

# Power Generation Voting Prediction Model of Floating Photovoltaic System

Ali Jassim Lari

*Electronic and Electrical Engineering  
Department  
Swansea University  
Swansea, UK  
940094@swansea.ac.uk*

Augustine Egwebe

*Electronic and Electrical Engineering  
Department  
Swansea University  
Swansea, UK  
augustine.egwebe@swansea.ac.uk*

Farid Touati

*Electrical Engineering Department  
Qatar University  
Doha, Qatar  
touatif@qu.edu.qa*

Antonio S. Gonzales Jr

*Electrical Engineering Department  
Qatar University  
Doha, Qatar  
antoniojr.gonzales@qu.edu.qa*

Amith Abdullah Khandakar

*Electrical Engineering Department  
Qatar University  
Doha, Qatar  
amitk@qu.edu.qa*

**Abstract**— Solar energy is the most promising renewable energy within the Gulf area as annual solar irradiance is among the highest in the world ( $>2000\text{kWh/m}^2$ ). Therefore, countries within the Gulf area have focused their energy investment on solar energy harvesting, especially Photovoltaics (PV). Photovoltaics (PV) power output is highly dependent on environmental conditions variability. Accurate PV generation power prediction models are essential to investigate the effects of varying environmental conditions and ensure solar power converters' optimum performance whilst meeting peak demand through various environmental conditions. The environmental data which is analysed and discussed in this paper includes air temperature, relative humidity, Photovoltaics (PV) surface temperature, irradiance, dust, wind speed, and output power. The model proposed in this paper optimises and trains three prediction algorithms, including Artificial Neural Network (ANN), Multi-Variate (MV), and Support Vector Machine (SVM). The model deploys three well-known prediction algorithms and voting algorithm to decide the optimum prediction of PV generation power. Furthermore, the voting algorithm shows high prediction accuracy of the output power given the environmental conditions. The Mean Square Error (MSE) for the Artificial Neural network (ANN), Multi-variate (MV), and Support Vector Machine (SVM) are 98, 81, and 82, respectively. In comparison, Mean Squared Error (MSE) of the voting algorithm is significantly lower which is just above 53. The proposed PV power generation prediction algorithm shows reliable outcome with respect to the environmental conditions in Qatar. This tool is expected to assist in the design process of Photovoltaics (PV) plants design where energy generation is highly predictive using proposed voting algorithm.

**Keywords**—PV, Machine Learning, Artificial Neural Network, Environment, Qatar

## I. INTRODUCTION

Renewable energy is clean, sustainable, and environmentally friendly alternative energy source. Since traditional fossil fuel consumption is the main contributor to global warming and climate change problem, investment in various alternative source of energy technologies has increased recently. The primary renewable energy source in the Gulf area is solar energy. Gulf area is one of the highest irradiance areas in the world, with solar irradiance of more than  $2000\text{ kWh/m}^2$ . The high solar irradiance nature of the Gulf area draws massive investment in a regional potential toward increase dependence on clean energy. Also, Photovoltaics power generation techniques gained potential in Gulf areas due to the Photovoltaics cost drop and increase in their efficiency [1]. The major standing issue with solar panels

is the dependency of the output power of the solar panel on environmental conditions. This variability in the output energy introduces to design challenges in solar energy farms to meet rising energy demand. However, the energy demand is predictable and requires predictable supply for a reliable network [2].

Several models have been proposed in the literature to predict PV power generation. The first approach to predict PV power generation is based on the equivalent electrical circuit and optical circuit of the PV cell. Many commercial software packages exist for this purpose, for example, System Advisor Model (SAM). Electrical equivalent circuit of the PV solar panel is good in predicting solar power generation for fixed weather conditions but fails to include environmental conditions [3] [4]. The second approach for predicting PV solar output power involves using statistical and stochastic models to fit the relationship between the PV power output and environmental conditions [5] [6]. Generally, the statistical and stochastic models can account for the uncertainty associated with PV output power. However, one shortcoming of typical statistical models is the low accuracy of predicting PV output power, usually due to the flexibility within the model. The relationship between environmental data and the PV output power is heavily studied to determine the significance of each environmental condition parameter on the PV output power. The study in [7] investigates the minimal environmental monitoring conditions required to predict PV output power. The main significant environmental contributor to the prediction of PV output power is solar irradiance [8]. In addition, several machine learning algorithms have been studied and used to predict renewable sources other than solar energy, including wind, tide, geothermal [9] [10]. Artificial Neural Network (ANN) is one of the most common machine learning algorithms used for power generation prediction of renewable sources. Still, it is usually tuned to fit the specific dataset, known as overfitted model, which minimises the model flexibility and reduces its capacity to predict data outside the training dataset [11].

In this paper, the authors reported the following contributions:

- Development of four statistical prediction models for forecasting the PV power output using the most dominate environmental parameters,
- Comparison of the statistical models using various evaluation metrics,

- Optimisation of the four statistical prediction models without overfitting provided data.

The paper layout is as follows: section II shows the data extraction experiment and discusses the dataset in detail, section III presents the development of the statistical models and optimisation processes, section IV discusses the output results of running the forecast algorithms over the dataset, section V concludes the work and proposes future work.

## II. EXPERIMENT AND DATASET

The environmental parameters with the PV set-up were collected by the author of [11] for typical environment in Qatar, Doha. A. Khandakar, et al. created a system to monitor and log seven parameters, including the power generated by the PV. The monitoring box is built using calibrated sensors, a low-cost (Arduino) microcontroller, LabView and Xbee wireless communication interface. The data were gathered using internet of things (IoT) technology.

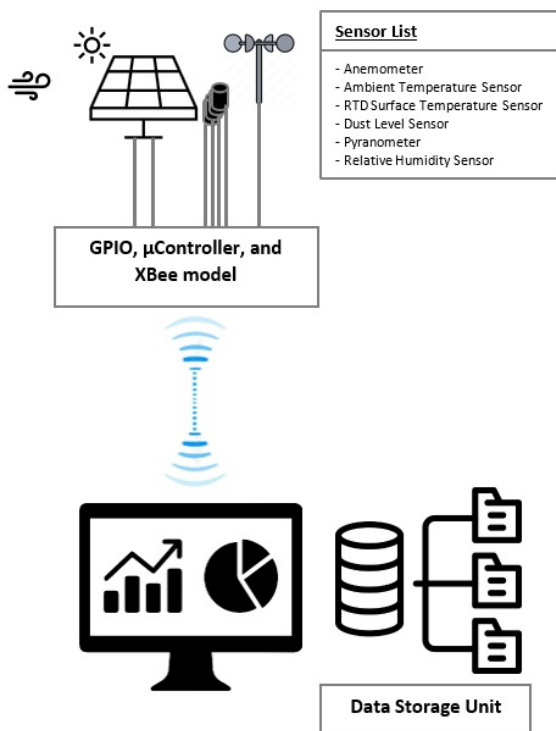


Fig. 1: Data Acquisition Block Diagram

The environmental parameters were chosen carefully to represent PV surrounding environment. The environmental parameters are:

- Air temperature surrounding the PV is well-known to influence any electronic device performance including PV panel. So, the air temperature within the PV set-up is measured and logged in degrees Celsius. The air temperature is measured using analogy temperature sensor LM35DT.
- Relative humidity of the air surrounding the PV is defined as the percentage of water vapor in the air. Humidity is known by its effect on the lifetime of the PV mainly and is expected to have a negative effect on the PV performance due to the disturbance to the sunlight hitting the PV surface. The relative humidity is measured using calibrated HSM-20G sensor.

- PV surface temperature is a very important to measure the effect of the overall heat accumulating on the surface and its effect on the performance. PT100 temperature sensor is used to record high temperature of the PV surface which is attached to the backside of the PV.
- Irradiance is expected to be the most effective environmental parameter on the PV performance. The irradiance measures the amount of solar irradiance covered by the PV surface. SP110 irradiance sensor is attached on the side of the PV surface and logged the irradiance in W/m<sup>2</sup>.
- Dust accumulation on the PV surface is a major contributor to the PV performance where dust blocks the irradiance to reach the PV. GP2Y1010AU0F dust sensor is used to measure the dust amount in milligram within cubic meter.
- Wind speed is measured using anemometer. The wind speed is not expected to have a direct effect on the PV efficiency, but it would indirectly because it will most probably bring dust or influence other parameters such as humidity or temperature. Wind speed is logged in km/h unit.
- Output power of the PV is measured using the measurement of output current and voltage of the PV. The PV output terminals are connected to IGBT which acts as direct load and uses maximum power point tracking to ensure the maximum power is drawn from the PV.

The PV used in this experiment is Poly-crystalline silicon PV module of a maximum power of 80-Watt, area of 0.6426 m<sup>2</sup>, open circuit voltage is 21 Volt, and Short circuit current is 5.24 A [11].

## III. METHOD

The method developed in this paper aims at building accurate prediction model for PV power generation based on novel voting mechanism of three different predictors: Support Vector Machine (SVM) predictor, Multi-variate predictor (MV), and Artificial Neural Network (ANN). The three prediction models are developed using 5 folds cross-validation technique. The raw data is split randomly into 5 folds where every four folds are used to train the model and the fifth fold is used to validate the model. The process is repeated five times and the average performance of the prediction method is computed.

1. 5-fold Cross-validation  
The dataset is partitioned into five folds. The data rows are split into the five folds randomly. The random generation seed is set to a fixed value to ensure that the tests are reproducible; this step would also provide proper and consistent compression of various runs and tests on the code.
2. Feature Selection  
The data consists of six features (presented in Table 1). The algorithm is configured to test all the combinations, and the result of the best combination is used. Since all environmental conditions are expected to contribute to the PV power output, the rank of features contribution to the PV power output

is listed to allow for a more straightforward data acquisition system.

### 3. Model Training

In this model, three different machine learning models are used to produce a voting system. The three models are classically used on their own but not as a combination.

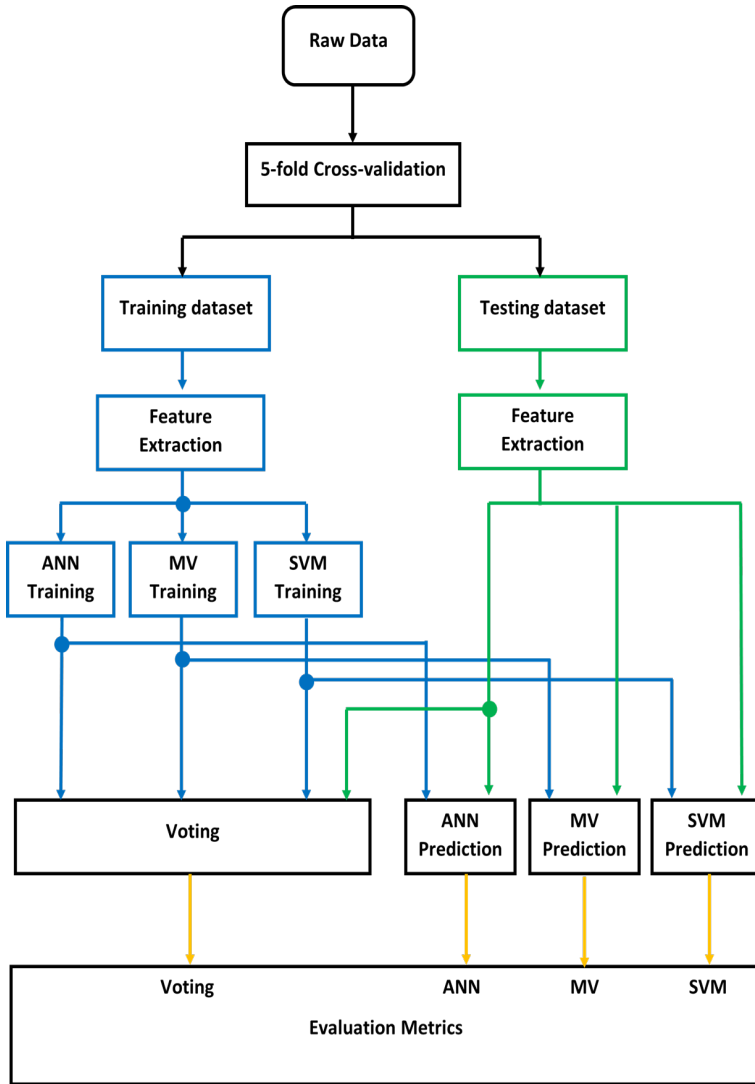


Fig. 2: PV Generation Power Prediction Block Diagram

The predictive models are Multi-Variate prediction, Support Vector Machine (SVM), and Artificial Neural Network (ANN). Each model is optimised to give the best performance.

#### 3.1 Multi-Variate Linear Regression

The simplest prediction method is linear prediction which models the relationship between the features and the dependent variable. Each feature is correlated with the dependent variable, which in this case is the PV power generation.

$$Y_i = \alpha_0 + \alpha_1 x_{i1} + \dots + \alpha_p x_{ip} \quad (2)$$

Where Y is the PV generation power, X is the vector of environmental conditions, and  $\alpha$  is the mapping factor. The model parameters are

selected using an iterative technique by minimising the sum of the least-squares error.

$$\text{Min} \sum_{i=1}^n \left( y_i - (\alpha_0 + \alpha_1 x_{i1} + \dots + \alpha_p x_{ip}) \right)^2 \quad (3)$$

#### 3.2 Support Vector Machine (SVM)

The Support Vector Machine (SVM) allows flexibility within the error of outliers. The optimisation of the relationship between the feature set and the dependent variable is done based on the constraint's optimisation equation.

$$\text{Min} \frac{1}{2} \|w\|^2 \quad (4)$$

where,  $|y_i - w_i x_i| \leq \epsilon$

Where U is the actual PV generation power, w is the mapping factor, and X is the vector of environmental conditions. The optimisation process allows for an acceptable margin of  $\epsilon$ . The margin parameter can be tuned to gain the most accurate forecast.

#### 3.3 ANN

The artificial neural network has three layers: input, hidden, and output layers. The number of neurons in each layer is selected in the optimisation process of the neural network to give the best forecast results. The learning algorithm used is the backpropagation method, where the data of the training dataset are fed back to the neural network to optimise the weight values.

### 4. Voting Mechanism

The three models are trained to produce the best forecast possible. The voting mechanism is suggested to compensate for one model over the other model within specific outcomes.

The voting block produces the average of the outcomes of the three models. Other voting mechanisms such as median value is also tested and reported in the results.

### 5. Evaluation

The results are being evaluated based on mean squared error (MSE) given in (5), root mean squared error (RMSE) shown in (6), relative mean squared error (rMSE) given in (7), normalised forecast metric (NFM) shown in (8) and coefficient of determination  $R^2$  given in (9).

All the statistical analyses listed above are used on each model prediction solely in addition to the voting output system.

$$\text{MSE} = \sum_{i=1}^n \frac{(Y_i - Y'_i)^2}{n} \quad (5)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^n \frac{(Y_i - Y'_i)^2}{n}} \quad (6)$$

$$rMSE = \frac{n-1}{n} \times \frac{\sum_{i=1}^n (Y_i - Y'_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (7)$$

$$NFM = \sum_{i=1}^n \frac{Y'_i - Y_i}{Y'_i + Y_i} \quad (8)$$

$$R^2 = 1 - rMSE \quad (9)$$

Although the mean squared error and root mean squared error give the overall image of the model prediction error, they do not provide the correct forecast since under predicted values are equally treated as over predicted values.

The normalised forecast metric (NFM) is used within a prediction window to evaluate the forecast either over forecasted or under forecasted as NFM metric value can range from -1 to 1 where a negative value indicates under-forecast, and a positive value indicates over-forecast. The zero value of the NFM indicates a perfect match between the forecast value and the actual value.

The relative mean squared error metric is used to reflect the error range within the data variation. The more significant output power variation should be more acceptable for a more extensive range of error values.

#### IV. RESULTS AND DISCUSSION

The machine learning algorithms were implemented on MATLAB R2020b system on Intel Core i5-3320M 2.4GHz CPU and 8GB of memory. The three main algorithms: SVM, Multi-variate Regression, and ANN, ran in series. The results of each algorithm were subsequently used as input for the proposed voting algorithm.

##### 1- ANN

The artificial neural network hyperparameters are being selected using sensitivity analysis method. The number of hidden layers increased gradually from 1 hidden layer up to 100 hidden layers in the ANN. The ANN MSE error is minimum at 22 hidden layers as shown in Fig. 3.

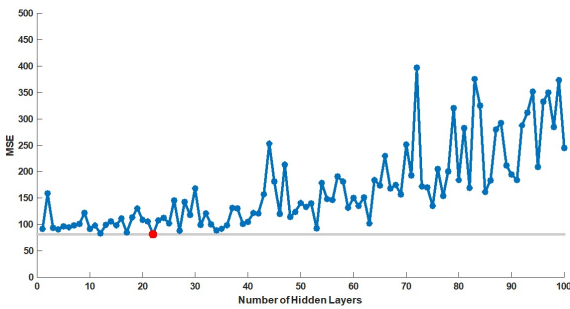


Fig. 3: Mean Squared Error vs Number of Hidden Layers

The best training algorithm is Bayesian Regularisation as shown in Fig. 4. Bayesian Regularisation training function is proven to be more robust than standard backpropagation training function [12], although the variation of the training algorithms is not large.

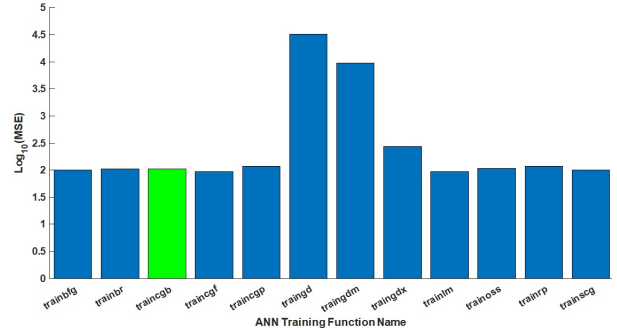


Fig. 4: Log of Mean Squared Error for Training Algorithms

The optimal parameters are selected to set-up the ANN. The training process shows quick learning progress at 609 Epochs. The ANN is set-up to be trained using randomly allocated 70% of the dataset, 15% for validation, and 15% for testing. Fig. 5 shows the performance progress with every training epoch iteration.

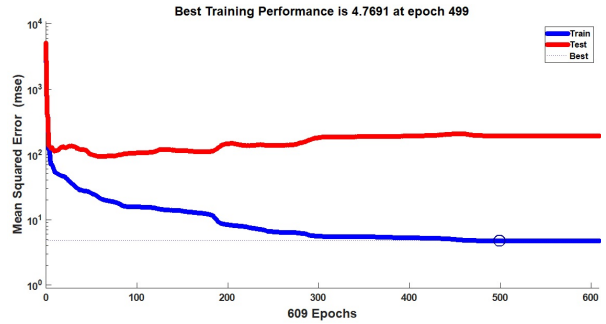


Fig. 5: Mean Squared Error vs Epochs

The error distribution is shown in Fig. 6. The largest number of instances have MSE around the zero error.

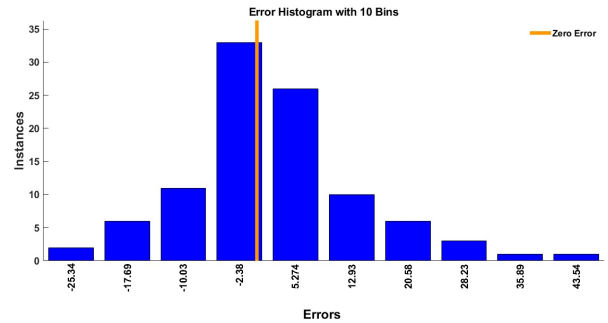


Fig. 6: ANN Output Error Histogram

The prediction plot shows high match between the prediction output and the real output power with correlation factor is above 85%. The most significant diverts within the prediction can be seen either at high power output or very low power output. This means the prediction system is less sensitive at both ends of the power spectrum.

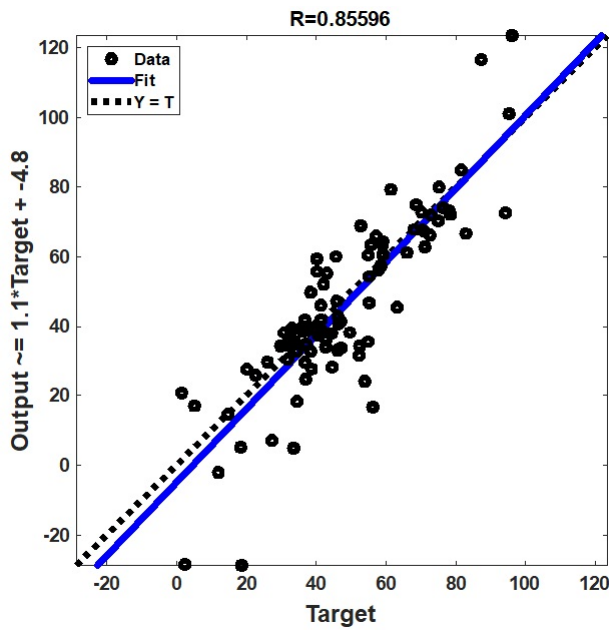


Fig. 7: ANN Prediction vs Target

### 2- Multi-variate

The multi-variate algorithm produces polynomial prediction parameters within iterations. The results of the prediction are shown in Fig. 8. The results show that the error of the prediction is centred around the zero MSE.

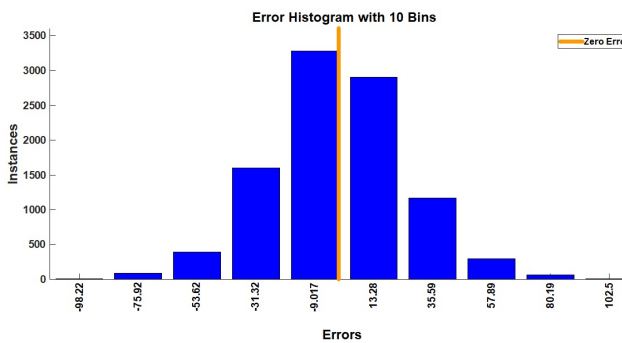


Fig. 8: Multi-Variate Error Histogram

### 3- SVM

The support vector machine algorithm fits the data within polynomial function with leverage of support vectors. The support vectors allow acceptable error when adjusting the polynomial prediction function. The most instances error is around the zero MSE.



Fig. 9: SVM Error Histogram

### 4- Voting System

The results of the three prediction methods are used to vote into more accurate prediction. The prediction output of the four methods is presented in the following graph.

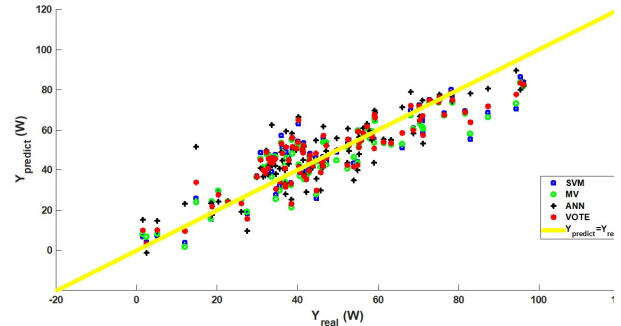


Fig. 10: Predicted Output Power vs Actual Output Power

Voting algorithm shows better results comparing with the rest of the three other methods. The algorithm guarantee the output to be better than the worst prediction algorithm because it produces the average of three algorithm.

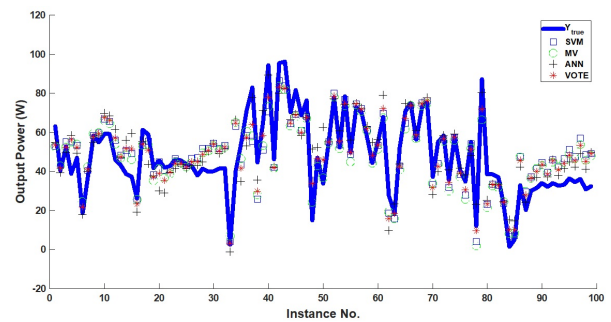


Fig. 11: Output Power for Each Instance

Table 1 : Forecasting Results

Algorithm\Metric	MSE	RMSE	rMSE	R <sup>2</sup>
ANN	98	9.87	0.26	0.74
Mutli-Variate	81	8.99	0.21	0.79
SVM	82	9.07	0.21	0.79
Voting	<u>53</u>	<u>7.33</u>	<u>0.14</u>	<u>0.86</u>

The test batch of data is presented in the graph above with the prediction output of the four algorithms. The average MSE, RMSE, rMSE, and R2 metrics are calculated for the four algorithms are presented in the following table. The Voting algorithm shows clearly less prediction error and higher correlation in the prediction comparing to the rest of individual methods.

The voting algorithm shows the lowest prediction mean square error by 40% of the average mean square error of the ANN, Multi-Variate, and SVM methods. The NFM graph, in Fig. 12, shows the over-forecasting and under-forecasting instances in the four algorithms. The graph shows a consistent error among the four algorithms for specific instances which could be related to the PV output power at this point.



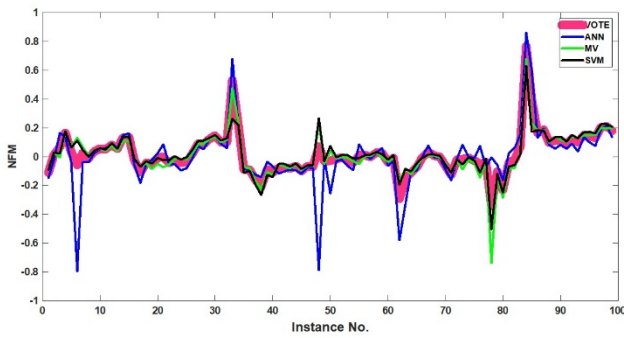


Fig. 12: Bias Calculation using NFM Technique

Fig. 13 shows the points where the data is over-forecasted or under-forecasted significantly where the PV output power at these points is within the least 10% reported PV output power. One possible explanation of this data forecasting trend is a systematic under or over report of PV output power caused by limits of the hardware acquisition system used to collect the dataset.

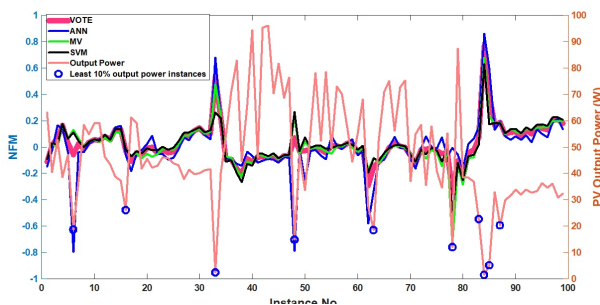


Fig. 13: NFM Calculation and PV Output Power

#### CONCLUSIONS AND FUTURE WORK

PV output power accurate prediction based on environmental data was developed using combination of machine learnings. The machine learning models were trained based on data collected by Qatar University. To conclude, the prediction model was developed using a voting algorithm for three machine learning algorithm output, ANN, Multi-Variate, and SVM. Each algorithm was configured to give least error over five folds of cross-validation. The ANN algorithm is constructed using 22 hidden layers which shows the best outcome based on sensitivity analysis method. The training algorithm used is Bayesian Regularisation not only because its robustness but also because it shows the best performance among 12 different training algorithms. The other two algorithms are multi-variate and SVM showed similar results which are more accurate than ANN. The proposed voting algorithm generates the output based on the average output of the three previous prediction methods. The proposed model is trained and built based on environmental data from Qatar area so the model can be used to optimise the solar power plant design. To determine the number of PV cells within the solar power plant, environmental conditions can be used as the input of the proposed model within the Gulf area to ensure required power demand is met. Further environmental data can be used to train the voting model to produce more accurate PV model. This environmental data includes time, date, GPS location of the solar panel, power drawn from the

panel, etc. The model can be used to compare various types of PV models and features such as moving PV panels, cleaning models, and cooling models.

#### BIBLIOGRAPHY

- [1] A. Sayyah, M. Horenstein and M. Mazumder, "Energy yield loss caused by dust deposition on photovoltaic panels," *Solar Energy*, vol. 107, p. 576–604, 2014.
- [2] A. Ennaoui, B. Figgis and D. Plaza, "Outdoor Testing in Qatar of PV Performance, Reliability and Safety.," in *Qatar Foundation Annual Research Conference Proceedings*, Doha, Qatar, 2016.
- [3] F. Touati, N. Chowdhury, K. Benhmed, A. Gonzales, M. Al-Hitmi, M. Benammar and A. Gastli, "Long-term performance analysis and power prediction of PV technology in the State of Qatar," *Renew. Energy*, vol. 113, p. 952–965, 2017.
- [4] Y. S. G. a. B. A. Sun, "Solar PV output prediction from video streams using convolutional neural networks," *Energy & Environmental Science*, vol. 11, pp. 1811-1818, 2018.
- [5] A. L. S. a. M. G. Dolara, "Comparison of different physical models for PV power output prediction," *Solar energy*, vol. 119, pp. 83-99, 2015.
- [6] A. S. T. F. T. a. K. C. Yona, "Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead PV power output correction," *IEEE Transactions on Sustainable Energy*, vol. 4, pp. 527-533, 2013.
- [7] W. J. E. T. G. S. M. S. T. H. B. M. S. a. S. A. VanDeventer, "Short-term PV power forecasting using hybrid GASVM technique," *Renewable energy*, pp. 367-379, 2019.
- [8] A. H. G. B. a. U. R. Drews, "Quality of performance assessment of PV plants based on irradiation maps.," *Solar Energy*, vol. 11, pp. 1067-1075, 2008.
- [9] S. Mishra and P. Dash, "Short term wind power forecasting using Chebyshev polynomial trained by ridge extreme learning machine," in *IEEE Power, Communication and Information Technology Conference (PCITC)*, Bhubaneswar, Odisha, India, 2015.
- [10] S. Netsanet, J. Zhang, D. Zheng, R. Agrawal and F. Muchahary, "An aggregative machine learning approach for output power prediction of wind turbines," in *IEEE Texas Power and Energy Conference (TPEC)*, College Station, TX, USA, 2018.
- [11] A. Khandakar , M. . E. Chowdhury , M.-K. Kazi, K. Benhmed, F. Touati, M. Al-Hitmi and A. . J. S. P. Gonzales, "Machine learning based photovoltaics (PV) power prediction using different environmental parameters of Qatar," *energies*, vol. 12, no. 14, p. 2782, 2019.
- [12] F. a. W. D. Burden, "Bayesian regularization of neural networks," *Artificial neural networks*, pp. 23-42, 2008.