Modified Support Vector Regression Model For Very Short-Term Solar Irradiance Forecasting

Thirunavukkarasu, Gokul Sidarth, Seyedmahmoudian, Mehdi, Jamei, Elmira, Horan, Ben, Mekhilef, Saad and Stojcevski, Alex (2022) Modified Support Vector Regression Model For Very Short-Term Solar Irradiance Forecasting. Webology, 19 (3). pp. 2027-2038. ISSN 1735-188X

Webology (ISSN: 1735-188X)
Volume 19, Number 3, 2022

Modified Support Vector Regression Model For Very Short-Term Solar Irradiance Forecasting

[1]Gokul Sidarth Thirunavukkarasu* , [1]Mehdi Seyedmahmoudian , [2]Elmira Jamei , [3]Ben Horan , [1]Saad Mekhilef , and [1]Alex Stojcevski

[1] School of Science, Computing and Engineering Technologies, Swinburne University of Technology, VIC 3122, Australia.

[2]College of Engineering and Science, Victoria University, VIC 3011, Australia.

[3]School of Engineering, Deakin University, Waurnponds, VIC 3216, Australia.

Abstract- One of the most reliable and prominent renewable energy resource which addresses the global energy demand is the Solar photovoltaics (PV) systems. Usage of the PV system substantially reduces the carbon footprint of the energy generation process leading towards an environmentally friendly alternative. The substantial rise in the adaptation of the distributed energy resources likes solar PV and batteries into conventional electric power grids, has increased the complexity of the energy management problem. Many Researchers across the world are working towards developing accurate forecasting models to predict the solar PV system's generation capacity in advance to effectively manage the supply and demand. In this research, a modified support vector regression (SVR) based solar irradiance forecasting model with a minute wise time horizon is proposed, developed, and tested using the data obtained from Bureau of Meteorology (BOM), Australia. The performance of the proposed approach is then validated using the benchmark statistical models like, the persistence algorithm, autoregression model, moving average model, a hybrid autoregressive moving average model and an autoregressive integrated moving average model. Metrics like root mean square error (RMSE), the mean absolute error (MAE) and mean bias error (MBE) are used to validate the performance of the proposed approach. Results indicated that the Modified SVR model reduces the forecasting error and the computational complexity substantially and outperforms the other conventional approaches. The increased performance of the forecasting method will assist in developing an efficient energy management system for future electricity grids.

Index terms: Machine learning, PV power forecasting, Renewable energy resources, Support vector Regression.

I. INTRODUCTION

Rapidly increasing population and exponentially growing energy demand has substantially increased the need for an environmentally sustainable alternative source of energy. The usage of renewable energy sources (RESs) has significantly addressed this problem in recent years [1, 2]. Many pieces of evidence in the literature have been identified that clearly depicts the renewable energy integration into smart power grids has reduced the carbon footprints of the energy generation [3, 4]. The solar PV energy systems are integrated into the conventional energy grids in the form of large-scale concentrated solar power plants (CSP) and the small-scale distributed generation (DG) units in the form of rooftop solar PV systems. In general, the solar PV systems convert the irradiance from the sun directly into electric power. The output of the solar PV system is intermittent and stochastic in nature because of the direct dependency of the output power with the random environmental conditions. The CSP's are considered as a potential replacement for the power plants and many investors are moving towards the move of implementing a solar farm. Especially in the context of Australia, the concept of solar farms acting as a node of the virtual power plant is accounted to be a reliable solution to address the energy demand in a more sustainable manner. The amount of the energy generated by the solar PV system is influenced by many external factors like partial shading conditions (PSC) that are created by the moving and transient clouds, shadows from neighboring buildings and tress. Many artificial intelligent and machine learning based techniques are used to address this issue of PSC by identifying the maximum power point tracking (MPPT) [5]. The high penetration of the solar PV systems in the electricity grids introduces many technical issues like reverse power flow, stability issues and voltage regulation problems into conventional power systems [6]. The stochastic natures of the renewable power systems create fluctuations in the stability of the power grid and introduce many problems associated with energy storage systems [7] and energy management systems [8]. To guarantee safe operation of the electric power grids with higher penetration of solar PV systems, accurate forecasting of the output power of the system is considered essential. Quite a few research works are identified in the literature to evaluate the performance of the solar forecasting algorithm using the direct horizontal irradiance (DHI) value which plays a vital role in accurately forecasting the output power of the large-scale PV systems which include solar farms. Support vector regression models have been one of the most commonly used technique that is addressing various applications which might include forecasting [9], demand response [10], control system [11], single-phase converter control [12], national language processing [13] and many other diversified applications. With the increase in the adaptation of large-scale PV systems in the form of the concentrated solar PV systems into the electric power systems, it is essential to have an accurate forecasting technique that predicts the output of the CSP's based on the forecasted direct nature irradiance (DNI) value. In this research, a modified support vector regression based solar irradiance forecasting model with a minute wise time horizon is proposed. Initially, persistence algorithm, auto regression model, moving average model, a hybrid auto regressive moving average model and an auto regressive integrated moving average model are considered as the base line approach for comparing and analyzing the performance of the proposed technique. The benchmark techniques are evaluated using the root-mean-squared error (RMSE), the mean absolute error (MAE) and mean bias error (MBE). The rest of the paper is organized as follows: Section 2 gives an overview of the different techniques used for forecasting the solar irradiance (DHI or DNI), after which in Section 3 the brief overview of the data set is presented. Later in Section 4 and 5 the proposed and the benchmark techniques are explained in detail. Finally, in Section 6 the results and discussion are presented and then the paper is concluded.

II. BACKGROUND

Accurate solar irradiance forecasting is an essential parameter for predicting the total energy generated from the solar PV systems such that the demand and supply curve is neutralized for an energy efficient system. Based on the application of the forecasted values the time horizon of the forecast varies from seconds to months. The forecasting is therefore categorized into four major categories, and they are: (a) long-term forecasting, (b) medium-term forecasting, (c) short-term forecasting and (d) very-short-term forecasting. Solar irradiance being very intermittent and stochastic in nature it needs an efficient forecasting technique to obtain

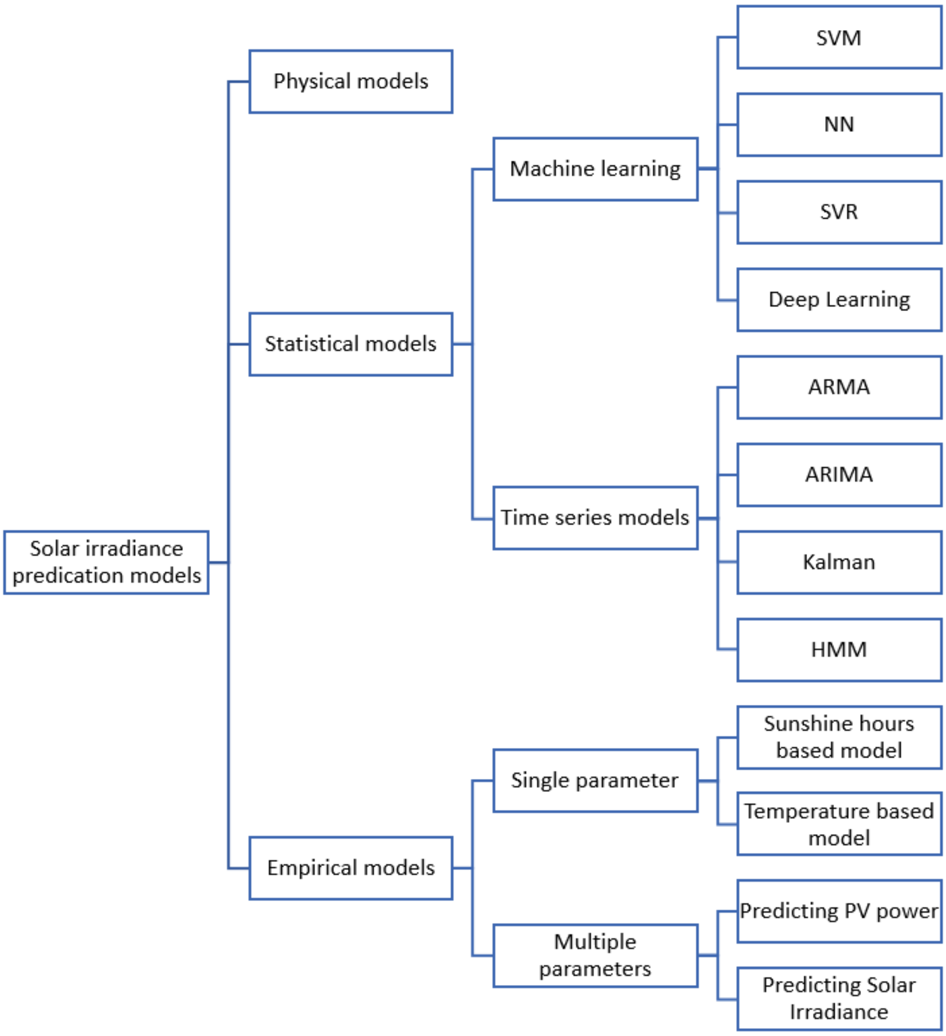

Figure 1 Different Forecasting techniques used in the literature

accurate predictions. This is a key concern for researchers in the field of renewable energy as this invariably contributes to the cost and efficiency of smart grids. Figure 1 illustrates the different algorithm used for forecasting of DNI in the literature. From the critical review conducted on the existing literature, it is evident that the many physical, statistical, and artificial intelligence methods exist [14]. The use of machine learning techniques hasn't been explored to the fullest potential and therefore in this research, we are aiming at exploring the performance of an SVR model using the time series forecasting technique.

Various statistical methods like persistence model [15], probabilistic models [16], auto-regressive moving average (ARMA), auto-regressive integrated moving average (ARIMA) and auto-regressive moving average model with exogenous inputs (ARIMAX) [17, 18], were highlighted in the literature based on the forecasting of accurate solar power output based on the historical time series analysis of the irradiance data. Followed by the statistical methods the artificial intelligence techniques like artificial neural networks [19], metaheuristic algorithms [20], fuzzy theory, Bayesian neural networks, hybrid wavelet analysis and ANN method and wavelet recurrent neural networks where identified and analyzed for getting an idea on how the time series weather data prediction has been used for forecasting the output power of the PV systems. Figure 1 shows the distribution of the techniques used for forecasting the solar PV output.

Figure 2 Location of the data

III. DATA USED IN THE STUDY

The minute wise solar irradiance data collected using the open-source solar statistics available from the Bureau of Meteorology (BOM), Australia is used in the study. BOM records the solar data in 29 locations in Australia and for this study, the data from the station situated at Melbourne Airport, VIC (37.69°S, 144.84°E) was used. The minute wise data recorded by BOM is not a real-time product. Even though the values are recorded minute wise, they are processed every 3 months to identify any abnormalities with the data and correct the long-term drifts appropriately. In this research, we use the data obtained from the 1st of January 2008 to 31st of March 2019. Various data preprocessing steps have been carried out to compile all the individual monthly DNI stats into a single file and then values without an entry are removed from the data set after it has been tagged with the corresponding date time. Figure 2 shows the location of the weather station, and it is observed that there is a lot of abnormalities in the data set and missing data points, so it is important to tag them based on time and use the data for training the forecasting model with a different time lag. The distribution of the data set with respect to days in a month, hours in a day and years considered is clearly shown in Figure 3. With respect to the distribution in terms of the minutes, we are not considering the time slots with complete minimal irradiance to have a better visualization of the data. Figure 4 illustrates the autocorrelation and the partial correlation outputs obtained from the considered dataset. It is clear from the data that there is a lot of abnormalities and some missing data which is replaced by an average value of the dataset.

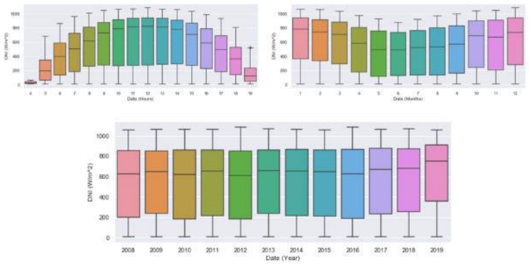

Figure 3 Distribution of the DNI data set obtained from BOM

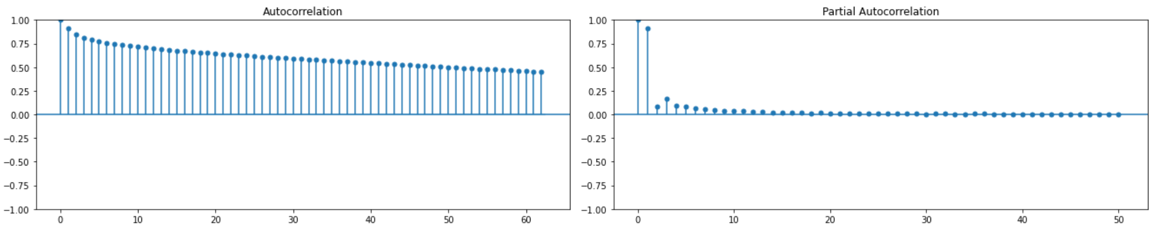

Figure 4 Correlation and Partial correlation of the dataset

IV. METHODOLOGY

In this research article, we present a support vector regression-based forecasting algorithm and the following methodology elaborated in Figure 5 is carried out. Initially, the data extraction from BOM was carried out. The datasets available was for individual months and these individual monthly datasets were extracted and the dataset also consisted of other weather parameters which were not considered in this research, so we had to develop methodology to combine the monthly dataset into a large 10-year minute wise dataset. Once the final dataset was obtained, we had to clean the data to figure out outliers and missing data points and replace it with a mean value. The obtained data set was then evaluated with the base line approaches like persistence model, AR, MA ARMA, and ARIMA models. More detailed explanation on the individual techniques is expressed in the next section, in this section we will have a detailed overview of the support vector regression model used for forecasting.

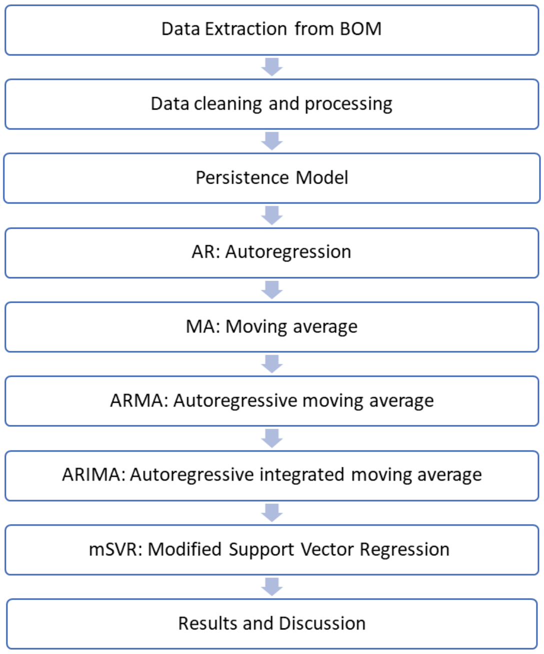

Figure 5 Methodology used for the development of the Modified support vector regression-based Forecasting algorithm

A. Proposed Modified Support vector regression model

Vapnik et. al. proposed the support vector machines in 1999. [21-24] and the concept of the support vector regression to correlate the nonlinearity of the original data x into a higher dimensional feature space as expressed in Figure 6 was later developed. The SVR models are used commonly in the classification problem and the forecasting problem. In the SVR model we consider a data set $G = (x_i, d_i)_{i-1}^{N}$ (where, the x_i is the input vector and d_i is the actual value of the parameter that you are classifying or forecasting, and the N is the total size of the dataset. In general, the SVR function is expressed as indicated in the following

$$y = f(x) = w\psi(x) + b \qquad (1)$$

The y is the output and the $\psi(x)$ is the function that maps the nonlinearity the dataset and identifies a relationship in the search space in relation to x which is the input considered for the model. The w and b are coefficients extracted from the process of the minimization of the risk function which is defined as follows:

$$R(C) = (C/N)\sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \parallel w \parallel^2/2 \qquad (2)$$

$$L_\varepsilon(d, y) = \begin{cases} 0 & |d-y| \leqslant \varepsilon \\ |d-y| - \varepsilon & \text{otherwise} \end{cases} \qquad (3)$$

The risk function considered for estimating the coefficients help us understand the relationship with the hyperplanes in the support vector regression approach as expressed in the figure 6. The loss function $L_\varepsilon(d, y)$ is ε-insensitive the flatness function $\parallel w \parallel^2/2$ in combination with the loss function formulates the risk functions. The term C is a specific tradeoff between the empirical risk function and the user defined parameters which are based on the slack that is added based on the actual values considered in the dataset and the corresponding boundary constraints.

$$R(w, \zeta, \zeta^*) = \| w \|^2/2 + C\left(\sum_{i=1}^{N} (\zeta_i + \zeta_i^*)\right) \quad (4)$$

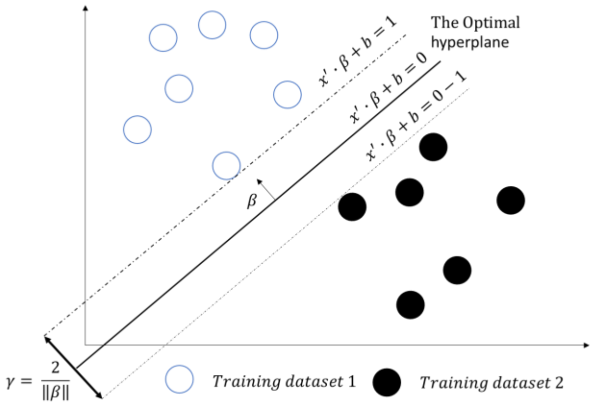

Figure 6 Illustration of the SVR model

The modified SVR model developed as the part of the research presented in this manuscript uses the radial bias function as a kernel and the expression of the Radial bias function kernel is expressed as follows and this assist in the approximation of the predicted values using the proposed algorithm.

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right) \quad (5)$$

Therefore, the modified support vector regression model developed using python libraries are used to evaluate the prediction using the dataset expressed in the previous section. The following sections explain the other benchmarking models considered in the study and in discussion section following that a detailed overview of the results observed from the study is presented.

V. BENCHMARKING MODELS

As a benchmark for evaluating the performance of the proposed mSVR based forecasting model, we considered five baseline or benchmark models, which are explained in detail in the following section.

A. Persistence Model

The most used benchmarking or baseline technique in the literature to evaluate the performance of the forecasting models is the persistence model. In this model, the minute wise granular irradiance value from the previous day $y_{(d-1,t)}$ is considered as the predicted value of the time t in day d.

$$\hat{y}_{d,t} = y_{d-1,t} \quad (6)$$

The persistence model is a simple, computationally less expensive but less accurate forecasting model that is free from training procedure and features like parameter setting. Equation 6 illustrates the mathematical equation governing the persistence model.

B. Autoregression (AR)

An autoregression model is used to predict the output of the system by identifying a linear combination of the input values. The prediction made by the AR model is mathematically represented as

$$\hat{y} = b_0 + b_1 * y_{t-1} + b_2 * y_{t-2} + \cdots + b_n * y_{t-n} \quad (7)$$

In equation (7), $\hat{y}$is the predicted value and $b_0, b_1, .., b_n$ are the coefficients of the input variables which are obtained by training the model. y is the input/historical values which will be used to predict the forecasted output. AR model is like the linear regression model but since it used the data from the same input for every previous time step. The AR model that is used as a baseline in this research, predicts the next day ahead solar irradiance value with a granularity of per minute based on the observations made over the training period.

C. Moving average (MA)

The moving average model uses the history of forecasted errors in a regression like the model to calculate the white noise e_t. Later, after identifying the white noise the forecasted value $\hat{y}$ is derived from the following expression.

$$\hat{y} = \theta_0 + \theta_1 * c_{t-1} + \theta_2 * c_{t-2} + \cdots + \theta_n * c_{t-n} \quad (8)$$

In the moving average model, we really do not literally observe the white noise, but instead, during the training phase, the model learns the pattern of error and uses it to predict the following time cycle.

D. Autoregressive moving average (ARMA)

The combination of the AR and the MA model results in the autoregressive moving average model. This is considered as one of the other baseline models which have been evaluated in this study.

$$\hat{y} = b_0 + b_1 * y_{t-1} + b_2 * y_{t-2} + \cdots + b_n * y_{t-n} + \theta_0 + \theta_1 * e_{t-1} + \theta_2 * e_{t-2} + \cdots + \theta_n * e_{t-n} \quad (9)$$

E. Autoregressive integrated moving average (ARIMA)

The autoregressive integrated moving average (ARIMA) model was introduced by Box and Jenkins [17, 18]. At the first stage of using the ARIMA forecasting model we need to find the stationarity of the dataset and is it is not stationary differentiation is taken over the data till it achieves the stationarity. The ARIMA model consists of three parameters (p, d, q). The number of times the differentiation is carried out it denoted by the integration factor d in the ARIMA model. Followed by the p and q that denotes the AR parameter or the lag of the system and the Moving average parameter which will be the window in which the moving average is performed.

The identified benchmarking algorithms and the proposed mSVR model is evaluated with the data set by splitting the data from 2008 to 2018(10 Year) as the training dataset and consider the prediction for the 1st of Jan 2019. The results explained in the following section are based on this consideration and the predictions for the 1st of Jan 2019 are presented in the study and the prediction for the entire month is carried out before analyzing the identified metrics expressed in the section below.

VI. RESULTS AND DISCUSSION

Three error metrics were used to evaluate the performance of the five different benchmarking algorithms along with the proposed modified SVR model. The results of the experimental trials are given in Table 1. Furthermore, along with the RMSE, the mean absolute error (MAE) and mean bias error (MBE) are considered for evaluating the performance of the proposed benchmark algorithms in addition to the RMSE. The RMSE, MAE and MBE are calculated using the following equations 10, 11 and 12.

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{t=1}^{\mathrm{N}}(\hat{y} - \mathrm{X})^2} \qquad (10)$$

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum_{t=1}^{\mathrm{N}}\mathrm{abs}(\hat{y} - \mathrm{X}) \qquad (11)$$

$$\mathrm{MBE} = \frac{1}{\mathrm{N}}\sum_{t=1}^{\mathrm{N}}\mathrm{X} - \hat{y} \qquad (12)$$

where, $\hat{y}$ is the forecasted value and X is the original value and N is the total number of samples forecasted. From the results, it was evident that the benchmarking agent had a lot of discrepancy and scope of improvement in terms of predicting the value accurately. Addressing these limitations, the proposed mSVR based approach is evaluated with the same dataset. Figure 7 shows the predicted and actual value comparison of the mSVR model. For the entire day ahead predictions (zoomed to fit in the time from 4:00pm to 6:00pm on Jan 1st, 2019, fig (7a) and for a specific 15-minute interval in fig (7b). The RMSE value of 66.763 is obtained and the results indicated that there is a substantial increase in the accuracy of the predications. In general, for the forecasting task, the nighttime data is ignored as this consideration increases the stochastic nature of the data set, but in this study, we are considering the zeros and the nighttime data to increase the complexity of the dataset. It is evident from the results indicated in Table 1, that the benchmarking algorithms like AR and MA are struggling to match the basic persistent algorithm in terms of the forecasting the data for the given dataset. With the increased randomness in the dataset, it was essential to test the performance of the computationally less expensive algorithm like the mSVR and evaluate the performance of the algorithm. Results indicated that the proposed mSVR based forecasting model performs about 25% better than the persistent model and outperforms the other statistical approaches. It is also noted that accuracy of the hybrid algorithms like ARMA and ARIMA are close to the proposed mSVR algorithms,

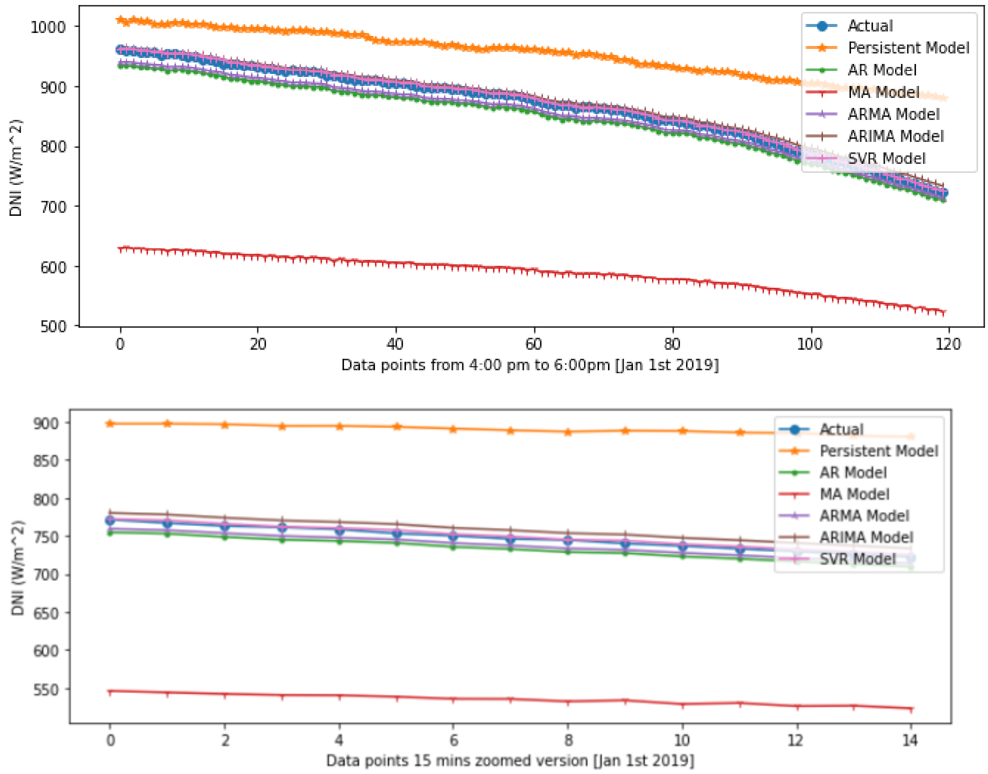

Figure 7 Comparison of the data forecast using the proposed SVR model and the benchmarking

but the proposed algorithm outperforms the other techniques in terms of the computational requirements.

Table 1 Comparison of the benchmark methods

Method	RMSE	MAE	MBE
Persistence	281.711	153.078	40.500
AR	76.953	35.954	22.424
MA	329.172	317.562	269.369
ARMA	72.220	31.251	17.767
ARIMA	68.979	16.891	1.760
SVR	66.763	14.938	2.280

VII. CONCLUSION

A time series forecasting model using a modified SVR approach to predict a day ahead minute wise DNI irradiance value was proposed in this manuscript. SVR model is robust and performs critically better with datasets that consist of more outliers. It's generalization capability with high predication accuracy makes it a more efficient algorithm in comparison to the other advanced machine learning approaches. The less complex nature of the SVR model also attributes to the low computational time required for predicting a day ahead day at a minute wise accuracy. Results indicated in the table 1 illustrates the performance of the proposed SVR model. In this paper, we have proposed a modified SVR model using a radial bias kernel based DNI forecasting technique with a very short-term time horizon, and this will be used to predict the output generated from the CSP by knowing the capacity of the CSP considered. The results indicate that the proposed mSVR model performs better than the baseline models compared in the research. The proposed model has a very small forecasting error when compared to the other benchmarking models considered and this accurate forecasting will help in accurate energy management and planning for addressing the energy demand effectively. In addition to the increase accuracy the computation requirements of the proposed approach are very minimal making it a better choice. Future work of the team will be to evaluate the performance of the proposed SVR models with some advanced machine learning models like recurrent neural networks and develop a hybrid algorithm with increased accuracy.

ACKNOWLEDGEMET

The authors would like to extend their sincere appreciation to the DCH5 AIRAH - ARENA grant and SUPRA for supporting the research.

REFERENCES

[1] O. Ellabban, H. Abu-Rub, and F. Blaabjerg," Renewable energy resources: Current status, future prospects and their enabling technology," Renewable and Sustainable Energy Reviews, vol. 39, pp. 748-764, 2014.
[2] N. Mahmud and A. Zahedi," Review of control strategies for voltage regulation of the smart distribution network with high penetration of renewable distributed generation," Renewable and Sustainable Energy Reviews, vol. 64, pp. 582-595, 2016.
[3] D. Connolly, H. Lund, B. V. Mathiesen, and M. Leahy," The first step towards a 100% renewable energy-system for Ireland," Applied Energy, vol. 88, no. 2, pp. 502-507, 2011.
[4] B. V. Mathiesen et al.," Smart Energy Systems for coherent 100% renewable energy and transport solutions," Applied Energy, vol. 145, pp. 139-154, 2015.
[5] M. Seyedmahmoudian et al.," Maximum power point tracking for photovoltaic systems under partial shading conditions using bat algorithm," Sustainability, vol. 10, no. 5, p. 1347, 2018.
[6] Y. S. Lim and J. H. Tang, "Experimental study on flicker emissions by photovoltaic systems on highly cloudy region: A case study in Malaysia," Renewable Energy, vol. 64, pp. 61-70, 2014.
[7] A. Woyte, V. Van Thong, R. Belmans, and J. Nijs," Voltage fluctuations on distribution level introduced by photovoltaic systems," IEEE Transactions on energy conversion, vol. 21, no. 1, pp. 202-209, 2006.
[8] S. Shivashankar, S. Mekhilef, H. Mokhlis, and M. Karimi," Mitigating methods of power fluctuation of photovoltaic (PV) sources–A review," Renewable and Sustainable Energy Reviews, vol. 59, pp. 1170-1184, 2016.

[9] M. Abdel-Nasser and K. Mahmoud," Accurate photovoltaic power forecasting models using deep LSTM-RNN," Neural Computing and Applications, vol. 31, no. 7, pp. 2727-2740, 2019.
[10] Y. Yao, X. He, T. Huang, C. Li, and D. Xia," A projection neural network for optimal demand response in smart grid environment," Neural Computing and Applications, vol. 29, no. 6, pp. 259-267, 2018.
[11] L. Jin, S. Li, H. M. La, and X. Luo," Manipulability optimization of redundant manipulators using dynamic neural networks," IEEE Transactions on Industrial Electronics, vol. 64, no. 6, pp. 4710-4720, 2017.
[12] X. Fu and S. Li," Control of single-phase grid-connected converters with LCL filters using recurrent neural network and conventional control methods," IEEE Transactions on Power Electronics, vol. 31, no. 7, pp. 5354-5364, 2015.
[13] S. Naz, A. I. Umar, R. Ahmed, M. I. Razzak, S. F. Rashid, and F. Shafait," Urdu Nasta'liq text recognition using implicit segmentation based on multi-dimensional long short term memory neural networks," Springer Plus, vol. 5, no. 1, p. 2010, 2016.
[14] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres," Review of photovoltaic power forecasting," Solar Energy, vol. 136, pp. 78-111, 2016.
[15] S. D. Campbell and F. X. Diebold," Weather forecasting for weather derivatives," Journal of the American Statistical Association, vol. 100, no. 469, pp. 6-16, 2005.
[16] R. J. Bessa, A. Trindade, C. S. Silva, and V. Miranda," Probabilistic solar power forecasting in smart grids using distributed information," International Journal of Electrical Power & Energy Systems, vol. 72, pp. 16-23, 2015.
[17] R. J. Hyndman and G. Athanasopoulos, Forecasting: principles and practice. OTexts, 2018.
[18] C. N. Babu and B. E. Reddy," A moving-average filter-based hybrid ARIMA–ANN model for forecasting time series data," Applied Soft Computing, vol. 23, pp. 27-38, 2014.
[19] A. Mellit, A. M. Pavan, and V. Lughi," Short-term forecasting of power production in a large-scale photovoltaic plant," Solar Energy, vol. 105, pp. 401-413, 2014.
[20] W. VanDeventer et al.," Short-term PV power forecasting using hybrid GASVM technique," Renewable energy, vol. 140, pp. 367-379, 2019.
[21] Jo, T., 2021. Support vector machine. In Machine Learning Foundations (pp. 167-188). Springer, Cham.
[22] Awad, M. and Khanna, R., 2015. Support vector regression. In Efficient learning machines (pp. 67-80). Apress, Berkeley, CA.
[23] Chen, Y., Xu, P., Chu, Y., Li, W., Wu, Y., Ni, L., Bao, Y. and Wang, K., 2017. Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. Applied Energy, 195, pp.659-670.
[24] Vapnik, V.N., 1997, October. The support vector method. In International Conference on Artificial Neural Networks (pp. 261-271). Springer, Berlin, Heidelberg.