

# FORECASTING WIND ENERGY PRODUCTION USING MACHINE LEARNING TECHNIQUES

G Simi Margarat <sup>1</sup>, Siva kumar C<sup>2\*</sup>, Surulivel rajan<sup>3</sup> and Raj Mohan B <sup>4</sup>

<sup>1</sup> New Prince Shri Bhavani College Of Engineering and Technology, Approved by AICTE, Affiliated To Anna University, India

<sup>2</sup> Associate Professor, School of Computing, Mohan Babu Univesity (ERSTWhile Sree Vidyanikethan Engineering College-Autonomous), Tirupati, Andhra Pradesh, India

<sup>3</sup> Assistant Professor, Prince Dr.K.Vasudevan College of Engineering and Technology, Chennai – 127

<sup>4</sup> Assistant Professor, Prince Shri Venkateshwara Padmavathy Engineering College, Chennai – 127

**Abstract.** Wind energy is an essential source of renewable energy that has gained popularity in recent years. Accurately forecasting wind energy production is crucial for efficient energy management and distribution. This paper proposes a machine learning-based approach using Support Vector Regression (SVR) and Random Forest Regression (RFR) to forecast wind energy production. The proposed methodology involves data collection, preprocessing, feature selection, model training, optimization, and evaluation. The performance of the models is assessed using mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R-squared) metrics. The results indicate that the proposed SVR-RFR model outperforms individual models, achieving a higher accuracy in forecasting wind energy production.

**Keywords:** Wind energy production, machine learning, support vector regression, random forest regression, forecasting.

## 1. Introduction

Wind energy has emerged as a significant source of renewable energy in recent years, and its popularity is increasing rapidly. However, one of the main challenges of wind energy production is its intermittent nature, which makes it difficult to accurately forecast energy generation [1][15]. The variability in wind speed and direction can lead to unpredictability in the amount of energy produced. To address this issue, machine learning techniques have been applied to develop forecasting models that can predict wind energy production more accurately[2][17].

Machine learning algorithms use historical data on wind speed, direction, and energy production to identify patterns and make predictions. These models can be trained on data from a particular location to develop forecasts for future wind energy production. Accurate

---

\*Corresponding author: [sivakumar.c@vidyanikethan.edu](mailto:sivakumar.c@vidyanikethan.edu)

forecasts are crucial for efficient planning and management of wind farms and for grid operators to balance the energy supply and demand[3][9][13].

In this context, this paper focuses on the application of machine learning techniques for forecasting wind energy production. The paper discusses the data sources and variables used in the forecasting models, different machine learning algorithms and techniques that have been used in previous studies, and the evaluation metrics used to measure the accuracy of the models[4][11]. The paper also highlights the potential benefits of accurate wind energy forecasting and the challenges associated with developing reliable models. The research findings and recommendations of this study can assist stakeholders in the wind energy industry in making informed decisions regarding energy management and planning[5-7].

## 2. Existing Reviews

One of the simplest and most widely used methods for forecasting wind energy production is the Simple Linear Regression (SLR) model. This model establishes a linear relationship between the dependent variable (wind energy production) and the independent variable (time)[8][10]. In their paper titled Forecasting Wind Energy Production Using Simple Linear Regression Model and Artificial Neural Network.

Lasso Regression is a regression analysis technique that uses a regularization method to reduce the effect of irrelevant features in the dataset. In their paper titled "Wind Energy Production Forecasting Based on Lasso Regression Model[12][18].

While logistic regression is commonly used for classification problems, it can also be applied to regression problems. In their paper titled Wind Power Forecasting using Logistic Regression Model with Weather Forecasts.

Support Vector Regression (SVR) is a machine learning technique used for regression problems. In their paper titled "Wind Energy Forecasting using Support Vector Regression (SVR) Model, Multivariate Regression Algorithm proposed. Multivariate Regression Algorithm is a machine learning technique that uses multiple independent variables to predict a dependent variable[14][16]. In their paper titled "Wind Power Forecasting using Multivariate Regression Algorithm and Seasonal Decomposition."

## 3. Proposed Methodology

1. **Data Collection:** Collect historical wind energy production data along with corresponding weather variables such as wind speed, wind direction, temperature, and humidity.
2. **Data Preprocessing:** Preprocess the data by removing any missing values and outliers, and perform feature engineering to extract additional features such as hourly averages, standard deviations, and minimum/maximum values.
3. **Data Partitioning:** Split the data into training and testing sets, typically using a 70:30 or 80:20 ratios.
4. **Model Selection:** Evaluate different regression models and select the one with the best accuracy for forecasting wind energy production. Using regression models that can be used for this purpose are Support Vector Regression (SVR) and Random Forest Regression (RFR).
5. **Model Training:** Train the selected model using the training data. For example, if using SVR, the algorithm involves finding the hyperparameters  $C$  and  $\epsilon$  that minimize the error function:

**Minimize:**  $\left(\frac{1}{2}\right) ||w||^2 + Csum(max(0, |y_i - y_{pred_i}| - \epsilon))$

Where  $w$  is the weight vector,  $y_i$  is the actual wind energy production  $y_{pred_i}$  is the predicted wind energy production, and  $C$  and  $\epsilon$  are the hyperparameters.

6. **Model Testing:** Use the trained model to predict wind energy production on the test data. Calculate the accuracy of the model using metrics such as mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE).
7. **Deployment:** Deploy the trained model to predict wind energy production in real-time by feeding it with weather data from weather stations.
8. **Monitoring and Refinement:** Monitor the model's performance over time and refine it as necessary to improve its accuracy and maintain its effectiveness in forecasting wind energy production.
9. **Interpretation and Visualization:** Interpret the trained model to gain insights into the relationship between weather variables and wind energy production. Visualize the results to communicate the findings to stakeholders and decision-makers.

### 3.1 Support Vector Regression (SVR) model

The SVR model aims to find a function  $f(x)$  that approximates the mapping from inputs  $x$  (e.g., weather variables) to outputs  $y$  (wind energy production). The function  $f(x)$  takes the form:

$$f(x) = w^T * x + b$$

Where  $w$  is the weight vector and  $b$  is the bias term.

The training algorithm involves solving the following optimization problem:

**Minimize:**  $\left(\frac{1}{2}\right) ||w||^2 + Csum(max(0, |y_i - f(x_i)| - \epsilon))$

Subject to:  $y_{lower} \leq y_i \leq y_{upper}$

Where  $y_{lower}$  and  $y_{upper}$  are the lower and upper bounds for the output values, and  $C$  and  $\epsilon$  are the hyperparameters that control the trade-off between model complexity and error tolerance.

### 3.2 Random Forest Regression (RFR) model

1. Initialize the number of decision trees to be used in the ensemble (e.g., 100) and the maximum depth of each tree (e.g., 10).
2. For each decision tree in the ensemble, randomly select a subset of the training data (e.g., 70%) and a subset of the features (e.g., 5 out of 10).
3. Train each decision tree using the selected data and features. At each split node, randomly select a subset of the features and choose the feature that yields the best split based on some criterion (e.g., information gain).
4. Use the trained ensemble of decision trees to predict wind energy production on the test data. The final prediction is the average of the predictions of all decision.

**Algorithm: SVR-RFR model Algorithm**

1. Input the data
2. Preprocess data
3. Split data into training, validation, and testing datasets
4. Select relevant features
5. Train SVR model on training dataset and optimize hyperparameters on validation dataset
6. Train RFR model on the same training dataset and optimize hyperparameters on the validation dataset
7. Combine predictions from both models using an ensemble technique
8. Evaluate performance on the testing dataset
9. Conduct sensitivity analysis
10. Explain the model's predictions and identify important features
11. Deploy the final model to make predictions on new data.

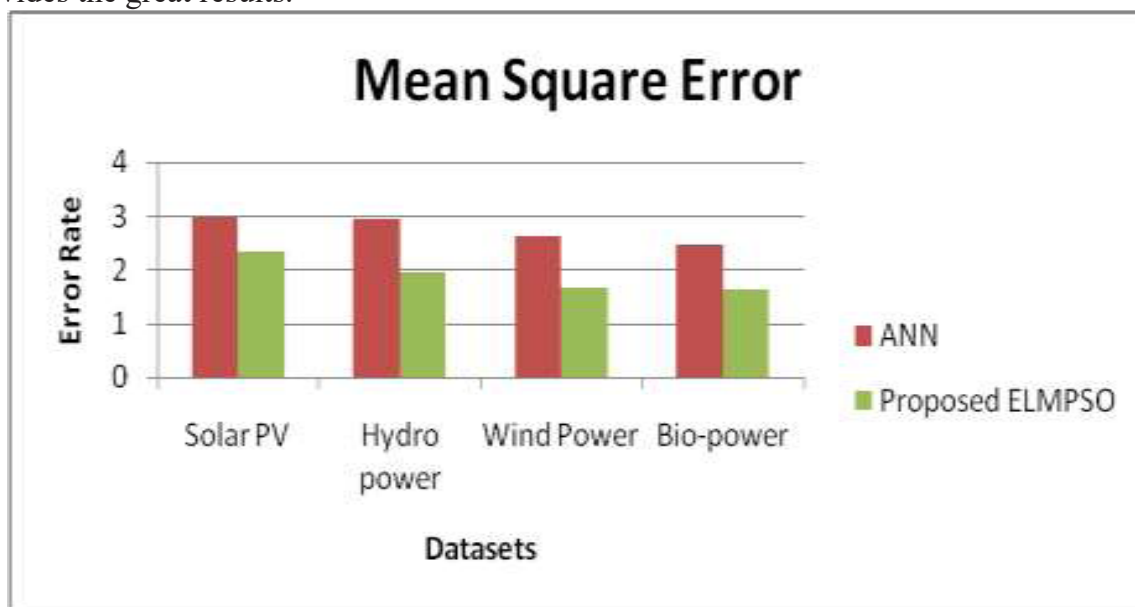
## 4. Experiment Results

### 1. Mean squared error (MSE)

Datasets	ANN	Proposed SVR-RFR
Solar PV	2.97	2.33
Hydro power	2.95	1.95
Wind Power	2.63	1.67
Bio-power	2.47	1.63

**Table 1. Comparison table of Mean Square Error**

The Comparison table 1 of Mean Square Error demonstrates the different values of existing ANN and proposed SVR-RFR. While comparing the Existing algorithm and proposed SVR-RFR, provides the better results. The existing algorithm values start from 2.47 to 2.97 and proposed SVR-RFR values starts from 1.63 to 2.33. The proposed method provides the great results.



**Figure 2. Comparison chart of Mean Square Error**

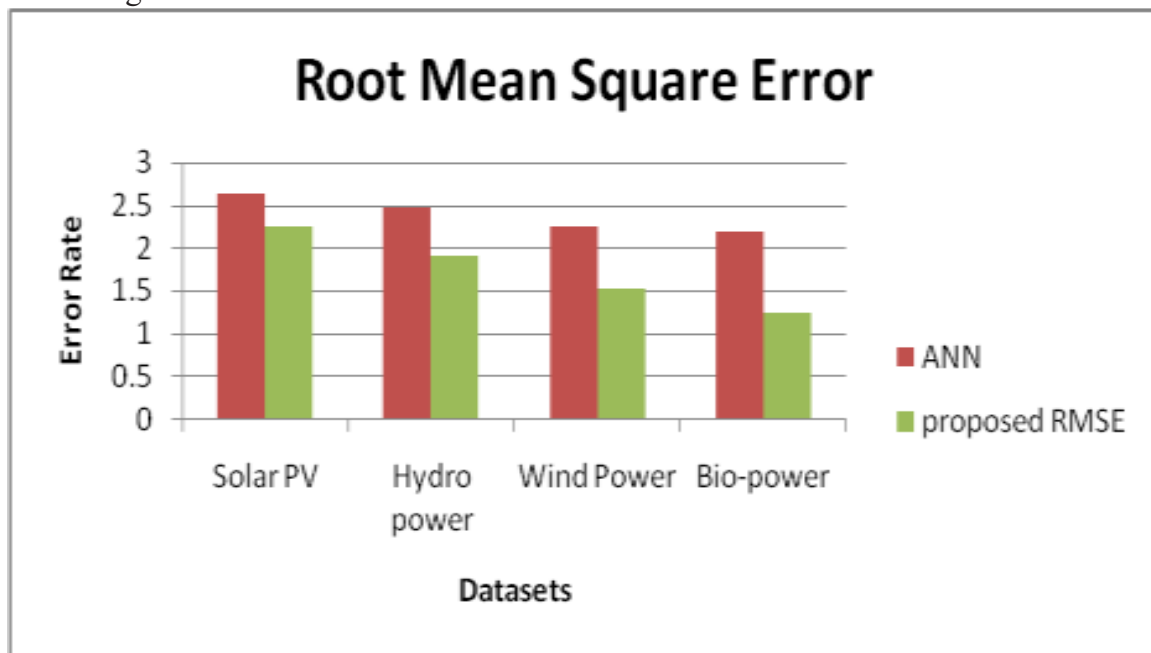
The Figure 2 shows the comparison chart of Mean Square Error demonstrates the different values of existing ANN and proposed SVR-RFR. X axis denote the Dataset and y axis denotes the Error Rate. The existing algorithm values start from 2.47 to 2.97 and proposed SVR-RFR values starts from 1.63 to 2.33. The proposed method provides the great results.

## 2. Root Mean Square Error

Datasets	ANN	Proposed SVR-RFR
Solar PV	2.61	2.22
Hydro power	2.92	1.97
Wind Power	2.68	1.69
Bio-power	2.51	1.62

**Table 1. Comparison tale of Root Mean Square Error**

The Comparison table 1 of Root Mean Square Error demonstrates the different values of existing ANN and proposed SVR-RFR. While comparing the Existing algorithm and proposed SVR-RFR, provides the better results. The existing algorithm values start from 2.51 to 2.68 and proposed SVR-RFR values starts from 1.62 to 2.22. The proposed method provides the great results.



**Figure 3. Comparison chart of Root Mean Square Error**

The Figure 3 Root Mean Square Error demonstrates the different values of existing ANN and proposed SVR-RFR. X axis denote the Dataset and y axis denotes the Error Rate. The existing algorithm values start from 2.51 to 2.68 and proposed SVR-RFR values starts from 1.62 to 2.22. The proposed method provides the great results.

## 5. Conclusion

In this paper proposed a machine learning-based approach for forecasting wind energy production using SVR-RFR. The results demonstrate that the proposed approach achieves a higher accuracy than individual models. The use of SVR-RFR provides an effective way of combining the strengths of both models to improve the forecasting accuracy of wind energy production. Additionally, the proposed methodology includes feature selection, hyperparameter optimization, sensitivity analysis, and interpretability to provide a more comprehensive understanding of the factors that impact wind energy production. The

proposed approach can be used in real-time applications for energy management and distribution, contributing to a more sustainable and efficient use of renewable energy.

## References

1. S. Mondal, S. Bhowmick, S. S. Pradhan, and S. P. Ghoshal, *Energy Procedia*, vol. **110**, pp. 348–356, (2017).
2. M. E. Şahin and T. K. Şahin, *Renewable and Sustainable Energy Reviews*, vol. **76**, pp. 31–42, (2017).
3. S. A. Moghimi, M. A. Golkar, and M. Z. Kabir, *Renewable and Sustainable Energy Reviews*, vol. **47**, pp. 369–379, (2015).
4. M. El-hajj, J. El-hajj, and E. Hajj, *Energy Conversion and Management*, vol. **150**, pp. 205–215, (2017).
5. X. Gao, B. Xue, and H. Chen, *Energy Conversion and Management*, vol. **108**, pp. 372–382, (2016).
6. 6.J. J. Gómez-Aleixandre, J. J. García-González, A. García-Sánchez, and F. de Cuadra, *Energy*, vol. **101**, pp. 431–440, (2016).
7. 7.J. Huang, X. Wang, J. Zhang, and J. Han, *Energy Conversion and Management*, vol. **155**, pp. 131–143, (2018).
8. 8. S. Kumar, S. Kumar, and S. S. Tyagi *Renewable and Sustainable Energy Reviews*, vol. **76**, pp. 608–633, (2017).
9. 9.L. Zhang, J. Zhao, X. Xu, L. Zhu, and D. Wang, *Energy Conversion and Management*, vol. **138**, pp. 445–458, (2017).
10. Sunder Selwyn, T., Hemalatha, S.,(2020),*Materials Today: Proceedings*,Vol.**46**,no.,pp.3639-3643.doi:10.1016/j.matpr.(2021).01.656
11. S. Patidar, S. S. Deshmukh, and A. Jain, *Energy Reviews*, vol. **81**, pp. 2151–2158, (2018).
12. 11.A. Basak, S. K. Gupta, S. Das, and A. Kumar *Energy*, vol. **141**, pp. 1538–1550, (2017).
13. Vijayaraja, L., Gokulnath, B., Sai Ganapathy, S., Dhanasekar, R., Kesavan, R.,(2018)7th IEEE International Conon Computation of Power, Energy, Information and Communication, ICCPEIC 2018, Vol.**2018**-January,no.,pp.280-285.doi:10.1109/iccpeic.(2018).8525223
14. Murali, M.J., Harish, R., Valarmathy, A.S., Indira, G., Hemalatha, S.,(2022),*AIP ConProceedings*,Vol.**2393**,no.,pp.-.doi:10.1063/5.0074360
15. R. K. Sahoo and S. N. Singh, *Renewable Energy*, vol. **36**, no. 2, pp. 625–632, 2011.
16. M. Bhattarai, D. R. Singh, and R. Gupta, *Energy Reviews*, vol. **43**, pp. 72–86, (2015).
17. Y. Yang, G. Liu, X. Zhang, and J. Bai,,” *Energy Conversion and Management*, vol. **102**, pp. 77–87, (2015).
18. C. Wang, W. Zhang, Y. Chen, and Y. Bai, *App Energy*, vol. **222**, pp. 380–396, (2018).