Journal of Telecommunication, Electronic and Computer Engineering
ISSN: 2180 – 1843 e-ISSN: 2289-8131 Vol. 16 No. 4 (2024) 31 – 36
jtec.utem.edu.my
DOI: https://doi.org/10.54554/jtec.2024.16.04.005

Machine Learning-Based Solar Irradiance Forecasting Model Using GPS

Ho. Y.H.[1]*, Tang. S.Y.[1], and Guo. J.Q.[2]
[1]*Centre of Technology for Disaster Risk Reduction (CDR), Universiti Teknikal Malaysia Melaka (UTeM), Melaka, 76100, Malaysia.*
[2]*School of Integrated Circuit, Shandong College of Electronic Technology, Jinan, 250200, China.*

Article Info

Article history:
Received Aug 16th, 2024
Revised Nov 7th, 2024
Accepted Dec 12th, 2024
Published Dec 20th, 2024

Index Terms:
Solar Irradiance Forecasting
Machine Learning Algorithms
Zenith Total Delay (ZTD)
Total Electron Content (TEC)

Abstract

Accurate solar irradiance forecasting is critical for optimizing photovoltaic (PV) systems, enhancing grid stability, and enabling effective energy management. This study explores the integration of machine learning (ML) techniques with Global Positioning System (GPS) data to improve the accuracy of solar irradiance prediction. By incorporating Total Electron Content (TEC) and Integrated Water Vapor (IWV) derived from GPS data, alongside meteorological variables such as pressure and temperature, a robust forecasting model was developed. Among the three backpropagation algorithms tested—Levenberg-Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient—the Bayesian Regularization algorithm with a 10-layer neural network achieved the best performance, with the lowest Mean Square Error (MSE) and highest Correlation Coefficient (R). The model's predictions closely aligned with measured solar irradiance, demonstrating its reliability. Despite challenges such as data availability and computational complexity, the study highlights the potential of integrating GPS-derived data into ML-based solar irradiance forecasting. This approach offers a promising solution for advancing renewable energy management and supporting the transition to sustainable energy systems.

*Corresponding Author: yihhwa@utem.edu.my

I. INTRODUCTION

The increasing reliance on renewable energy sources, particularly solar energy, has necessitated the development of accurate forecasting models for solar irradiance. Solar irradiance forecasting is essential for optimizing the performance of photovoltaic (PV) systems, enhancing grid stability, and facilitating effective energy management [1]. As solar energy generation is inherently variable, driven by factors such as weather conditions and geographical location, precise forecasting becomes crucial for integrating solar power into existing energy infrastructures. This section discusses the significance of solar irradiance forecasting, the role of Machine Learning (ML) techniques, and the innovative incorporation of Global Positioning System (GPS) data in enhancing forecasting accuracy.

Solar irradiance, the power per unit area received from the Sun, is a key parameter influencing the efficiency and output of solar energy systems [2]. Accurate predictions of solar irradiance allow for better planning and management of energy resources, enabling utilities and operators to optimize the scheduling of energy generation and consumption, reduce reliance on fossil fuels by maximizing the use of renewable energy, enhance grid reliability by anticipating fluctuations in solar power production [3].

The variability of solar irradiance poses significant challenges for grid operators, particularly during peak demand periods or adverse weather conditions. Thus, effective forecasting models are imperative for mitigating these challenges and ensuring a stable energy supply.

Historically, solar irradiance forecasting has relied on physical models and statistical methods, which often struggle to capture the complex, non-linear relationships inherent in solar radiation data. These traditional methods typically use meteorological variables such as temperature, humidity, and cloud cover to predict solar irradiance. However, their performance can be limited, particularly in rapidly changing weather conditions [4 – 7].

In contrast, ML offers a robust alternative by leveraging large datasets to identify patterns and relationships that may not be evident through conventional methods. ML algorithms can learn from historical data and adapt to new information, making them particularly well-suited for dynamic environments like solar energy generation [8]. Recent studies have demonstrated that ML models, including Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and ensemble methods, can significantly improve forecasting accuracy compared to traditional approaches [9][10].

Recent literature has highlighted the growing interest in applying advanced machine learning techniques to solar irradiance forecasting. A notable trend is the use of deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which have shown promise in capturing complex patterns in large datasets. These models can process multiple input variables, including historical irradiance data, meteorological conditions, and GPS-derived information, to produce highly accurate forecasts.

A study by Almarzooqi et al. [11] introduced a hybrid ML model that combines various historical weather parameters with GPS data to improve the accuracy of solar irradiance predictions. The model demonstrated superior performance in forecasting under varying weather conditions, showcasing the potential of integrating diverse data sources.

Furthermore, research by Meenu et al. [12] explored the application of deep learning algorithms for very short-term solar irradiance forecasting. The authors emphasized the importance of spatial and temporal correlations in improving prediction accuracy, suggesting that incorporating GPS data could further enhance the effectiveness of their models.

Despite the advancements in ML-based solar irradiance forecasting, several challenges remain. One significant hurdle is the availability and quality of data. High-resolution, long-term datasets that combine solar irradiance measurements with GPS data are often scarce, limiting the training and validation of forecasting models. Additionally, the computational complexity of advanced ML algorithms can pose challenges in terms of resource requirements and scalability.

II. METHODOLOGY

A. Total Electron Content

Total Electron Content (TEC) refers to the total number of free electrons present in a column of the Earth's ionosphere, typically measured along the path of radio signals transmitted from satellites to ground receivers [13]. The existence of electrons influences radio wave propagation. TEC quantifies the electron density in the ionosphere, representing the cumulative effect of the electron concentrations across the ionospheric layer. It is commonly expressed in electrons per square meter, 1 TEC unit (1 TECU = 10^{16} electrons/m²) [14]. TEC can be divided into two parts: Slant TEC (sTEC) and Vertical TEC (vTEC). The equation for sTEC is as follows:

$$\mathrm{sTEC} = [9.483 \times (PR_{L2} - PR_{L1} - \Delta)] + \mathrm{CAL} \tag{1}$$

where:
PR_{L2} - L_2 pseudo-range in meters
PR_{L1} - L_1 pseudo-range in meters
Δ - Input bias between the C/A and P code chip transitions in meters
CAL - TEC result due to internal receiver L_1/L_2 delay and the offset

The sTEC stated in Equation 1 measures the total electron content of the ionosphere along the ray path from the satellite to the receiver and is measured at various elevation angles, as shown in Figure 1 [15].

Although sTEC is measured at differing elevation angles, vTEC is usually modelled and widely used. The representation of vTEC at an elevation angle of χ is depicted in Figure 1 and expressed as shown in Equation 2.

$$\mathrm{vTEC} = \mathrm{sTEC}(\cos\chi') \tag{2}$$

with

$$\cos\chi' = \sqrt{1 - \sin^2\chi'} \tag{3}$$

$$\sin\chi' = \frac{R_E}{R_E + h_m}\sin\chi \tag{4}$$

where:
χ and χ' - zenith angles at the receiver site and at the ionospheric pierce point, IPP
R_E - Mean Earth radius
h_m - Height of maximum electron density (450km)

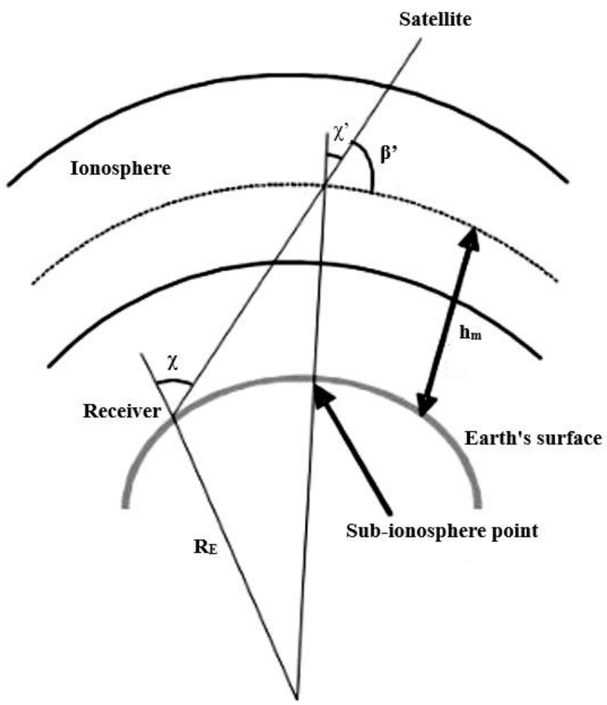


Figure 1. Ionospheric Single layer Model [15]

Seemala [16] developed a software application utilizing these equations to compute sTEC, vTEC and mean TEC. Mean TEC or average Total Electron Content, is often derived from multiple vertical Total Electron Content (vTEC).

B. Integrated water vapor (IWV)

Integrated Water Vapor (IWV) refers to the total amount of precipitable water in an atmospheric column between the Earth's surface and the top of the atmosphere [17]. It is expressed in units of kilogram per square meter (kg/m^2). The Zenith Total Delay (ZTD) can be derived from GPS data processing and converted IWV using the following formula:

$$\mathrm{IWV} = \frac{10^6}{R_v \cdot [k_2 + k_3/T_m]} \cdot (\mathrm{ZTD} - \mathrm{ZHD}) \tag{5}$$

where:
R_v - Gas constant for water vapor
k_2 and k_3 - Atmospheric refractivity constants [18]
T_m - Weighted mean temperature

Musa et al. [19] developed a regional model for T_m in Peninsular Malaysia, as shown in Equation 6:

$$T_m = 0.36T_s + 182.4 \tag{6}$$

where T_s represents the surface temperature.

As GPS signals travel from satellites to ground receivers, they experience a delay caused by the atmosphere. This total delay is referred to as ZTD. Within this delay, a portion known as the Zenith Hydrostatic Delay (ZHD) is attributed to dry gases like oxygen and nitrogen. The remaining part, known as the Zenith Wet Delay (ZWD), is due to water vapor. Equation 7 shows how ZWD can be computed by taking the difference between ZTD and ZHD.

$$\mathrm{ZWD} = \mathrm{ZTD} - \mathrm{ZHD} \tag{7}$$

ZHD can be calculated using local measured surface pressure. The formula for ZHD is given in Equation 8 [20]:

$$\mathrm{ZHD} = 2.2768\frac{\rho_s}{1 - 2.66 \times 10^{-3}\cos(2\varphi_s) - 2.8 \times 10^{-7}H_s} \tag{8}$$

where ρ_s is the pressure at the GPS station, φ_s is the latitude of the station, and H_s is the height of the station.

This delay can be mathematically represented by acquiring ZTD, which demonstrates a close proportionality to IWV. Therefore, ZTD can be computed using measured temperature and pressure.

In this study, TEC was obtained using historical GPS data from the *Fakulti Teknologi dan Kejuruteraan Elektronik dan Komputer (*FTKEK) GPS station (located at coordinates N 2.314100, E 102.318353). Meanwhile ZTD was derived from FTKEK weather station, located at the same coordinates. The dataset for this study comprised six months of data, specifically from February, March, April, August, September, and October 2022.

The input of the ML model included TEC, ZTD, pressure, and temperature to forecast solar irradiance. The dataset was divided such that 70% of the data formed the training set for model development, while the remaining 30% was divided evenly into two subsets: 15% for testing set and another 15% for validation to evaluate the model's performance.

Three algorithms (Levenberg-Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient) were used to achieve a good fit for the data. Mean Square Error (MSE) and Correlation Coefficient (R) were used to compare the performance of these algorithms. This validation process involved utilizing the processed dataset to plot a graph in MATLAB, comparing measured solar irradiance with predicted solar irradiance.

III. RESULTS AND DISCUSSION

The preprocessed dataset was trained using three backpropagation algorithms, each utilizing different layer sizes: 5, 8, and 10, as illustrated in Table 1.

Table 1
MSE And R Results for Different Algorithms and Layer Sizes

Algorithm	Layer Size	Mean Square Error (MSE)	Correlation Coefficient (R)
Levenberg-Marquardt	5	22635.0758	0.85149
	8	21671.6315	0.85816
	10	20894.1821	0.86083
Bayesian Regularization	5	21740.4728	0.85549
	8	20899.5107	0.86136
	10	20882.4233	0.86138
Scaled Conjugate Gradient	5	22413.4795	0.84889
	8	25279.5041	0.83225
	10	24324.8056	0.83575

During training, MSE was used to evaluate the correlation of the neural network with the training data. A lower MSE indicates better model performance during training. Additionally, the correlation coefficient (R) was used to verify how well the predicted values from the neural network model align with the actual target values. The correlation coefficient, ranging from -1 to 1, measures the closeness of the relationship between variables. An R value closer to 1 indicates a stronger association between the variables, implying better model performance. Table 1 shows that the Bayesian Regularization algorithm with a layer size of 10 is the best model, achieving the lowest MSE and the highest R value among the tested algorithms.

A. Mode Training

Figure 2 shows the neural network training state, which refers to the condition or status of the neural network during the training process. The gradient represents the slope of the tangent to the graph of the function, indicating the direction where the function exhibits a notable increase. The parameter '*mu*' served as a governing factor in the back-propagation neural network, directly influencing error convergence. Validation checks were employed to halt the neural network's learning process. The count of validation checks corresponded to the number of consecutive iterations of the neural network.

In Figure 4, the dotted line shows the best training performance plot (MSE plot), achieving a value of 20882.4233 at epoch 144. This value represents the average of the squared errors or deviations.

The histogram in Figure 5 illustrates the error distribution during model training. The distribution exhibits a symmetric shape, with a significant concentration of data points centered around zero error. Both the mean and median values are close to zero, implying a central tendency of the data around this point. This suggests that a substantial proportion of the dataset records errors near zero, potentially indicating accurate predictions or measurements for the given phenomenon.

Figure 6 presents the regression plot of the training model, showing the relationship between the measurement solar irradiance(T) and the predicted solar irradiance (Y). A strong positive correlation is observed in the regression plot. However, as the plot extends to the right, a deviation occurs where the regression line falls below the ideal linear regression line Y=T.

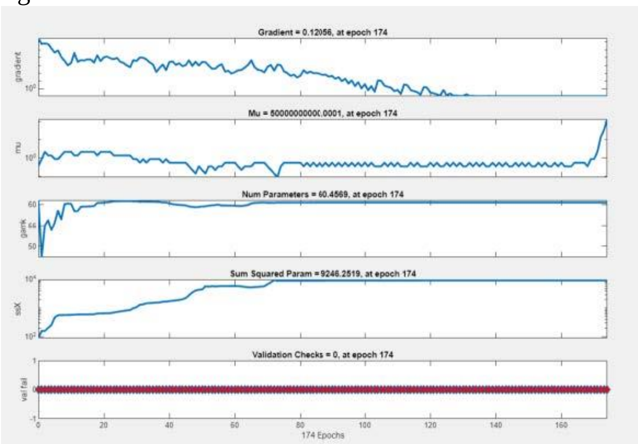


Figure 2 Training State Plot

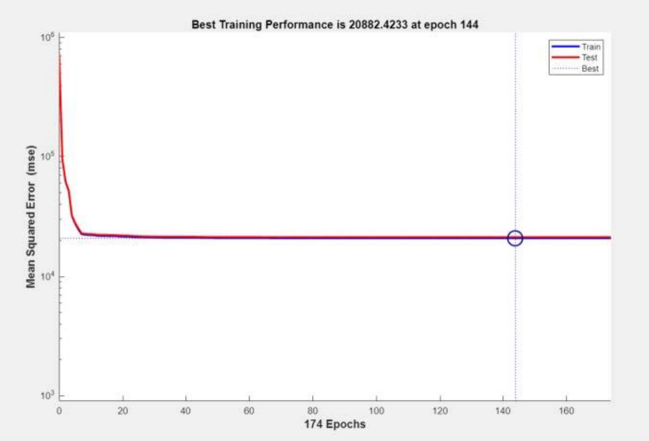


Figure 3 MSE Plot

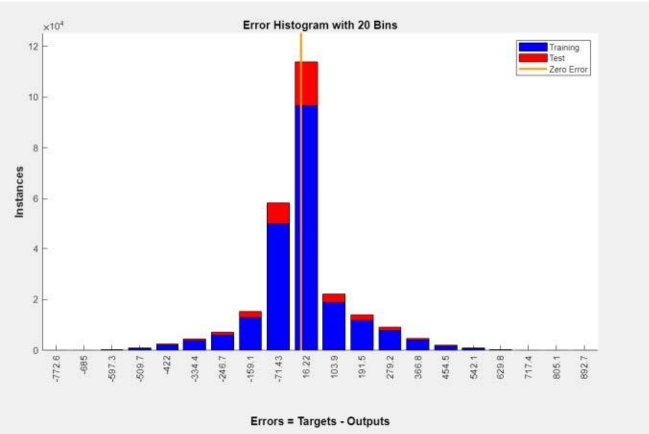


Figure 4 MSE Plot

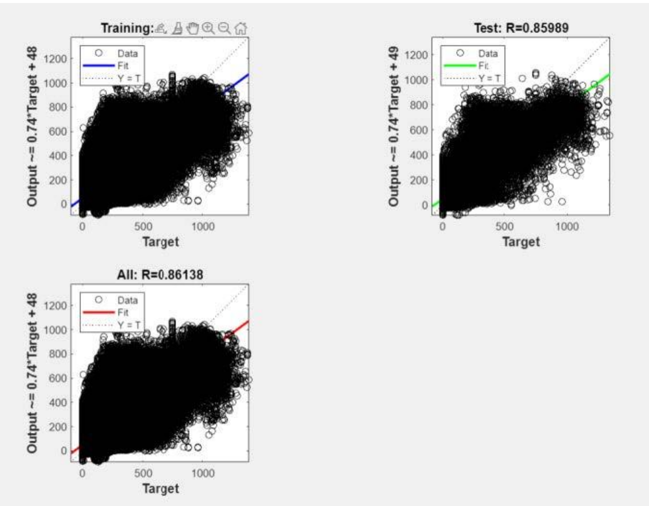
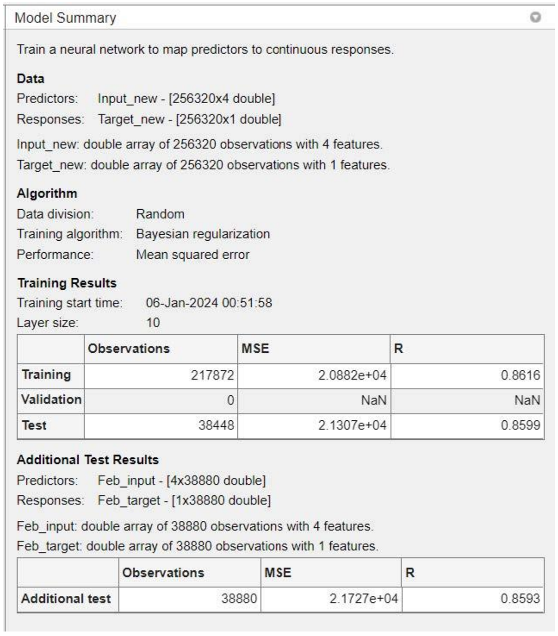


Figure 5 Error Histogram Plot

B. Model Testing

After the training process, one month of data from February was used for model testing. Figure 6 displays the model summary, providing a comprehensive overview and essential information about the trained neural network model, including input data, training algorithm, training results, and additional test results. Predictors represent the input data, while responses refer to the output (solar irradiance data).

Figure 7 presents the February error histogram plot, which exhibits a symmetric shape with a significant concentration of data points centered around zero error. Meanwhile, Figure 8 shows the regression plot of the additional testing. A moderate to weak positive correlation is observed in the regression line. However, as the plot extends to the right, the regression line falls below the ideal linear regression line Y=T.


Model Summary

Train a neural network to map predictors to continuous responses.

Data

Predictors: Input_new - [256320x4 double]

Responses: Target_new - [256320x1 double]

Input_new: double array of 256320 observations with 4 features.

Target_new: double array of 256320 observations with 1 features.

Algorithm

Data division: Random

Training algorithm: Bayesian regularization

Performance: Mean squared error

Training Results

Training start time: 06-Jan-2024 00:51:58

Layer size: 10

	Observations	MSE	R
Training	217872	2.0882e+04	0.8616
Validation	0	NaN	NaN
Test	38448	2.1307e+04	0.8599

Additional Test Results

Predictors: Feb_input - [4x38880 double]

Responses: Feb_target - [1x38880 double]

Feb_input: double array of 38880 observations with 4 features.

Feb_target: double array of 38880 observations with 1 features.

	Observations	MSE	R
Additional test	38880	2.1727e+04	0.8593

Figure 6 The model summary

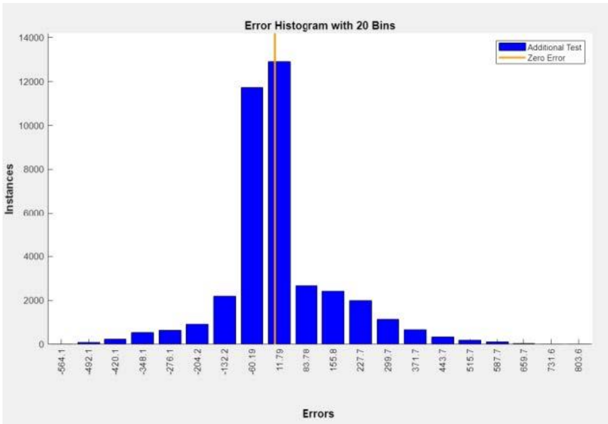


Figure 7 Additional testing Error Histogram Plot

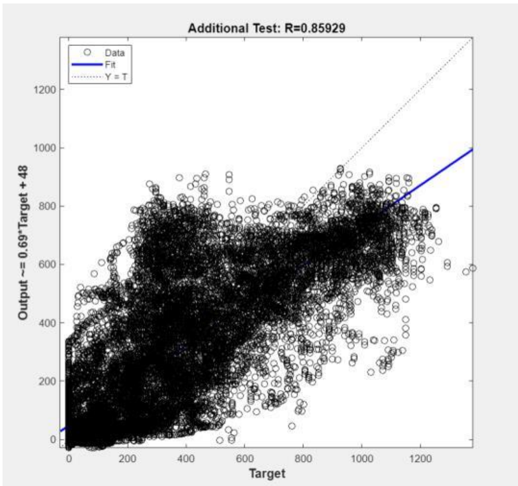


Figure 8 Regression Plot for additional test 1 month data (February)

C. Comparison between Measurement and Predicted Solar Irradiance

To assess the performance of the solar irradiance forecasting model, datasets from a single day (February 1,

2022) and the entire month of February were employed to visualize both the measured and predicted solar irradiance values. Figure 9 shows the comparison between measured and predicted solar irradiance for a single day (February 1, 2022).

Figure 10 shows the comparison between measured and predicted solar irradiance for the entire month of February. This comparison shows that both the measured and predicted data follow nearly identical graph shapes, indicating that the model's output closely matches the actual solar irradiance data. Moreover, the error histogram below the waveform plot shows the data difference between measured and predicted solar irradiance values. The histogram illustrates a symmetric shape with a substantial concentration of data points centered around zero, indicating a favorable result. This suggests a close alignment between the measured and predicted solar irradiance values, demonstrating the model's accuracy.

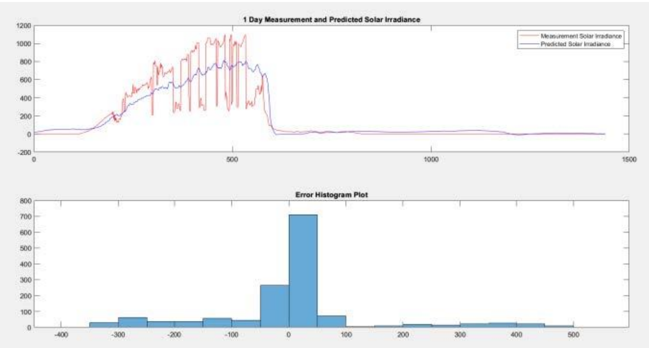


Figure 9 Comparison between measurement and predicted solar irradiance plot for 1 day data

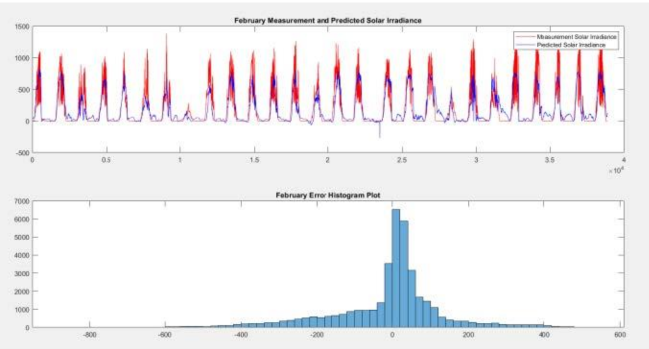


Figure 10 Comparison between measurement and predicted solar irradiance plot for one month data (February).

IV. CONCLUSION

The integration of ML techniques, coupled with the innovative use of GPS data, has shown promising results in enhancing forecasting accuracy. By incorporating TEC and IWV derived from GPS data, alongside meteorological variables such as pressure and temperature, this study successfully developed a robust model for predicting solar irradiance.

Among the three backpropagation algorithms tested—Levenberg-Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient—the Bayesian Regularization algorithm with a 10-layer neural network structure demonstrated the best performance, achieving the lowest Mean Square Error (MSE) and the highest Correlation Coefficient (R). The results suggest that this approach is particularly effective in capturing complex, non-linear relationships within solar irradiance data.

Moreover, the model's validation and testing phases further confirmed its reliability, with the predicted solar irradiance closely aligning with the measured values. This close alignment, as evidenced by the error histograms and regression plots, underscores the model's potential for practical application in real-world scenarios.

ACKNOWLEDGMENT

The authors wish to acknowledge the Universiti Teknikal Malaysia Melaka (UTeM) for providing the research platforms and facilities.

CONFLICT OF INTEREST

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

AUTHOR CONTRIBUTION

The authors confirm contribution to the paper as follows: Ho Y.H. led the project design, conceptualized the research methodology, and supervised the integration of GPS data with ML algorithms. Tang S.Y. contributed to the data acquisition and preprocessing, handled meteorological data integration and assisted with model validation. Guo J.Q. managed the model's technical implementation, focusing on coding, training, and testing the neural networks in MATLAB. All authors participated in analyzing the results and contributed to writing, reviewing, and revising the manuscript for publication.

REFERENCES

[1] C. Voyant, G. Notton, S. Kalogirou, M.L. Nivet, C. Paoli, F. Motte and A. Fouilloy, "Machine learning methods for solar radiation forecasting: a review," Renewable Energy, vol. 105, 2017, https://doi:10.1016/j.renene.2016.12.095.

[2] P. Kumari, and D. Toshniwal, "Deep learning models for solar irradiance forecasting: a comprehensive review," Journal of Cleaner Production, vol. 318, 2021, https://doi:10.1016/J.JCLEPRO.2021.128566.

[3] N. Michael, M. Mishra, S. Hasan, and A. Al-Durra, "Short-term solar power predicting model based on multi-step CNN stacked LSTM technique," Energies, 2022, https://doi:10.3390/en15062150.

[4] H. Verbois, R. Huva, A. Rusydi and W. Walsh, "Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning", Solar Energy, vol. 162, pp. 265-277, 2018, https://doi:10.1016/J.SOLENER.2018.01.007.

[5] J. Thaker and R. Höller, "A comparative study of time series forecasting of solar energy based on irradiance classification," Energies, 2022, https://doi:10.3390/en15082837.

[6] H. Jeon, M. Choi, and O. Lee, "Day-ahead hourly solar irradiance forecasting based on multi-attributed spatio-temporal graph convolutional network," Sensors, 2022, https://doi:10.3390/s22197179.

[7] A. Razagui, K. Abdeladim, K. Bouchouicha, N. Bachari, S. Semaoui and A. Arab, "A new approach to forecast solar irradiances using WRF and libRadtran models, validated with MERRA-2 reanalysis data and pyranometer measures," Solar Energy, vol. 221, pp. 148-161, 2021, https://doi:10.1016/J.SOLENER.2021.04.024.

[8] S. Tajjour, S. S. Chandel, M. A. Alotaibi, H. Malik, F. P. García Márquez and A. Afthanorhan, "Short-term solar irradiance forecasting using deep learning techniques: a comprehensive case study," IEEE Access, vol. 11, pp. 119851-119861, 2023, https://doi: 10.1109/ACCESS.2023.3325292.

[9] M. B. U. Shahin, A. Sarkar, T. Sabrina, and S. Roy, "Forecasting solar irradiance using machine learning," in Proc. of 2020 2nd International

Conference on Sustainable Technologies for Industry 4.0 (STI), 2020, pp. 1–6, https://doi: 10.1109/STI50764.2020.9350400.
[10] A. M. Assaf, H. Haron, H. N. A. Hamed, F. A. Ghaleb, S.N. Qasem and A.M. Albarrak, "A review on neural network based models for short term solar irradiance forecasting," Appl. Sci., vol. 13, 2023, https://doi.org/10.3390/app13148332.
[11] A. M. Almarzooqi, M. Maalouf, T. H. M. El-Fouly, V.E. Katzourakis, M. S. El Moursi and C. V. Chrysikopoulos, "A hybrid machine-learning model for solar irradiance forecasting," Clean Energy, vol. 8, no. 1, pp. 100–110, 2024, https://doi.org/10.1093/ce/zkad075.
[12] A. Meenu, M. Martínez-Ramón, and Manel, "Deep learning algorithms for very short-term solar irradiance forecasting: a survey," Renewable Energy, vol. 123, pp. 2198-2210, 2023, https://doi.org/10.1016/j.renene.2023.01.012.
[13] E. D. L. Izurieta, E. T. Guamanarca, and H. Barbier, "Ionospheric total electron content (TEC) above ecuador," Journal of Physics: Conference Series, vol. 2238, no. 1, 2022, https://doi:10.1088/1742-6596/2238/1/012010.
[14] B. Rana et al., "Study of total electron content and electron density profile from satellite observations during geomagnetic storms", JNPS, vol. 5, no. 1, pp. 59-66, 2019.
[15] Abdullah, "TEC and scintillation study of equatorial ionosphere: a month campaign over Sipitang and Parit Raja stations, Malaysia," American Journal of Engineering and Applied Sciences, vol. 2, no. 1, pp. 44–49, 2009, https://doi:10.3844/ajeas.2009.44.49
[16] G. Seemala, "GPS-Tec Analysis Software," Accessed: Jan. 9, 2024, [Online] Available: https://www.academia.edu/23913567/GPS_TEC_analysis_software.
[17] E. K. Makama and H. S. Lim, "Variability and trend in integrated water vapour from era-interim and IGRA2 observations over Peninsular Malaysia," Atmosphere, vol. 11, no. 9, 2020, https://doi:10.3390/atmos11091012.
[18] P. Yuan et al., "Characterisations of Europe's integrated water vapour and assessments of atmospheric reanalyses using more than 2 decades of ground-based GPS," Atmospheric Chemistry and Physics, vol. 23, no. 6, pp. 3517–3541, 2023, https://doi:10.5194/acp-23-3517-2023.
[19] T. A. Musa, M. H. Mazlan, Y. D. Opaluwa, I. A. Musliman, and Z. M. Radzi, "Water vapour weighted mean temperature model for GPS-derived integrated water vapour in Peninsular Malaysia," Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., pp. 127–135, 2017, https://doi.org/10.5194/isprs-archives-XLII-4-W5-127-2017.
[20] J. Morland and C. Mätzler, "Spatial interpolation of GPS integrated water vapour measurements made in the Swiss alps," Meteorological Applications, vol. 14, no. 1, pp. 15–26, 2007, https://doi:10.1002/met.2Z.