, and to solve difficult problems which are hard to model and analyze [4]. The “vanilla” feed-forward neural network is a common and simple type of Artificial Neural Network (ANN). It has a single hidden layer that connects the input and output layers, and it is widely used for tasks like pattern recognition and classification [5]. Application of such technology offers a powerful machine learning approach by replicating how the human brain operates, where it is known to capture complex patterns processing, integrating, and coordinating the data. By training an ANN model on historical power data and incorporating relevant meteorological, temporal, and contextual factors, it is possible to develop accurate and robust power forecasting models. The use of AI also enables efficient inverter control of photovoltaic (PV) systems [6].

The research methodology will involve the collection and analysis of historical data of AC power generation and other relevant variables. The models will be trained, validated, and tested on data from the large scale solar (LSS) farm to assess their accuracy and predictive capabilities. The anticipated outcomes of this research include the development of an accurate power generation forecasting model using ANN to assist decision-makers in optimizing energy operations, ensuring reliable energy supply, and facilitating the integration of renewable energy sources into the grid. Furthermore, this study aims to provide insights into the potential challenges, limitations, and best practices associated with ANN-based power forecasting [7].

2. Methodology

2.1 Dataset

The purpose of the study is to employ an artificial neural network (ANN) model to forecast the electricity generation from ac power generation data. Figure 1 shows the flowchart of the whole project, where the project starts with collecting the historical data, constructing the model architecture, training the ANN model, testing and deploying the model performance and evaluating the evaluation metrices. The data is collected from 50MW solar farm in Eastern Coast region of Malaysia. The data contains the 14,544 historical record of the power generation and the variables that monitored such as average output power, ambient temperature, PV module temperature, total global horizontal irradiance, and wind speed. The data is recorded every five minutes, but the dataset used 30 minutes times interval after the data cleaning. The data is recorded from 2020 up until 2023. The data will be split into two, train and testing [8]. The model use data from (2-8-2020 to 31-7-2021) to perform the training and testing of the performance of ANN. Then, the model will be deployed to the data from same dataset ranging from (1-9-2022 to 30-4-2023).

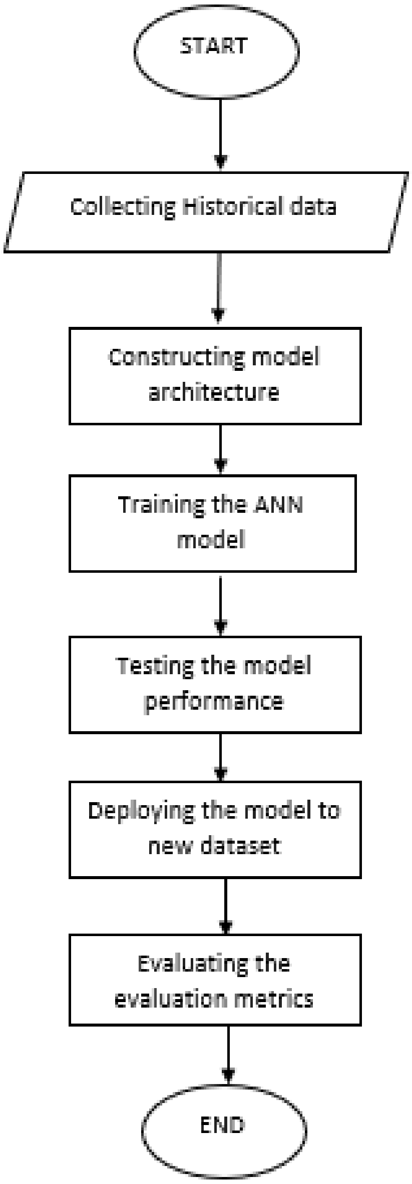


Fig. 1. Flowchart of the project

2.2 Model Architecture

The model architecture is constructed using Keras API. The Keras Model API retrieves the trained features and target of the selected layer of the model given by the user [9]. The design primarily encompasses an input layer, a hidden layer, an output layer, connection weights, biases, an activation function, and a summation node [10]. The model architecture is then defined using the Sequential model, stacking layers sequentially. Each layer, created with the Dense class, conducts a linear transformation on the input data and applies a specified activation function. Each layer consists of trainable parameters that are adjusted during the training process to optimize the model's performance on a given task. The code snippet exemplifies this by defining a model with input shape (2,) and successive layers containing 180, 65, 30, 30, and 1 node, all employing the 'ReLU' activation function.

2.3 Model Evaluation

The model evaluation is done using the test set. The model predicts the power generation for the test set and compares it to the actual power generation. The model evaluation is quantified using metrics such as root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE), which reflect how near the predictions are to the actual data. The model evaluation can be viewed by using matplotlib to plot the anticipated values against the actual values. It can provide how well the model fits the data overall, how much it deviates from the data on average, and how sensitive it is to outliers and variations. By assessing the overall model using the metrics of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) about the genuine values, the accuracy of the model can be determined [11]. The calculation is based on the formula given:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1} n \, (y_i - \hat{y}_i)^2} \quad (1)$$

$$MSE = \frac{1}{n}\sum_{i=1} n \, (y_i - \hat{y}_i)^2 \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1} n \, |y_i - \hat{y}_i| \quad (3)$$

N is the total number of observations in the test set, y_i is the actual ac power generation and $\hat{y}_i$ is the forecasted ac power generation.

$$RMSE\ \% = \frac{RMSE}{\max(y) - \min\,(y)} x\ 100 \quad (4)$$

$$MSE\ \% = \frac{MSE}{(\max(y) - \min(y))^2} x\ 100 \quad (5)$$

$$MAE\ \% = \frac{MAE}{\max(y) - \min\,(y)} x\ 100 \quad (6)$$

3. Results and Discussion

3.1 Features Selection

Figure 2 is a correlation heatmap used to display the strength and direction of the relationships between different variables. The variables are AC power, ambient temperature, PV module temperature, total global horizontal irradiance, and wind speed. These variables are related to the performance of a photovoltaic (PV) for AC power generation. The color indicators show the significance and direction of the correlation. The color indicators show the significance of the correlation. Blue indicates low correlation, light blue for medium correlation and blue indicates high correlation. Based on the findings from the heatmap, the correlation between AC power generation and both total global horizontal irradiance and PV module temperature have a strong positive relationship. PV module generates the maximum amount of electricity when the incident light is orthogonal to the plane of the module. Therefore, based on this analysis of PV modules Pmax has a stronger correlation GHI. Besides, the irradiation shining on the panel turns into heat, therefore, increasing the temperature of the solar PV panel [12]. This means that as these two variables increase, so does the AC power output.

From Figure 2, the heatmap shows that the correlation coefficients for PV module temperature and total global horizontal irradiance are 0.9 and 0.91 respectively, which are very close to 1. This indicates that these two features have a high correlation with the target variable, which is the average output power of the PV system. The temperature of the PV module is a crucial factor that strongly affects the behavior of a photovoltaic (PV) system. It has a significant impact on the system's efficiency and the amount of energy it produces [13]. The irradiance element which mainly contributes to solar concentration, which has a direct relationship with solar power generation, is of especially important in the context of power generation [14]. Figure 2 also shows the wind speed correlation with AC power generation, it produces value of 0.43 which indicates a positive relationship, but it's not extremely strong. The relationship of wind speed and AC power generation is low is when the wind speed increases, it causes the surface temperature of PV modules to decrease, ultimately resulting in an increase in power generation [15]. This finding also shows a positive correlation of ambient temperature with AC power generation with value of 0.68. as the most suitable temperature for effective electricity generation in solar PV panels is 25 °C [16]. Therefore, the selection of optimal correlation features of total global horizontal irradiance and PV module temperature are used as features to be use for the model forecasting.

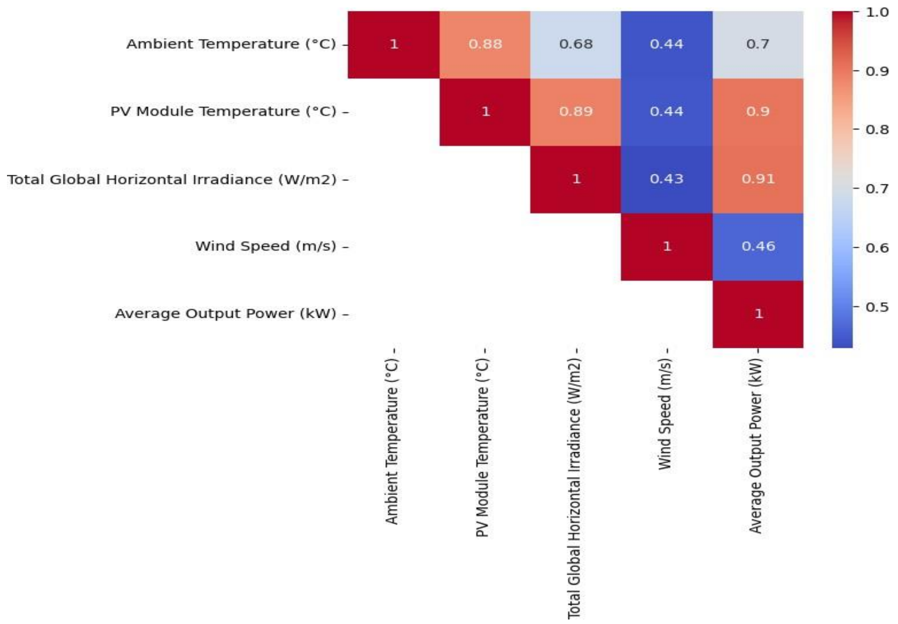


Fig. 2. Heatmap correlation for AC power generated

3.2 Model Performance

Figure 3 shows the graphs forecasted and actual values of AC power generated for a week from Monday to Sunday (Day 45- 51). The x-axis represents the time in hours, and the y-axis represents the power in watts (w). The AC power generation gradually increased at 9.00am. The blue line indicates the actual data, and the red line indicates the forecasted data. For weekday, except Wednesday and Thursday, there are small difference between the actual and forecasted data while Wednesday and Thursday have bigger difference when the data is change drastically. This can be seen on Wednesday at 16:00pm indicates flat forecasted data instead of following the increase in actual the increase and decrease of the power generation are mainly influenced by the solar irradiance total global horizontal irradiance.

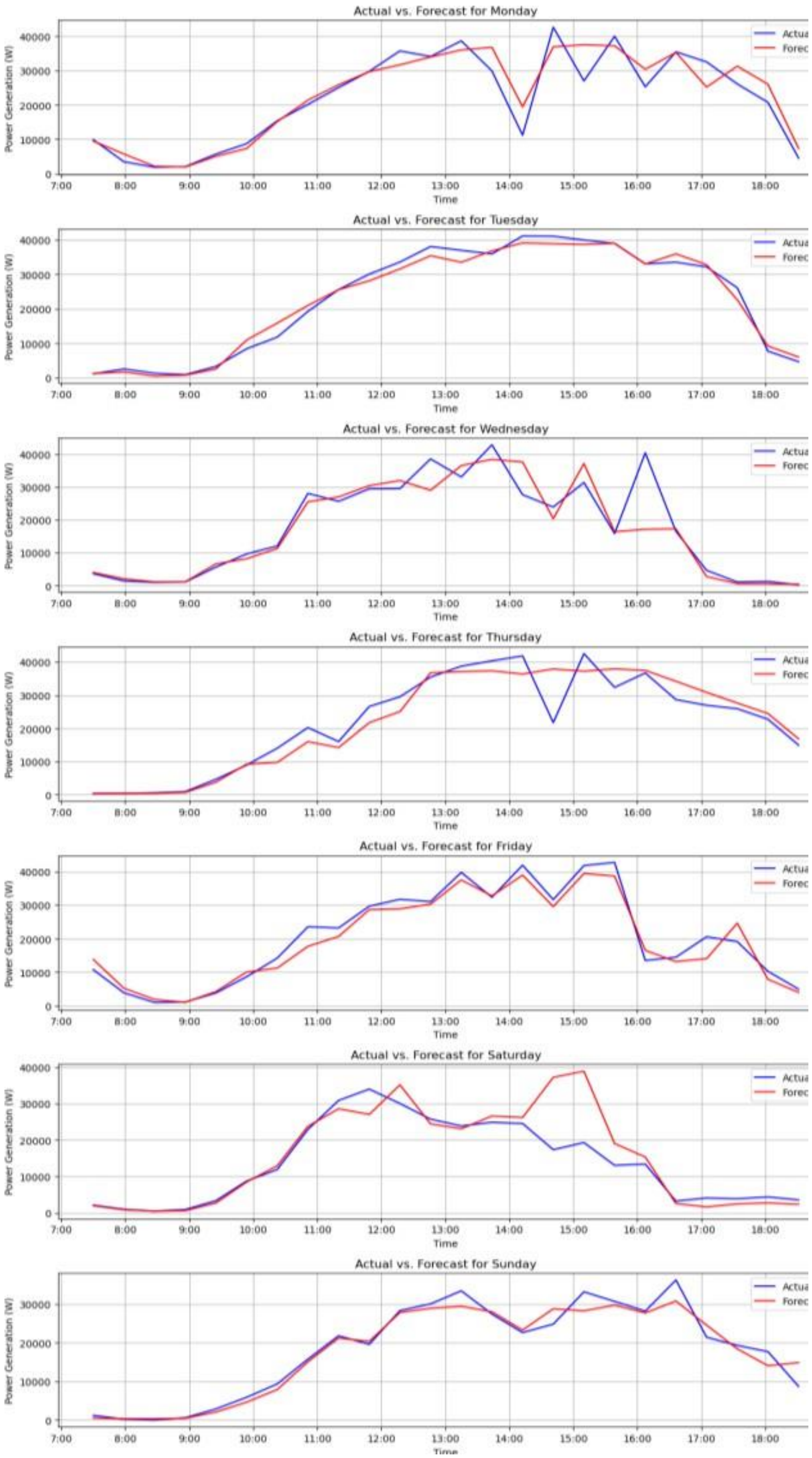


Fig. 3. Plot of forecast and actual data of the model

Table 1 shows the average values of RMSE, MSE, MAE, and R^2 for the forecasting model. These metrics measure the accuracy and error of the model by comparing the predicted power generation values with the actual power generation values. The RMSE (Root Mean Square Error) calculates the difference between each predicted data point and its corresponding actual data point, squares these errors, computes the average of the squared errors, and ultimately takes the square root of the result [17]. The lower the values, the better the model performance. The average RMSE for the model is 8.513% indicates the average percentage difference between your model's predictions and the actual values, the average MSE for the model is 0.829%, and the average MAE represents the average absolute difference between predictions and actual values for the model is 5.631% as shown in Table 1.

Table 1
Evaluation Metrics Table for Model

Evaluation metrics	Percentages
RMSE	8.513%
MAE	0.829%
MSE	5.631%
R^2	0.891

3.3 Model Deployment

To deploy the artificial neural network (ANN) model, we need two features: the temperature of the photovoltaic (PV) module and the total global horizontal irradiance (GHI) on the surface of the module. These features are used to predict the alternating current (AC) power generation of the PV system. Figure 4 shows the graphs of forecasted and actual lines when we deploy the model to a new dataset from Monday to Sunday. Based on the shapes of the graphs, they show that the forecasted data have significant differences from the actual data. The differences between the forecasted and actual data vary across different days and times. For instance, on Monday and Tuesday the differences are relatively small and consistent throughout the day. On Thursday, Friday Saturday and Sunday, the differences are larger difference, especially around midday. On Saturday, the differences are smaller than on weekdays, but still noticeable. On Sunday, the differences are very large after midday, when the actual data drops sharply while the forecasted data remains high.

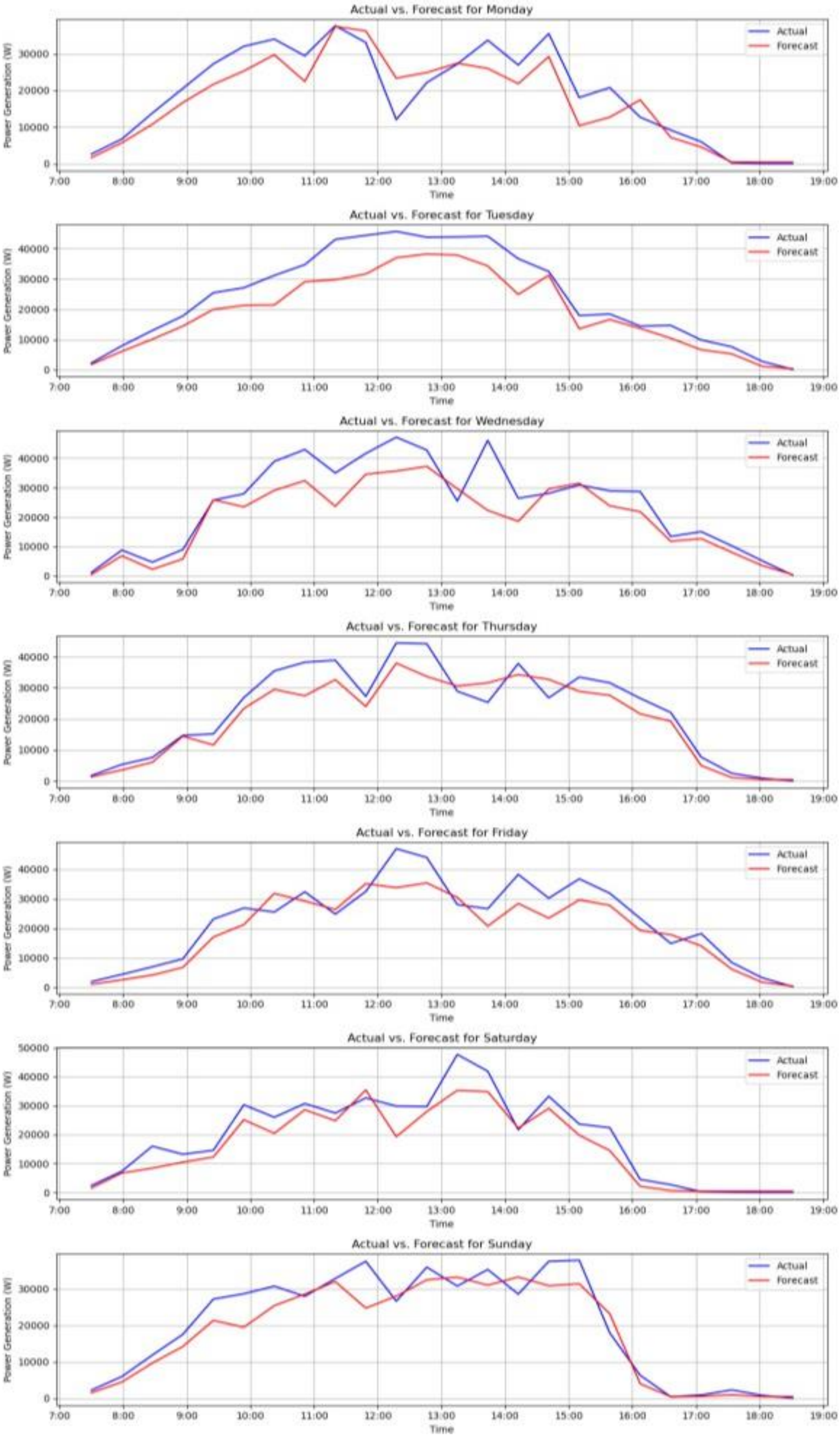


Fig. 4. Plot of forecast and actual data for model deployment

The objective of this research is to create an accurate model for forecasting AC power generation of a large-scale solar (LSS) farm in Malaysia's eastern coast region. A performance matrix analysis, comprising root mean squared error, mean squared error, and mean absolute error, was carried out. The study also investigates the possible effects short information on PV power prediction [18]. Furthermore, the solar farm required major maintenance and repair between 2021 and 2022, which resulted in a shortage of data. The data was insufficient long to capture seasonality and cycles, and key patterns and correlations were not collected during training to be transferred into the test dataset. By deploying ANN to the LSS forecasting method, it will not only have simple architecture but high accuracy. Despite the small inaccuracies in the model, ANN have big potential to contribute of forecasting AC power generation from the solar farm [19]. The model's performance has implications and limitations for the application of ANN to PV power forecasting. The model demonstrates that ANN can achieve high accuracy in predicting the AC power generation of a PV system with slightly error using only two input variables. On the other hand, the model also shows that ANN may not be able to capture the variability and uncertainty of the PV power generation, especially when applied to new and unseen data [20].

The model deployment indicates the true performance of the model. The model will forecast new historical dataset to evaluate the performance of the model. Based on the findings, the model's RMSE, MSE, MAE and R^2 are evaluated based on the performance of model training and testing with the actual deployment of the model. The average RMSE for the model is 10.955%, the average MSE for the model is 1.200%, the average MAE for the model is 6.805% and R^2 is 0.840 as shown in Table 2.

Table 2
Evaluation Metrics Table for Model Deployments

Evaluation metrics	Percentages
RMSE	10.955%
MAE	1.200%
MSE	6.805%
R^2	0.840

Table 3 indicates the difference in RMSE, MAE and MSE value for fundamental model and after model deployment. As shown in the table, the value of RMSE for model deployment is 10.95% which is slightly higher than the training model which is 8.513%. The value of MAE in model deployment is 1.200% higher compared training model 0.829%. The MSE of the model deployment is 6.805% which is also higher than training model with value of 5.631%. The R^2 decrease from 0.891 to 0.840. Based on the comparison, we can conclude the errors for each metrics is slightly increase including the changes in R^2.

Table 3
Evaluation Metrics Table for Model Deployments

Evaluation Metrics	Training Model	Model Deployment
RMSE	8.513%	10.955%
MAE	0.829%	1.200%
MSE	5.631%	6.805%
R^2	0.891	0.840

4. Conclusion and Recommendation

In this study, an artificial neural network (ANN) model is developed to forecast the alternating current (AC) power generation of a photovoltaic (PV) system. The model uses two input variables: temperature of the PV module and the total global horizontal irradiance (GHI) on the surface of the module. The model performance is evaluated by using three error metrics: root mean squared error (RMSE), mean squared error (MSE), and mean absolute error (MAE). We compare the model to the actual power output obtained from the PV system. The model is also applied to a new dataset to assess its ability to generalize to unseen data. The study shows that our ANN model can achieve high accuracy in predicting the AC power generation of a PV system, using only two input variables.

To improve the accuracy of our forecasting model, we recommend paying close attention to fine-tuning. This involves carefully adjusting the model's settings, such as learning rates, layer structures, and activation functions. By experimenting and finding the right balance, we can make the model more attuned to the unique patterns in our data. Regular fine-tuning sessions will help us zero in on the most effective configuration, narrowing the gap between our predicted and actual data. Expanding and enhancing the dataset with additional relevant information is a key step in boosting the model's predictive abilities [21]. Additionally, the ANN model can be deployed to the actual forecast to predict. Other than that, to improve or validate the model, we could use more input variables, such as weather data, solar angle, or module efficiency, to capture the complexity and dynamics of the PV power generation. We could also use cross-validation, regularization, or ensemble methods to reduce the overfitting and increase the generalization of the model. We could also compare the model with other forecasting methods, such as linear regression, support vector machines, or random forests, to evaluate its relative performance and suitability [22].

Acknowledgements
The author would like to thank Universiti Teknologi MARA (UiTM); and FRGS/1/2021/TK0/UITM/02/12 Ministry of Higher Education (MOHE); Malaysian Government for the financial support.

References
[1] Nair, Divya R., S. Devi, Manjula G. Nair, and K. Ilango. "Tariff based fuzzy logic controller for active power sharing between microgrid to grid with improved power quality." In *2016 International Conference on Energy Efficient Technologies for Sustainability (ICEETS)*, pp. 406-409. IEEE, 2016. https://doi.org/10.1109/ICEETS.2016.7583789

[2] National Renewable Energy Laboratory. "Energy Systems Integration Facility." *NREL*, 2018. https://www.nrel.gov/esif/.

[3] Singh, Astha, and Kishan Bhushan Sahay. "Short-term demand forecasting by using ANN algorithms." In *2018 International Electrical Engineering Congress (iEECON)*, pp. 1-4. IEEE, 2018. https://doi.org/10.1109/IEECON.2018.8712265

[4] Sözen, Adnan, Erol Arcaklıoğlu, Mehmet Özalp, and Naci Çağlar. "Forecasting based on neural network approach of solar potential in Turkey." *Renewable Energy* 30, no. 7 (2005): 1075-1090. https://doi.org/10.1016/j.renene.2004.09.020

[5] Abuella, Mohamed, and Badrul Chowdhury. "Solar power forecasting using artificial neural networks." In *2015 North American Power Symposium (NAPS)*, pp. 1-5. IEEE, 2015. https://doi.org/10.1109/NAPS.2015.7335176

[6] Youssef, Ayman, Mohammed El-Telbany, and Abdelhalim Zekry. "The role of artificial intelligence in photo-voltaic systems design and control: A review." *Renewable and Sustainable Energy Reviews* 78 (2017): 72-79. https://doi.org/10.1016/j.rser.2017.04.046

[7] Rahman, M. F. A., Rozana Zakaria, and Rosli Zin. "The importance of life cycle cost components for green highway and road management: A review." *Journal of Advanced Research in Technology and Innovation Management* 2, no. 1 (2022): 13-21.

[8] Das, Utpal Kumar, Kok Soon Tey, Mehdi Seyedmahmoudian, Saad Mekhilef, Moh Yamani Idna Idris, Willem Van Deventer, Bend Horan, and Alex Stojcevski. "Forecasting of photovoltaic power generation and model optimization:

A review." *Renewable and Sustainable Energy Reviews* 81 (2018): 912-928. https://doi.org/10.1016/j.rser.2017.08.017

[9] Jin, Yong-Chao, Qian Cao, Ke-Nan Wang, Yuan Zhou, Yan-Peng Cao, and Xi-Yin Wang. "Prediction of COVID-19 Data Using Improved ARIMA-LSTM Hybrid Forecast Models." *IEEE Access* 11 (2023): 67956-67967. https://doi.org/10.1109/ACCESS.2023.3291999

[10] Yadav, Amit Kumar, and S. Singh Chandel. "Solar radiation prediction using Artificial Neural Network techniques: A review." *Renewable and Sustainable Energy Reviews* 33 (2014): 772-781. https://doi.org/10.1016/j.rser.2013.08.055

[11] Narayanan, Niranjhana, Zitao Chen, Bo Fang, Guanpeng Li, Karthik Pattabiraman, and Nathan Debardeleben. "Fault injection for TensorFlow applications." *IEEE Transactions on Dependable and Secure Computing* (2022). https://doi.org/10.1109/TDSC.2022.3175930

[12] Leitão, David, João Paulo N. Torres, and João F. P. Fernandes. "Spectral irradiance influence on solar cells efficiency." *Energies* 13, no. 19 (2020): 5017. https://doi.org/10.3390/en13195017

[13] Rani, S. P., S. M. Giridhar, and S. R. Prasad. "Effect of temperature and irradiance on solar module performance." *IOSR Journal of Electrical and Electronics Engineering* 13, no. 2 (2018): 36-40.

[14] Wang, Kejun, Xiaoxia Qi, and Hongda Liu. "Photovoltaic power forecasting based LSTM-Convolutional Network." *Energy* 189 (2019): 116225. https://doi.org/10.1016/j.energy.2019.116225

[15] Hassan, Mohamed K., and A. N. Aldughmi. "Measurement of current-voltage characteristics, energy yield and the efficiency of the photovoltaic solar cell, as well as its applications in Saudi Arabia." *Journal of Mechanical Engineering Research and Developments* 44, no. 11 (2021): 186-196.

[16] Young, M. "The Technical Writer Handbook. Mill Valley." *CA: Science University* (1989).

[17] Khan, Akhlaque Ahmad, Ahmad Faiz Minai, Laxmi Devi, Qamar Alam, and Rupendra Kumar Pachauri. "Energy demand modelling and ANN based forecasting using MATLAB/simulink." In *2021 International Conference on Control, Automation, Power and Signal Processing (CAPS)*, pp. 1-6. IEEE, 2021. https://doi.org/10.1109/CAPS52117.2021.9730746

[18] Shim, Chung Siong, Chia Yee Ooi, and Giap Seng Teoh. "Low Power Integrated Circuit Design of Extreme Learning Machine using Power Gating Methodology." *Journal of Advanced Research in Computing and Applications* 31, no. 1 (2023): 13-19.

[19] Letchumanan, L. Thiruvarasu, Noordin Mohd Yusof, Hamed Gholami, and Nor Hasrul Akhmal Bin Ngadiman. "Green Lean Six Sigma: A Review." *Journal of Advanced Research in Technology and Innovation Management* 1, no. 1 (2021): 33-40.

[20] Abd Halim, Suhaila, Zuraida Alwadood, Norlenda Mohd Noor, and Hanifah Sulaiman. "Development of Quadric Surfaces Learning Tool for Engineering Students using MATLAB GUI." *Journal of Advanced Research in Computing and Applications* 17, no. 1 (2019): 6-13.

[21] Noor, Norlenda Mohd, Zuraida Alwadood, Nuramelissa Ahmad Murad, and Nur Maisarah Mohamad Termizi. "Optimization of Private Bus Scheduling in UiTM Shah Alam Using Integer Linear Programming." *Journal of Advanced Research in Computing and Applications* 15, no. 1 (2019): 26-34.

[22] Endut, Nor Adora, Mohammad Fahmi Mohd Fo'ad, Nor Azylia Ahmad Azam, Nor Ashitah Abu Othman, Siti Rahayu Abdul Aziz, and Anis Shobirin Abdullah Sani. "Real-Time Water Monitoring System for Fish Farmers Using Arduino." *Journal of Advanced Research in Computing and Applications* 14, no. 1 (2019): 10-17.