EVOLUTION IN ELECTRICAL AND ELECTRONIC ENGINEERING
e-ISSN: 2756-8458

EEEE

Vol. 6 No. 2 (2025) 356-365
https://publisher.uthm.edu.my/periodicals/index.php/eeee

Neural Network Based Prediction of Solar Power Generation

Meeganthiran[1], Dirman Hanafi[1]*, Ariffuddin Joret[2]

[1] *Department of Electrical Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400, Batu Pahat, Johor, MALAYSIA*
[2] *Department of Electronic Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia,86400, Batu Pahat, Johor, Malaysia*

*Corresponding Author: dirman@uthm.edu.my
DOI: https://doi.org/10.30880/eeee.2025.06.02.043

Article Info

Received: 19 June 2025
Accepted: 26 August 2025
Available online: 30 October 2025

Keywords

Neural Network, Solar Power Forecasting, Arduino Mega, Environmental Sensors, NARX, Time-Series Prediction, MATLAB, MSE.

Abstract

This project shows the development of solar power forecasting by using a recurrent neural network-based time-series model, typically a Nonlinear Autoregressive model with External Inputs (NARX) to predict the solar power produced by using real-time environmental factors. Multiple sensors are integrated with an Arduino Mega 2560 to detect important environmental factors such as sunlight intensity, temperature, humidity, wind speed, rainfall, atmospheric pressure, voltage and current. These data were logged to an SD card every 5 minutes over two months period and pre-processed for the training. The set of data is then imported to MATLAB to train NARX model using the Levenberg-Marquardt algorithm. This model achieved a best validation performance of 0.00035075 MSE, showing a high accuracy in analysing the non-linear relationships between environmental conditions and solar power output. The time series graph also shows that the predicted values are closer to the actual values during training, validation and testing stages. This forecasting method highlights the importance of using the NARX model in improving the solar energy under varying weather conditions.

1. Introduction

Solar energy holds the highest energy capacity of any renewable energy source (RES), and PV arrays direct electricity generation from sunlight. The sun is the primary energy supply for a lot of environmental processes such as heat, wind, and rain. Additionally, solar power is green and eco-friendly. The prediction of power output from solar energy is seeking special attention due to the increased power output from solar energy [1].

Artificial neural networks are widely used nowadays in various fields, including financial traders, data analysis, managers, doctors. They are mainly used for the voice and image recognition task because they provide accurate results. Despite being based on mathematical neuron models which was found half a century old, ANNs remain its popularity due to their simplicity and satisfying results [2]. So, their application makes them suitable tool to predict solar power output which influences by different environmental parameters.

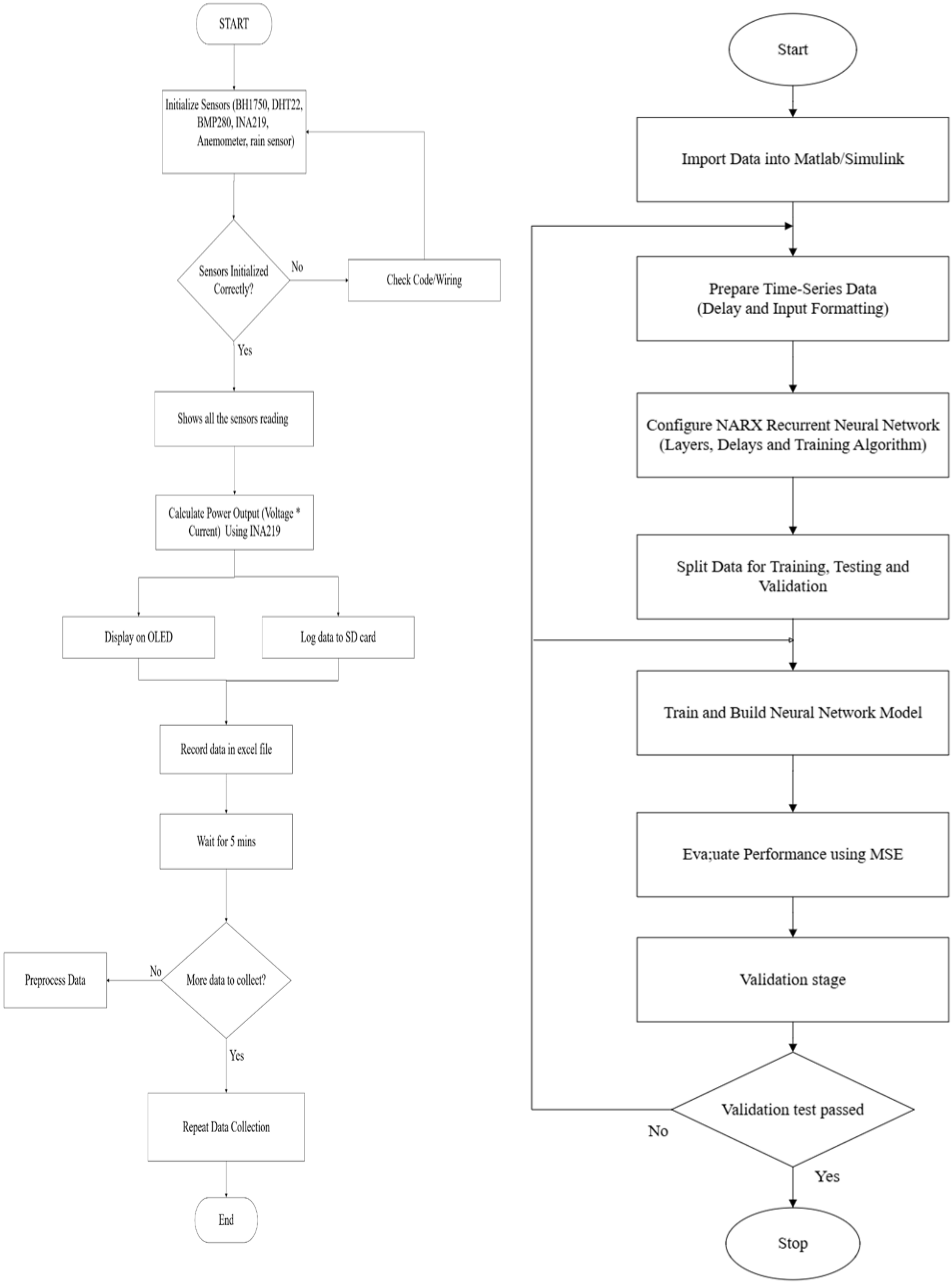
ANN's method of prediction was suggested to determine the heat from the sun [3]. The photovoltaic panel power output depends on the solar radiation, temperature, and humidity [4]. Neural networks can simulate the learning function like the human brain with minimum computational effort and so it can recognize an input-output relationship for both linear and non-linear systems. Neural networks are known as the machine learning model that usually deals with the nonlinear relationship system. Traditional methods for solar power prediction not usually accurate due to its complexity and non-linear relationship. For weather prediction to become more accurate, it is beneficial for making the neural network as one of the suitable methods to determine the solar energy output with more accuracy [5].

As the world moves toward renewable energy, improving solar energy generation has become vital for both economic and environmental sustainability of the environment. However, environmental parameters such as the intensity of sunlight, temperature, humidity, raindrops, wind speed, and atmospheric pressure crucially impact the solar panel's power output. These environmental factors interact in a complex, nonlinear way, making exact prediction of panel power output under various weather conditions complex [6]. To overcome this problem, three objectives were implemented, which are (1) to employ sensors and an Arduino Mega 2560 for collecting comprehensive datasets and solar panel output power consistently, ensuring a complete dataset for neural network training, testing and validation, (2) to train a NARX time-series based recurrent neural network model to predict the solar panel power output accurately based on the environmental parameters input such as sunlight intensity, temperature, humidity, wind speed, raindrops, and atmospheric pressure, voltage and current, (3) to determine how climate factors influence the solar panel power generation and optimize the solar energy usage and management. An Arduino Mega 2560 microcontroller was used as the interface for the hardware and is responsible for collecting data from multiple environmental sensors. The scope of this project builds a NARX recurrent neural network model in MATLAB 2024b using the Levenberg-Marquardt algorithm, where data from multiple sensors is recorded every 5 minutes to maintain the precision for training and collect a comprehensive dataset over a 1 to 2-month period to enhance the accuracy and reliability of the neural network model. Additionally, this method not only improves the understanding of how climate factors affect the solar output but also supports solar energy usage and improves its efficiency.

This project uses Artificial Neural Network to train the based on the real-time environmental data such as sunlight intensity, temperature, humidity, voltage, current, raindrop, windspeed and atmospheric pressure. To perform the recurrent neural network, MATLAB NARX with exogenous inputs was chosen to perform the time series statistical model for training, testing and validation. The aim is to predict the solar power output accurately which forecast under varying environmental conditions, allowing more systematic renewable energy planning.

2. Methodology

The Fig 1 (a) outlines a flowchart that starts with the initialization of the hardware system, where all the sensors are initialized to begin the data acquisition. Once all the sensors are working perfectly, the data collection begins and the INA219 sensor measure the real time voltage and current to calculate the power output from the solar panel. After the data is displayed and collected, it is logged to the SD card and directly saved into the CSV format. The data is collected for every 5 minutes to ensure consistent time intervals for data logging. Next, Fig 1(b) illustrates the flowchart for the neural network solar prediction system by using a NARX (Nonlinear Autoregressive with Exogenous Inputs) which is a time-series recurrent neural network model. It begins by importing the data into the MATLAB in CSV format after preprocessing the data. The next phase is the preparing the time-series data. The next stage is the configuration of NARX recurrent neural network, including the adjustment of the number of hidden layers, neurons and delays. The network was trained by using Levenberg-Marquardt algorithm and adjust the weight and biases. Meanwhile, Mean Squared Error (MSE) is used to evaluate the system performance. it proceeds to the prediction stage, where new environmental data is use to predict the solar power expected output if not the model is retuned or retrained. This entire flowchart outlines how a neural network used to solve the challenges in forecasting solar power under extreme climate changes.



(a) **(b)**

Fig. 1 *Flowchart (a) Data Collection process of the environmental; (b) NARX neural network process in MATLAB*

Evolution in Electrical and Electronic Engineering Vol. 6 No. 2 (2025) p. 356-365

2.1 Prototype of the project and hardware connections

The Fig 2 shown illustrates a sensor-based weather monitoring system for predicting solar panel power output by using an Arduino Mega 2560. Multiple sensors are interfaced with Arduino Mega 2560, including the DHT22, BH1750, BMP280, 3-Cup Anemometer and raindrop sensor. A solar panel of 12V 10W, which is connected to INA219 with a resistor as a load of 20 W 15 ohm to measure voltage and current values and calculate the power produced in real-time environmental conditions was also interfaced with Arduino. Additionally, a SD card module was used to log the data collected and an OLED display was included.

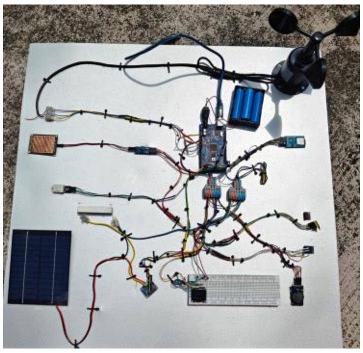


Fig. 2 *Hardware Integration of environmental sensors with Arduino Mega 2560*

2.2 Designing NARX model parameters

The Table 1 outlines the NARX model parameters and specifications that is required to build the recurrent neural network in MATLAB which is shown below.

Table 1: NARX model parameters

Dividing Data	Dividing Function: Dividerand
	Training Ratio: 70%
	Validation Ratio: 15%
	Testing Ratio: 15%
Network Architecture	Input Delay: 1:2
	Feedback Delay: 1:2
	Number of Hidden Neurons: 10
Training Algorithm	Levenberg-Marquardt backpropagation (trainlm)
Performance Function	Mean squared error
Plot Functions	Performance, training state. error histogram. regression, time-series response, error autocorrelation, input-error cross-correlation
Transfer Function	Hidden Layer: Hyperbolic tangent sigmoid
	Output Layer: Hyperbolic tangent sigmoid
Training Network	Open Loop Network
	Closed-Loop Network

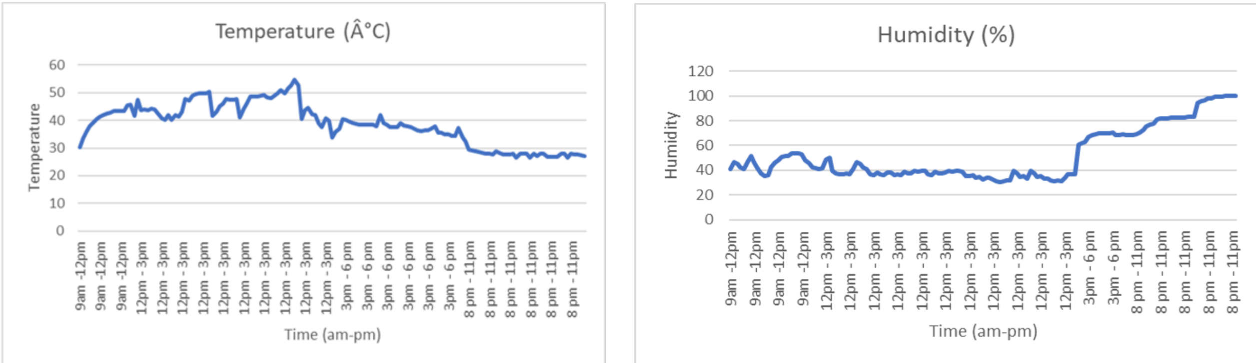
3. Results and Discussion

3.1 Analysis of the environmental parameters at different time intervals

In this case, recording the maximum and minimum values for each environmental factor is crucial to define the operating range of solar power output and to determine the best- and worst-case conditions during day and night. The data is collected continuously to obtain accurate neural network training, which enables the model for better training, more realistic and improves the training accuracy.

3.1.1 Temperature and Humidity VS Time

Fig 3 (a) shows the line graph of temperature over time. It shows the patterns of the temperature change at different time intervals. The hottest time was observed during the period from 12pm to 3pm, with an average temperature of 45.77°C. The highest temperature recorded at this time is due to the sun being at its peak, so the environment becomes hotter. The lowest average value of 27.25°C were recorded from 8pm to 11pm because the sun had disappeared at night. Fig 3 (b) shows the data of humidity recorded at different time intervals and shows the pattern of how humidity changed when time changes. The highest humidity recorded at 91.6% from 8pm to 11 pm because the humidity levels meant to be highest due to cool surroundings. Meanwhile, the lowest humidity recorded at 45.3% from 12pm to 3pm because of the heat at its peak so evaporation occurs in that situation.

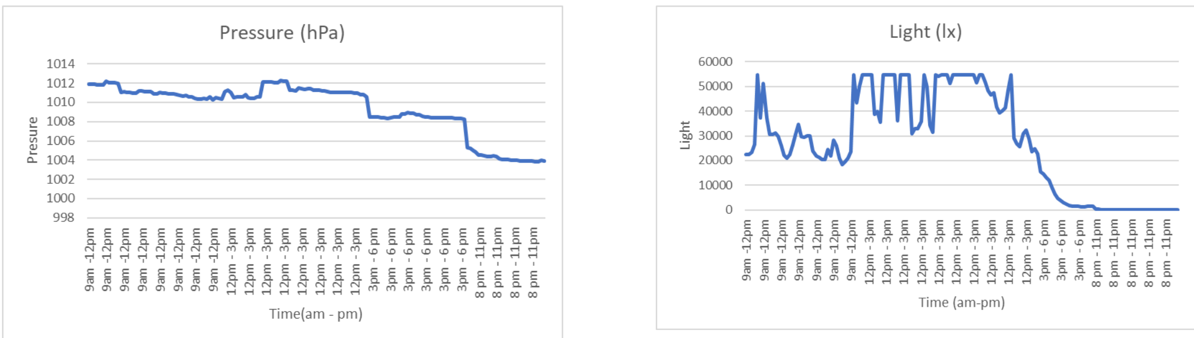


(a) (b)

Fig. 3 *Analysis of environmental factors vs time (a) Temperature; (b) Humidity*

3.1.2 Atmospheric pressure and Sunlight Intensity VS Time

Fig 4 (a) shows the atmospheric pressure over time graph which highlights how the pressure changes at different time intervals. The highest pressure recorded from the time interval of 12pm to 3pm which is 1011.20 hPa and the highest average recorded being at 1011.88 hPa. Higher pressure during midday is common as the atmosphere is more stable. Meanwhile, the lowest value is recorded during the time interval of 8pm to 11pm where the value is being recorded at 1006.3 hPa and the average value is 1004.25 hPa. Next, Fig 4 (b) shows the light intensity measured in lux, where it changes according to the day and night light intensity. The midday afternoon from 12pm - 3pm has the highest average light intensity of 46,971.42 lx because of the sun is at its peak and the light intensity is at its strongest at this time. The highest value recorded at 54,612.5 lx during this time interval which indicates the greatest intensity during this time interval.

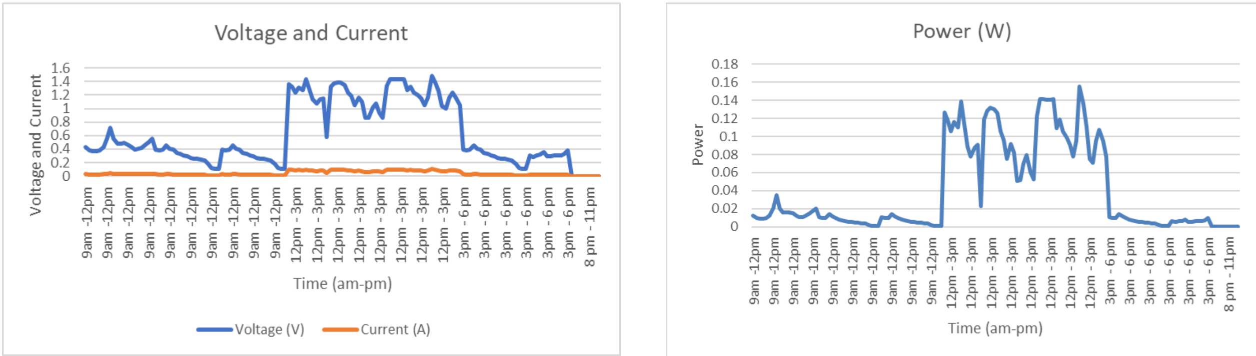


(a) (b)

Fig. 4 *Analysis of environmental factors vs time (a) Pressure; (b) Sunlight Intensity*

3.1.3 Voltage, Current and Power VS Time

Fig 5 (a) shows the voltage and current graph vs time at different time intervals which shows how solar panels functions by absorbing the heat from the sun to produce voltage. During the early afternoon from 12pm to 3pm, the voltage and current reaches the highest values. The highest voltage reading is at 1.49 V and the current is at 0.104 A. As the sun is at peak, the solar energy production is maximum at this time intervals. The average for the voltage and current is 1.21 V and 0.0828 A from 12pm to 3pm. Meanwhile, the lowest reading recorded at night which is 0 because there is no sun so no heat is generated during night. Next, Fig 5 (b) shows the highest power output during early afternoon from 12 pm to 3pm where the sun is at its peak. The average power recorded at this time interval is 0.1023 W, with the maximum and minimum values recorded at 0.155 W and 0.0226W. The values are high because the sunlight hitting direct to the panels, maximize the energy production when the sun is hot. At night period from 8pm to 11 pm, the power output becomes zero since there is not sunlight during nighttime. The average, minimum and maximum value recorded is 0 W.

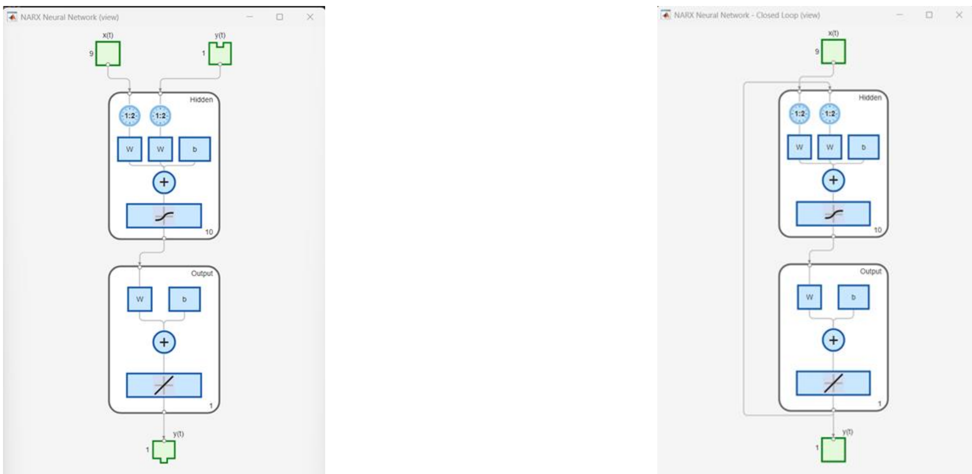


(a) **(b)**

Fig. 5 *Analysis of environmental factors vs time (a) Voltage and Current; (b) Power*

3.2 Open and Closed loop of NARX model

Fig 6 (a) shows the architecture of a NARX (Nonlinear Autoregressive with Exogenous Inputs) Neural Network for a time-series forecasting data prediction. This is an open-loop network where it is trained without the feedback to allow for simple training so that the model can learn the basic relationship between inputs and outputs. Additionally, the architecture consists of three parts including input layer, hidden layer, and an output layer. The input layer refers as x(t) with a size of 9 which represents the input environmental variables that influences the time series. There are 10 neurons in the hidden layer in the network to handle complex relationship. Furthermore, the ratio 1:2 is the feedback delays and input delays. The output layer y(t) which produce the final output using a linear activation function. The fig (b) shows closed-loop network shows the output y(t) has feedback into the system after each prediction which allows the network to use its own output as the input for future prediction. This feedback loop function is to make multi predictions for this model where each prediction influences the next.



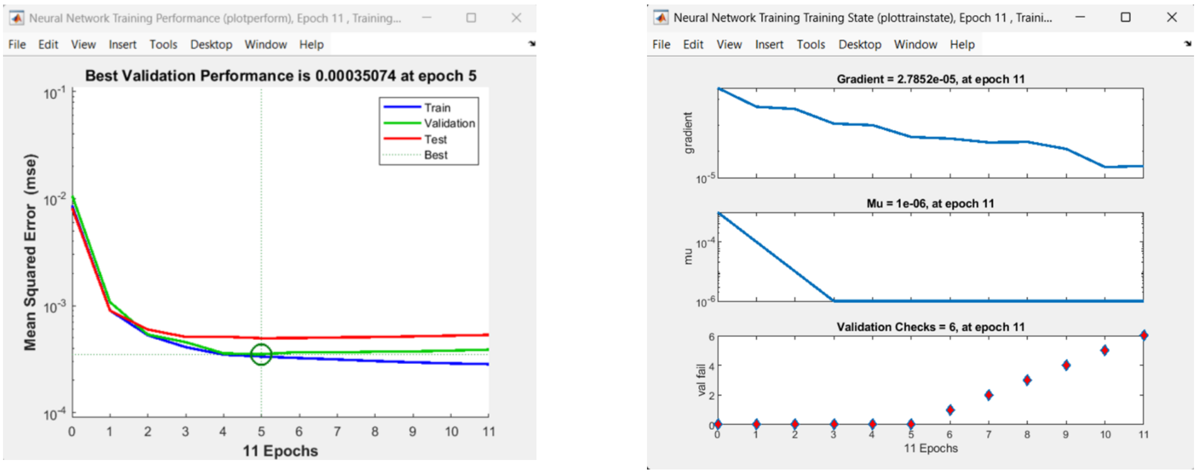
(a) **(b)**

Fig. 6 *NARX model (a) Open-Loop; (b) Closed-Loop*

Penerbit UTHM

3.3 Performance plot and Training state plot

Fig 7 (a) illustrates the performance graph of Mean Squared Error (MSE) for the train, validation, and test datasets vs the epochs. The blue curve represents the training section of the MSE where it decreases quickly at the initial epochs. This shows that the model learns quickly predictions during the training of the data. Next the green represents the validation of MSE, which monitor how great the model generalizes to a separate dataset validation. The validation graph decreases sharply until epoch 5, where at epoch 5 it shows the best validation performance at the value of 0.00035. This indicates that the model performed well even after it reached the best validation set at epoch 5. Furthermore, the red curve shows the test of MSE for the unseen data. At the beginning, the error of the test decreases until epoch 5. Next, Fig 7 (b) shows the Training State graph which provides a detailed view regarding the learning progress of the neural network. The graph on the top shows the gradient values vs 11 epochs. By the epoch 11, the gradient is at the value of 2.79e-05 which indicates the model learning rate has been slow down. Furthermore, the middle graphs the Mu vs the 11 epochs graph. Initially, the value of Mu is at higher value, where the weight is updated at the early stages. The Mu decreases as the model reaches the optimal and by epoch 11, it becomes stable at 1e06. The bottom graph shows the validation condition graph vs 11 epochs. The graph shows 6 times validation checks were performed and complete all the target value so that the model did not fail on these checks.

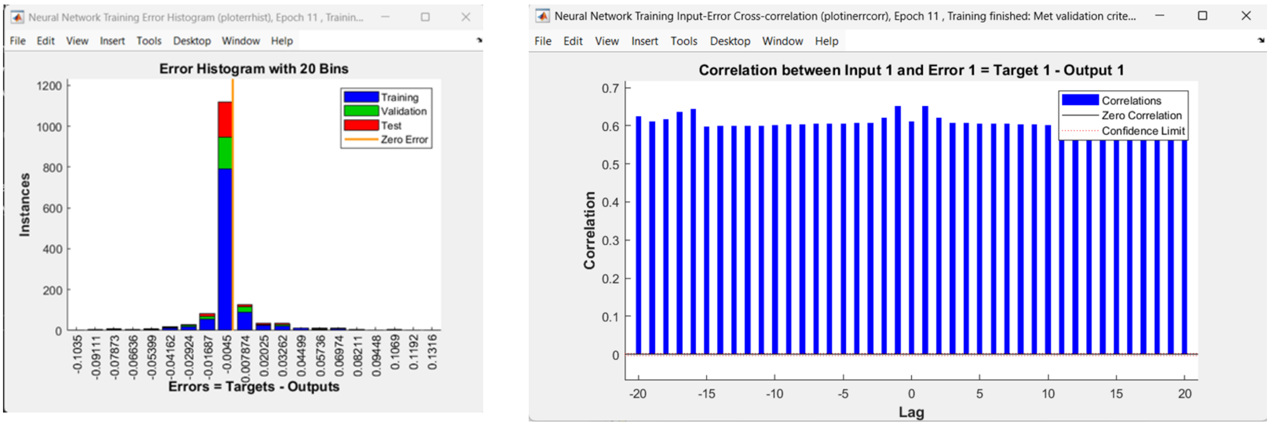


(a) **(b)**

Fig. 7 *Plot (a) Performance; (b) Training State*

3.4 Error Histogram Plot and Input-Error Cross-correlation Graph

The Fig 8 (a) shows the error histogram which distributes the errors (target - output). The errors are with 20 bins which means it show how the prediction of the model deviate from actual targets. The blue bar graph represents the training dataset which consist of large portion which is close to zero around the middle of the histogram, allowing the model to make accurate predictions. Next, the green bar is also same as the blue bar which concentrated at zero even though the validation bar is fewer instances compared to the training bar. Furthermore, the red bar represents the test dataset with fewer instances of zero errors. The Fig 8 (b) shows the relationship of input data and the error calculated between the target and output. This type of analysis used in time series forecasting so it can determine how the data in the past or future influences the prediction error. The graph shows that the correlation is quite strong at multiple lags because the values of the bar is close to 0.6 or higher. This means the error is heavily influenced by the input of the data from negative lags and positive lags. Overall, the graph highlights the prediction and the error are nearly linked to the input data at different lags, which helps to identify some areas for further improvement in the model's structure. From the input correlation graph, it can be observed that the temperature and sunlight intensity recorded the highest influences on the output, followed by humidity and atmospheric pressure. Wind speed and rainfall has the lesser impact on the output of the solar due to limited changes during the collection period.



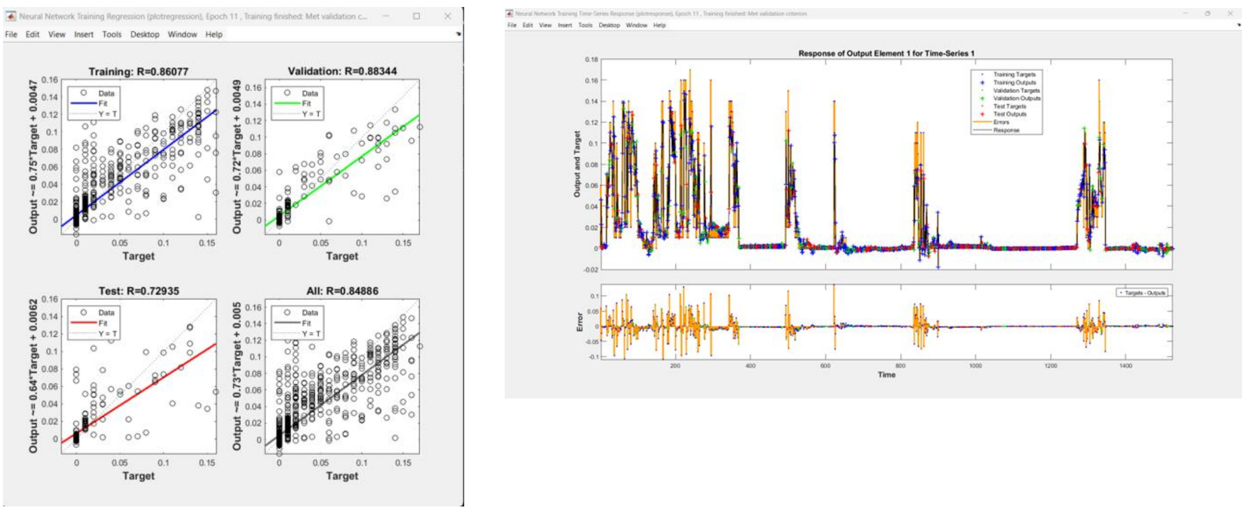
(a) **(b)**

Fig. 8 *Graph (a) Error Histogram; (b) Correlation between Input and Error*

3.5 Regression Plot and Time-series Response Graph

Fig 9 (a) shows the regression training graphs of four type of phases including training, validation, test and entire dataset. The R-value is the correlation coefficient where it predicts how close the predicted values are to the target values. In this case, all the values are nearly closer to 1, considered better performance. For training data regression plot, the R value is 0.86077, the model achieved a prediction accuracy of 84.87%, showing a strong correlation between the predicted and actual values. For the training data regression, the R value is 0.72935 which drops a bit, showing a slightly drop in performances on the test data but it is not a major problem because the test data is till decent. For the Validation data regression, the value of R value is 0.88344 which is better in performances than the training set. Finally, for the All the data combined regression plot, the R value is 0.84866, which indicates a strong line when combining with training, validation, and test datasets. The grey regression line fits very well with data points.

The Fig 9 (b) shows time-Series Response plot that provides useful information regarding the performances of the NARX neural network model. The blue positive sign (small) represents the training targets (actual values), and the blue big positive sign represents the predicted outputs for the training set of the model. The model trains the data fairly well but there is error spike that is noticeable shown by the orange lines, where the model struggles a bit to train which can be due to rapid changes in the data. Following, the green sign represents the validation target and the green plus bigger signs show the validation output for predicted. The model performs well but encountered major differences during sudden changes. The red small sign represents the test target while the red big plus sign represents the test outputs predicted. The test predictions shift more than actual test values when compared to training and validation datasets because the data is volatile in some specific regions. The graph in the bottom shows the error over time. The orange lines which represent the error where it fluctuates when it predicts the values over time. The errors are around time steps of 400 to 700.



(a) **(b)**

Fig. 9 *Graph (a) Training Regression; (b) Time Series Response*

3.6 Results Comparison with other Models and error metrics

Table 2 shows the comparison of results with others and error metrics vs this project. When comparing the Mean Squared Error (MSE) on different forecasting models, it is shown that the NARX model is the best among the others in the term of accuracy. This project achieves MSE of 0.00035 which is considered as the best among the listed papers.

Table 2: Results Comparison with other Models and error metrics

Author	Model Type & Inputs	Target Variable	Error Metrics
Boussaada et al. (2018) [7]	NARX (cloud cover + past radiation), 15 hidden neurons, sigmoid/tansig	solar radiation	MSE 0.00279
Mamoon et al. (2024) [8]	LSTM, GRU, CNN-LSTM, Random Forest,	Combined solar and wind forecast	MSE 0.010 for LSTM
Khan et al. (2024) [8]	LSTM	Solar and wind power production	MSE 0.00876
Sharma et al. (2016) [9]	NARX ANN	PV output p ower	MSE 7.5
Ahmad and Anderson (2015) [10]	NARX	Global solar radiation	MSE 0.0591-0.0722
Sandhya and Kavitha (2015) [11]	NARX	Solar radiation	MSE 1.5-6 and Regression Analysis 0.16-0.78
Cai et al. (2010) [12]	NARX ANN MLP ANN	PV output power	MAPE 16.47%-30.72%
Our project	NARX ANN	Solar power output	MSE 0.00035

4. Conclusion

This project has successfully achieved all three primary objectives including the developing and validating the NARX time-series based recurrent neural network forecasting model to predicts the solar power produced using the real-time environmental data. The sensors are integrated with an Arduino Mega 2560 to collect the climate values accurately including the temperature, humidity, wind, light, atmospheric pressure, voltage, current and raindrop. The collected values of the data were logged in the SD card and then it was processed for NARX model training of recurrent neural network in MATLAB. The model was trained using Levenberg-Marquardt Algorithm and validated by time-series analysis. The model trained shows the output results with accuracy where the predicted power output was close to the measured values across training, validation and testing stages even when the extreme weather changes. The model achieved a low Mean Squared Error (MSE) of 0.00035075 where it indicates 99.96% of accuracy considered a high prediction model. This project shows the NARX network model is able to predict the solar power produced. The three objectives of this project were successfully achieved. Firstly, complete hardware was implemented to collect the data. Next, the NARX was effectively trained and validated, the third objective is the effect of the environmental conditions towards solar output was clearly analyzed. The methodology also shows that these outcomes show the effectiveness of using the NARX time series recurrent model for handling the nonlinear relationship between environmental inputs and solar output power.

Acknowledgement

The authors would like to thank the Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia for its support.

Conflict of Interest

Authors declare that there is no conflict of interest regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Meeganthiran, Dirman Hanafi;* ***data collection:*** *Meeganthiran;* ***analysis and interpretation of results:*** *Meeganthiran, Dirman Hanafi, Ariffuddin;* ***draft manuscript preparation:*** *Meeganthiran, Dirman Hanafi. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] L. Mart´ın, L. F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, and M. Cony, “Prediction of global solar irradiance based on time series analysis: application to solar thermal power plants energy production planning,” Solar Energy, vol. 84, no. 10, pp. 1772–1781, 2010.

[2] K. S. Sayarkin, A. V. Popov and A. A. Zhilenkov, "Spiking neural network model MATLAB implementation based on Izhikevich mathematical model for control systems," 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), Moscow and St. Petersburg, Russia, 2018, pp. 979-982, doi: 10.1109/EIConRus.2018.8317253.

[3] T. Khatib, A. Mohamed, K. Sopian, and M. Mahmoud, “Assessment of artificial neural networks for hourly solar radiation prediction,” International Journal of Photoenergy, vol. 2012, Article ID 946890, 7 pages, 2012.

[4] A. V. Timbus, R. Teodorescu, F. Blaabjerg, and U. Borup, “Online grid measurement and ENS detection for PV inverter running on highly inductive grid,” IEEE Power Electronics Letters, vol. 2, no. 3, pp. 77–82, 2004.

[5] J. Gaboitaolelwe, A. M. Zungeru, A. Yahya, C. K. Lebekwe, D. N. Vinod andA. O. Salau, "Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison," in IEEE Access, vol. 11, pp. 40820-40845, 2023, doi: 10.1109/ACCESS.2023.3270041.

[6] Das, M. R. (2019). The effect of various environmental factors on the efficiency of solar panels. International Journal of Innovative Technology and Exploring Engineering, 8(11), 15–18. [Online] Available: https://doi.org/10.35940/ijitee.j9889.0981119

[7] Boussaada, Z., Curea, O., Remaci, A., Camblong, H., & Mrabet Bellaaj, N. (2018). A Nonlinear Autoregressive Exogenous (NARX) Neural Network Model for the Prediction of the Daily Direct Solar Radiation. Energies, 11(3), 620. [Online] Available: https://doi.org/10.3390/en11030620

[8] Khan, S., Mazhar, T., Khan, M. A., Shahzad, T., Ahmad, W., Bibi, A., Saeed, M. M., & Hamam, H. (2024). Comparative analysis of deep neural network architectures for renewable energy forecasting: enhancing accuracy with meteorological and time-based features. Discover Sustainability, 5(1). [Online] Available: https://doi.org/10.1007/s43621-024-00783-5

[9] Sharma, G., Pandey, A. and Chaudhary, P. 2016. Prediction of output solar power generation using neural network time series method. International Conference on Electrical Engineering (ICEENG), 19-21 April, 10, Cairo, Egypt, 1-5.

[10] Ahmad, A., Anderson, T. and Lie, T. 2015. Hourly global solar irradiation forecasting for New Zealand. Solar Energy, 122, 1398-1408.

[11] Sandhya, T. and Kavitha, V. 2015. Estimation of solar radiation with various climatic parameters based on neural network. International Journal of Emerging Technology in Computer Science \& Electronics (IJETCSE), 13, 151-155.

[12] Cai, T., Duan, S. and Chen, C. 2010. Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement. International Symposium on Power Electronics for Distributed Generation Systems, 2, 773-777.