

Economic Impacts of Performance Optimization in Large-Scale Solar Farms: A Case Study Using Artificial Neural Networks in Eastern Malaysia

Muhammad Azfar Shamil Abd Aziz¹, Nurfadzilah Ahmad²

¹Solar Research Institute (SRI), Universiti Teknologi MARA (UiTM) Shah Alam, Selangor

²School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM) Shah Alam, Selangor

Corresponding Author Email: nurfadzilah6344@uitm.edu.my

To Link this Article: <http://dx.doi.org/10.6007/IJAREMS/v13-i3/22258> DOI:10.6007/IJAREMS/v13-i3/22258

Published Online: 28 August 2024

Abstract

This study aims to improve the performance analysis of a large-scale solar farm in Eastern Malaysia by using an Artificial Neural Network (ANN) approach. The study's goal is to optimize the solar farm's energy output by considering critical aspects such as weather, string level, shadowing, and dirt buildup. An ANN-based predictive model will be created to properly forecast energy output, allowing for exact diagnosis of possible difficulties and the implementation of corrective steps to improve overall system performance. The project will also entail the creation of an innovative monitoring and control system designed to continually track and analyze the solar farm's performance indicators. This technology will allow for real-time modifications, ensuring that the solar farm functions at maximum efficiency. The suggested technique intends not only to increase the solar farm's performance and dependability, but also to contribute to its long-term and economically feasible operation. This study aims to establish a baseline for the successful management of solar farms in the region by incorporating ANN approaches for predictive analysis and real-time monitoring, thereby minimizing environmental impact and promoting renewable energy. From the result, it is suggested that the model's predictions closely align with the actual power output trend throughout the day, indicating the neural network's ability to capture the general pattern and variability in solar power generation.

Keywords: Artificial Neural Network (ANN), Large-Scale Solar Farm.

Introduction

In recent years, there has been a rising appreciation for the critical role that renewable energy sources play in tackling the global energy problem and minimizing the negative consequences of climate change. Solar energy has emerged as a prominent and ecologically friendly alternative to traditional fossil fuels. Its relevance stems not only from its ability to generate clean energy, but also from its potential to transform energy production and consumption patterns. This paradigm shift towards solar energy is especially noticeable in areas with ample

sunshine, such as Eastern Malaysia. Here, a rare opportunity exists to use solar power to fulfil rising energy demands while also lowering greenhouse gas emissions that contribute to climate change (Alam et al., 2021). Despite the obvious promise of solar energy in Eastern Malaysia, its implementation is hampered by several complicated issues concerning the performance and efficiency of large-scale solar farms. The successful use of solar electricity is dependent on a complex interaction of numerous elements, such as the unpredictable nature of weather conditions, the complexity of local terrain, and the complexities inherent in the solar farm's system design (Amer et al., 2023). These factors impact the energy output and overall efficiency of solar farms. As a result, it is critical to conduct an extensive examination, measurement, and optimization of their performance.

This study intends to solve these issues by using an Artificial Neural Network (ANN) approach to improve the performance analysis of a large-scale solar farm in Eastern Malaysia. The study's goal is to create a prediction model that can properly estimate energy output by considering critical aspects such as weather conditions, string level, shadowing, and dirt buildup (Aswin et al., 2022). This model will help identify possible faults and conduct corrective procedures to improve the solar farm's performance. Furthermore, the creation of a sophisticated monitoring and control system will allow for continuous performance tracking, guaranteeing that the solar farm functions at maximum efficiency[4]. This research aligns well with Malaysia's strategic goals set in its renewable energy roadmap while also contributing to the worldwide drive towards clean and sustainable energy (Bimenyimana et al., 2017). By addressing current difficulties and meeting crucial demands in the quest for a sustainable energy environment, the proposed project demonstrates its commitment to national and international renewable energy goals. The use of the Artificial Neural Network (ANN) approach in this context is also remarkable, demonstrating a dedication to using modern and cutting-edge technology (Boukelia et al., 2021).

Renewable Energy (RE) refers to all powerful sources of energy that provide energy without the use of fossil fuels and pose no environmental or health risks. Solar energy is a key component of renewable energy, and photovoltaic (PV) modules use sunlight to generate electricity (Dandil & Gürgen, 2017). Solar energy has all the characteristics listed above and offers the most significant contribution to RE. However, there are a few issues linked with PV modules. Economic and technical aspects have a significant impact on the environmental and technological benefits of a PV system. It can only be made practical for users if public financing is provided, or government programs are implemented to boost reliance on renewable energy (Dahlan et al., 2022).

Weather factors including irradiance, temperature, humidity, and wind speed all have a substantial impact on PV module production. Some parameters have a stronger influence than others. The worst-case situation happens when a band of clouds sweeps over a big area containing multiple PV modules, resulting in a significant loss in power delivery to the grid. This condition is worsened during peak load hours, which has a significant influence on a PV module's ideal performance [8]. To avoid such instances, accurate forecast performance of PV module output power is critical for planning and operations. Efficient technologies such as Machine Learning (ML) and Deep Learning (DL) are required to achieve this consistent prediction performance (Hussain et al., 2023).

Several strategies have been used to maximize power generation from PV modules. The literature discusses Maximum Power Point Tracking (MPPT) and a variety of empirical methodologies[10]. However, these solutions need a thorough understanding of the physical properties and manufacturing specifications of PV modules, which are not always available. Adaptive approaches, like regression models and neural networks, have been used to forecast PV module performance (Liu et al., 2020). Studies have revealed that irradiance and temperature have a substantial impact on PV power output forecast. However, for accurate power output estimation, all environmental characteristics must be considered, and a standardized dataset comprising this information is required (Lu et al., 2022).

In this study, all essential environmental characteristics, such as temperature, irradiance, wind speed, and humidity, are used to accurately estimate PV module output power. The humidity parameter is later filtered out since it has a negative association with power production. An efficient ANN is trained on an experimental database of 46,682 data samples using a 15-layer structure and the Levenberg-Marquardt technique. The findings show a low MSE and an average regression value of 0.996 across the training, validation, and testing stages, demonstrating the ANN's ability to reliably forecast solar power output.

Methodology

Project Flow

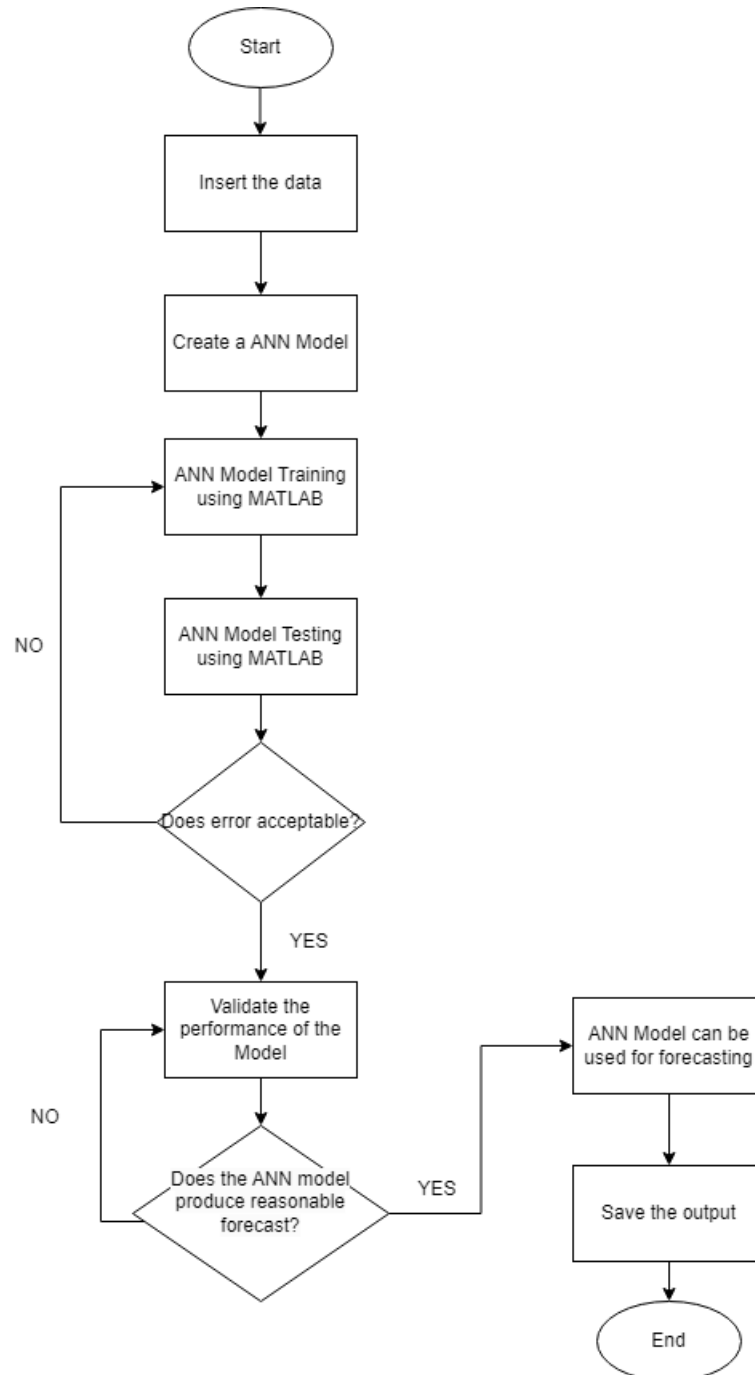


Figure 1: Flowchart Artificial of Neural Network for this Project.

Flowchart in Figure 1 shows the flowchart for analysing the performance of a solar farm with an Artificial Neural Network (ANN) follows a methodical set of stages. It begins with the collecting of historical data on critical variables such as weather conditions, module quality, shading, temperature, and energy production from the solar farm. The collected data is then pre-processed to guarantee cleanliness and uniformity, including the removal of missing values and outliers. Then comes feature selection, which involves identifying relevant input features for the ANN model, such as weather conditions and operating factors. ANN model development includes creating the architecture and training the network repeatedly with historical data. The validation and testing phases evaluate the model's generalizability and

predictive capabilities. After training, the performance analysis step assesses the accuracy of the ANN model by comparing its predictions to actual energy output using measures such as Mean Absolute Error or Root Mean Squared Error. Corrective steps are implemented in response to identified inconsistencies or concerns based on the analysis results. An important component is the implementation of a continuous monitoring and control system that incorporates the optimized ANN model. This technology provides continuous monitoring of solar farm performance, allowing for real-time modifications for optimal operation. Finally, the flowchart displays a cyclical process in which the ANN is regularly enhanced, and the solar farm's performance is constantly optimized based on the analysis's feedback loop

Data Description

The data used in this research for performance analysis is collected from historical records of solar power generation at the UiTM Large Scale Solar farm in Gambang. The data consists of PV module temperature, Global horizontal irradiance (GHI), Slope transient Irradiance for the inputs. The outputs are from PV module generated in Power (MW) which are from Ipv01 until Ipv15. The collected raw data consists of 46682 sets of 5 minutes' interval ranging from the month of December 2022 until December 2023. The usage of an Artificial Neural Network (ANN) as a robust regression model, emphasizing its capacity to make precise predictions. A major component of the approach is careful consideration of the specific geographical location, with the UiTM Solar Farm in Gambang serving as the focal point. This feature is essential for recording environmental information such as local weather conditions, solar radiation patterns, and topographical features, all of which have a significant influence on power generation. The study aims to demonstrate the utility of the modified ANN in delivering accurate forecasts tailored to the peculiarities of the UiTM Solar Farm in Gambang, therefore contributing significantly to the area of solar energy forecasting.

Data Pre-Processing

The data sets consist of multiple time series and each further partitioned into training and testing. The data used in this study comes from the UiTM Large Scale Solar Farm in Gambang, where historical records of solar power output were collected from December 2022 to December 2023. However, the dataset contains missing numbers, notably in July and August 2023, due to PV module failures. To solve this, missing data points were discovered and imputed using techniques such as linear interpolation, forward-fill, and Seasonal Decomposition of Time Series (STL). Data cleaning entailed deleting duplicates, finding and treating outliers using statistical methods, and performing consistency checks to verify data matched predicted physical and environmental parameters.

Data cleaning involved several steps to eliminate inconsistencies and ensure the dataset's integrity. Duplicate records were identified and removed to ensure that each data point was unique. Outliers, which could indicate measurement errors or unusual environmental conditions, were detected using statistical methods such as Z-scores and IQR analysis. Depending on their nature, these outliers were either corrected or excluded from the dataset. Consistency checks were performed to ensure the data aligned with expected physical and environmental conditions; for example, GHI values were cross-checked to ensure they fell within feasible ranges based on known solar radiation patterns for the region.

The data was then integrated and transformed, which included normalizing input variables using Min-Max scaling, storing temporal elements such as day of the week and hour of the day, and creating extra features such as moving averages to emphasize longer-term patterns. This extensive pre-processing guarantees that the input to the Artificial Neural Network (ANN) is strong and dependable. The rigorous cleaning, integration, and modification of the data allows the ANN to properly capture the underlying patterns and relationships in the solar power generating process. This pre-processing step is critical for the model to deliver precise and tailored forecasts for the UiTM Solar Farm in Gambang, thereby making a significant contribution to the field of solar energy forecasting by demonstrating the utility of the modified ANN in delivering accurate forecasts tailored to the specific geographical location.

Neural Network Architecture

The neural network architecture used in this study is a Feedforward Neural Network (FNN), which is meant to estimate solar power production based on input characteristics like PV module temperature. The network design starts with an input layer that uses the PV module temperature as its key characteristic. To achieve successful training, the `mapminmax` function normalizes the input data to a range of $[0, 1]$. The hidden layer consists of ten neurons selected by empirical testing to balance complexity and training time. Each neuron in the hidden layer has a sigmoid activation function, which introduces nonlinearity to capture complicated patterns in the data.

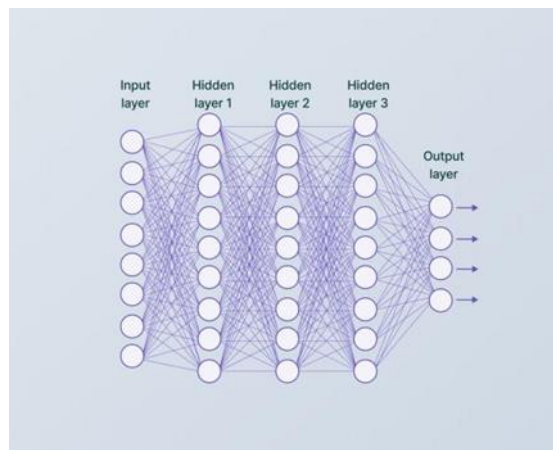


Figure 2: Artificial Neural Network Model

The output layer is made up of 10 neurons, which correspond to the power outputs from `lpv01` to `lpv15`. The output data, like the input data, is normalized using the `mapminmax` function. The network is trained with the Levenberg-Marquardt backpropagation method (`trainlm`), which is ideal for small to medium-sized datasets because to its rapid convergence. The dataset is split into three subsets: 70% for training, 15% for validation to avoid overfitting, and 15% for assessing the network's performance after training. The Mean Squared Error (MSE) metric is used to evaluate performance, since it calculates the average of the squares of the errors between projected and actual values (Pattanayak et al., 2023).

The neural network was developed using MATLAB's neural network toolbox. The training procedure included normalising the input and output characteristics, designing the network

architecture, and arranging the data division for training, validation, and testing. Following training, the network's performance was assessed, and the findings were displayed by graphing the anticipated power outputs versus the actual power outputs. The performance measurements demonstrated the network's capacity to generalise successfully to previously encountered data, with precise values for training, validation, and test performance

Model Evaluation and Analysis

The evaluation of the performance of the Artificial Neural Network (ANN) model was thoroughly assessed utilising a variety of critical metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and correlation coefficient (R). The MSE as shown in equation 1 calculates the average squared difference between expected and actual values, showing the total prediction error. The RMSE, or square root of the MSE, provides a more interpretable statistic by expressing the average size of the mistake in the same units as the target variable. Furthermore, the correlation coefficient (R) evaluates the linear connection between anticipated and actual values, with values approaching 1 indicating a significant positive association.

$$MSE = (1/n) * \sum(y_i - \hat{y}_i)^2 \quad (1)$$

- n is the number of data points (predictions).
- y_i is the actual value for the i -th data point.
- \hat{y}_i is the predicted value for the i -th data point.
- \sum denotes the summation over all data points.

At the performance phase, the neural network aims to minimize the MSE between the actual output power and the predictions value. The MSE is calculated for each instance in the training data, and the weights and biases of the neural network are adjusted through backpropagation to minimize this error.

```
% Sum the actual and predicted outputs
actualTotal = sum(filtered_data(:, {'Ipv01', 'Ipv02', 'Ipv03', 'Ipv04'},
predictedTotal = sum(predictedTargets, 1);

% Calculate MSE
mseValue = mean((actualTotal - predictedTotal').^2);
```

Figure 3: Snippet of code for calculation of MSE

The Mean Squared Error (MSE) calculation is used to assess the accuracy of the neural network model by comparing the expected and actual total power outputs. The computation starts with extracting and adding the actual and expected power outputs from the 15 individual PV modules. Specifically, the `actualTotal` variable stores the total of the actual power outputs for each time interval, whereas the `predictedTotal` variable stores the sum of the neural network's predicted power outputs for each period. The MSE is then calculated using a method that takes the average of the squared discrepancies between the actual and anticipated values. The code implements this as `mseValue = mean((actualTotal - predictedTotal').^2)`. The difference between actual and predicted values is squared to emphasise greater mistakes, and then averaged to give the MSE. This determined MSE value is presented in the plot title as a fast reference to the model's correctness. The MSE is an important indicator in this context since it quantifies prediction error, with lower values

suggesting better accuracy. Furthermore, because MSE is sensitive to bigger mistakes caused by difference squaring, it effectively emphasises major forecast disparities. Before examining the results, it is crucial to outline the experimental configuration. The simulation focused on evaluating the actual output power of Ipv01-Ipv05 on the predictive accuracy of the neural network. Additionally, the machine learning which is ANN approaches was used to predict the output power of solar photovoltaic panels. The feasibility of these approaches was shown by comparing the predicted data with the real one.

Evaluation Metric

Before focusing into the results, it is necessary to describe the experimental setup employed in this work. The experiment sought to assess the prediction accuracy of the neural network in anticipating solar power output, with an emphasis on the effect of the number of training epochs. In this arrangement, the learning rate, a critical parameter impacting convergence rate, was kept constant at 0.01. The number of epochs was adjusted to 100 to guarantee that the neural network, which included an 8-element hidden layer, had enough training iterations. The Mean Squared Error (MSE) was the major statistic utilised to evaluate the performance of the neural network.

This statistic quantifies the difference between the estimated and actual total power output of photovoltaic (PV) modules. The MSE is calculated in the given code by first adding the actual and anticipated power outputs from 15 individual PV modules, and then taking the mean of the squared discrepancies between these numbers. The resultant MSE number, which appears in the plot title, represents the model's predicted accuracy. A lower MSE indicates more accurate predictions, revealing the model's efficacy in capturing the general trend and variability in power generation.

Result and Discussions

Model ANN Performance

Table 1

Tuning Parameters and Options Chosen for ANN

Features	Values
Historical Data	46686
Training Algorithms	Levenberg-Marquardt
Performance	Mean Squared Error
Layer Size	15
Input and Output	GHI and PV Output

Hyperparameters were heuristically chosen in each trial for the Artificial Neural Network (ANN) implementation to maximize regression results while minimizing Mean Squared Error (MSE) throughout all three phases: training, validation, and testing. This entailed modifying parameters like the number of hidden layers, neurons per layer, learning rate, and training epochs by trial-and-error (Roy et al., 2021). The goal was to obtain a strong linear connection between anticipated and actual values (high regression values), while minimising average squared prediction errors (low MSE).

Table 2
Analysis of ANN results

Phases	Regression(R)
Training	0.99696
Validation	0.99589
Testing	0.99828
Total	0.99695

Regarding the alternatives mentioned in Table 2, the 'Levenberg-Marquardt' (LM) training algorithm was chosen due to its computational speed, making it the quickest of the available methods. The Mean Squared Error (MSE) was chosen as the performance metric since it is simple and useful for assessing forecast accuracy (Study of the effects of partial shading on PV array, n.d.). A layer size of 15 was picked because it gave higher training, validation, and testing accuracies than alternative configurations, resulting in improved overall performance in the neural network model.

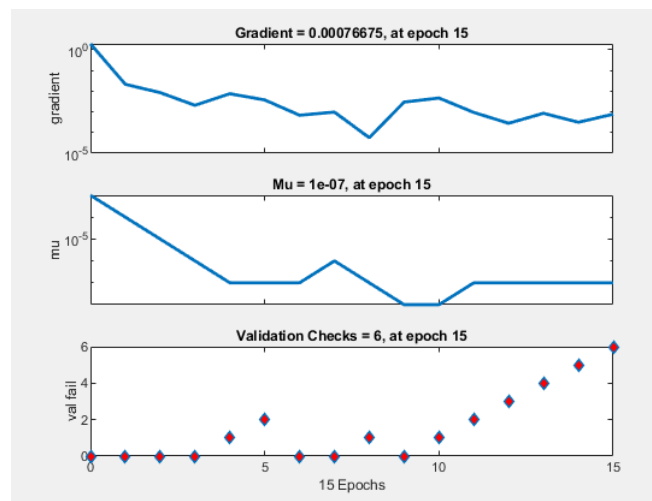


Figure 4: Gradient, Mu, and Validation Fail in ANN

Based on figure 4, at training phase, some criterion sets force training to be stopped when the validation criterion, epoch, performance, gradient, or mu values reach their predetermined levels. The training is halted at the 15th epoch owing to the maximum number of failures, i.e. a hyperparameter specified in tuning is fulfilled at the 21st iteration, i.e. 6, the gradient reached the value of 0.00076075, and mu reached the value of 1e-07 at that 15th iteration.

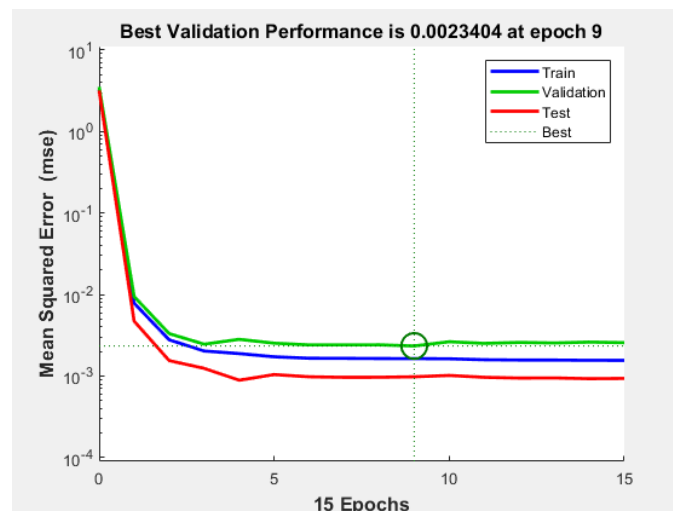


Figure 5: Plot of MSE vs No. of Epochs

Figure 5 shows a plot of Mean Squared Error (MSE) with the number of epochs for the training, validation, and testing stages, demonstrating how the MSE value changes with each iteration. This graphic depicts the model's learning curve and the point at which it performs best. The image shows that the best validation performance is reached at epoch 15, when the MSE value hits its lowest point of 0.0023404. This means that at this iteration, the model properly predicts the output while avoiding overfitting, as the validation error is minimized. This epoch is an important reference point for assessing the model's success and identifying the ideal stopping point during training to achieve peak performance.

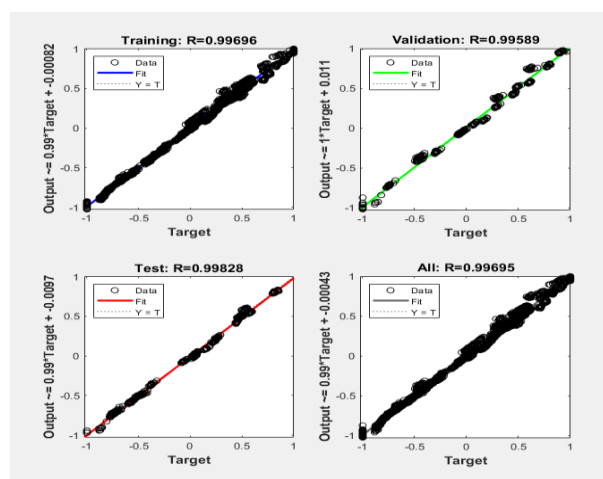


Figure 6: Regression Analysis Plot

The regression parameter is critical in evaluating the neural network model's effectiveness; an ideal value of 1 indicates a perfect match between anticipated and actual values. Figure 6 shows the link between predicted and target values using a regression plot. This plot has a fit line in each of the four subplots, which serves as a standard for determining prediction accuracy. Data points closer to this line show a better fit, with regression coefficients closer to the ideal of 1. The clustering of data samples in these subplots also illustrates the large dataset

used to train the neural network, which ensures the model's predictions are resilient and reliable (Classification and summarization of solar irradiance and power forecasting methods: A thorough review, n.d.).

During the training phase, the regression value attained was 0.99696, showing a very strong correlation between predicted and observed values. During the validation phase, the regression result is somewhat lower at 0.9959, indicating good predictive accuracy. The testing step produces the greatest regression value, 0.99828, confirming the model's outstanding generalization to new, previously encountered data. The total regression value, which includes all three stages, is 0.99695. This continuously high regression value across the training, validation, and testing phases demonstrates the model's efficacy and durability, proving its ability to reliably estimate solar power production across many datasets.

Graph Actual Output Power Ipv01-15 vs Predicted Output Power

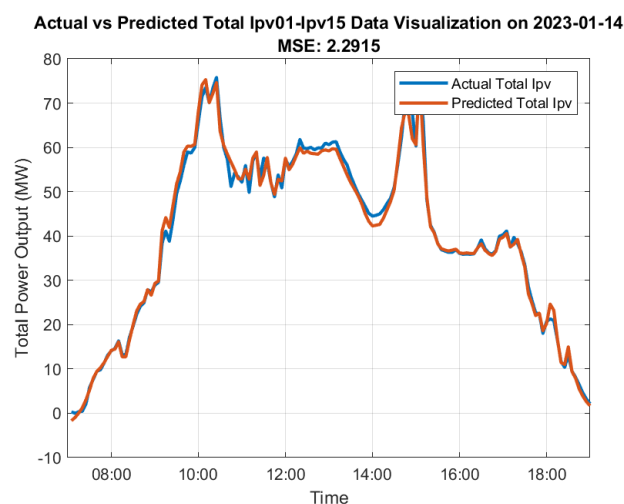


Figure 7: Actual vs Output Power Ipv1-15 in January 2023

Figure 7 shows a comparison of the actual and expected total power output (Ipv01-Ipv15) on January 14, 2023. The model's predictions closely align with the actual power output trend throughout the day, indicating the neural network's ability to capture the general pattern and variability in solar power generation. The alignment of the actual and anticipated values indicates that the model effectively learns from the input data and appropriately anticipates the power output. However, modest disparities at times suggest that the model's predictions differ significantly from the actual results. Despite these aberrations, the model's overall performance is still strong, as indicated by the tight clustering of predicted values around actual values.

The Mean Squared Error (MSE) value of 2.2915 is the average squared difference between the actual and predicted values, serving as a quantitative assessment of the model's accuracy. A lower MSE value implies higher predictive performance, and in this example, an MSE of 2.2915 shows that the model is doing effectively with few prediction mistakes. The graphic shows when the model's predictions are most accurate and where improvements might be made. For example, modest overestimations or

underestimations at peak power production times may indicate model adjustments, such as adding new input characteristics or fine-tuning the network design. Overall, the plot and MSE value validate the model's ability to estimate solar power generation while suggesting areas for additional improvement.

Conclusion

Based on the extensive study and findings given, using an Artificial Neural Network (ANN) as a predictive model for solar power generation at the UiTM Large Scale Solar Farm in Gambang produced promising results. The study aimed to exploit the ANN's capacity to reliably predict power outputs based on important environmental factors such as PV module temperature, GHI, and slope transient irradiance. The ANN was successfully trained using the Levenberg-Marquardt backpropagation method after thorough data preprocessing, which included normalisation and partition into training, validation, and testing sets.

In conclusion, this study highlights the potential of ANN-based techniques to improve solar energy forecasting accuracy. The work increases our understanding of solar power dynamics by combining powerful machine learning algorithms with extensive environmental data. It also gives practical insights for optimising energy generation and management tactics. Future study might look at further adjustments to the ANN design, the incorporation of other environmental characteristics, and validation across many geographical areas to widen the usefulness and resilience of such prediction models in renewable energy industries.

Contribution

This research significantly contributes to optimizing the performance of large-scale solar farms in Eastern Malaysia by employing the Artificial Neural Network (ANN) method. It enhances performance modeling by accurately predicting energy output based on complex environmental and operational factors. The study provides region-specific insights, crucial for tailoring solar farm operations to Eastern Malaysia's unique conditions. It offers practical recommendations for improving efficiency, such as optimizing equipment and site selection. Additionally, the research advances the application of ANN in renewable energy, highlighting its potential for broader energy forecasting and management.

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 and 100-TNCPI/INT 16/6/2 (003/2024); Universiti Teknologi MARA (UiTM) for the financial support

References

- Alam, M. N., Aziz, S., Karim, R., & Chowdhury, S. A. (2021). Impact of solar PV panel cleaning frequency on the performance of a rooftop solar PV plant. In 2021 6th International Conference on Development in Renewable Energy Technology (ICDRET) (pp. 1–4). <https://doi.org/10.1109/ICDRET54330.2021.9752681>
- Amer, H. N., Dahlan, N. Y., Azmi, A. M., Latip, M. F. A., Onn, M. S., & Tumian, A. (2023). Solar power prediction based on artificial neural network guided by feature selection for large-scale solar photovoltaic plant. *Energy Reports*, 9, 262–266. <https://doi.org/10.1016/j.egy.2023.09.141>
- Aswin, K. S., Purushothaman, M., Sritharani, P., & T. S. (2022). ANN and deep learning classifiers for BCI applications. In 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICT) (pp. 1603–1607). <https://doi.org/10.1109/ICICT54557.2022.9917834>
- Bimenyimana, S., Norensse, G., Asemota, O., Lingling, L., & Li, L. (2017). Output power prediction of photovoltaic module using nonlinear autoregressive neural network. *Journal of Energy, Environmental & Chemical Engineering*, 2, 32–40. <https://doi.org/10.11648/j.jeece.20170204.11>
- Boukelia, T. E., Kabar, Y., & Djimli, S. (2021). Development of an ANN model for forecasting time series of cooling performances in CSP plants - Solar tower power plant (Gemasolar) as a case study. In 2020 6th International Symposium on New and Renewable Energy (SIENR) (pp. 1–5). IEEE. <https://doi.org/10.1109/SIENR50924.2021.9631921>
- Dandil, E., & Gürgen, S. (2017). Prediction of photovoltaic panel power output using artificial neural networks learned by heuristic algorithms: A comparative study. In 2017 International Conference on Computer Science and Engineering (UBMK) (pp. 397–402). <https://doi.org/10.1109/UBMK.2017.8093423>
- Dahlan, N. Y., Zamri, M. S. M., Zaidi, M. I. A., Azmi, A. M., & Zailani, R. (2022). Forecasting generation of 50MW Gambang large scale solar photovoltaic plant using artificial neural network-particle swarm optimization.
- Hussain, M. T., Sarwar, A., Tariq, M., Urooj, S., BaQais, A., & Hossain, M. A. (2023). An evaluation of ANN algorithm performance for MPPT energy harvesting in solar PV systems. *Sustainability*, 15(14), 11144. <https://doi.org/10.3390/su151411144>
- Liu, D., Yang, Q., & Yang, F. (2020). Predicting building energy consumption by time series model based on machine learning and empirical mode decomposition. In 2020 5th IEEE International Conference on Big Data Analytics (ICBDA) (pp. 145–150). <https://doi.org/10.1109/ICBDA49040.2020.9101335>
- Pattanayak, B., Nanda, S., & Kumar, N. (2023). Performance review of a PSO trained ANN based MPPT in a grid-connected 3MW solar power plant. In 2023 IEEE 2nd International Conference on Industrial Electronics: Developments & Applications (ICIDeA) (pp. 187–192). <https://doi.org/10.1109/ICIDeA59866.2023.10295257>