International Journal of Power Electronics and Drive Systems (IJPEDS)
Vol. 15, No. 3, September 2024, pp. 1925~1933
ISSN: 2088-8694, DOI: 10.11591/ijpeds.v15.i3.pp1925-1933

Machine learning applications for predicting system production in renewable energy

Deepa Somasundaram[1], R. Muthukumar[2], N. Rajavinu[3], Kalaivani Ramaiyan[4], P. Kavitha[5]
[1]Department of Electrical and Electronics Engineering, Panimalar Engineering College, Chennai, India
[2]Department of Electrical and Electronics Engineering, Erode Sengunthar Engineering College, Perundurai, India
[3]Department of Electrical and Electronics Engineering, S. A. Engineering College, Chennai, India
[4]Department of Electronics and Communication Engineering, Rajalakshmi Engineering College, Chennai, India
[5]Department of Electrical and Electronics Engineering, R. M. K. Engineering College, Kavaraipettai, India

Article Info

Article history:

Received Dec 27, 2023
Revised Mar 6, 2024
Accepted Apr 5, 2024

Keywords:

Energy consumption
Energy management
Performance metrics
Regression model
Renewable energy
Solar system
Wind system

ABSTRACT

Renewable energy systems play pivotal role in addressing the global challenge of sustainable energy production. Efficiently harnessing energy from renewable sources requires accurate prediction models to optimize system production. This paper delves into the realm of predictive modeling, focusing on the utilization of machine learning techniques to forecast system production in renewable energy systems. The investigation incorporates a range of factors such as wind speed, sunshine, air pressure, radiation, air temperature, and relative air humidity, alongside temporal data ('Date-Hour (NMT)'). These factors undergo rigorous curation and preprocessing to ensure the reliability and quality of the predictive model. Various machine learning algorithms, including linear regression, decision tree, random forest, and support vector machine (SVM), are employed to examine the relationships between these factors and system production. The findings are assessed using metrics such as mean squared error, mean absolute error, and R-squared. Through comparative analysis, the study illuminates the strengths and limitations of each algorithm, providing valuable insights into their suitability for renewable energy forecasting. This paper adds to renewable energy research by examining how machine learning predicts system production. The insights are valuable for researchers, practitioners, and policymakers in sustainable energy development.

Corresponding Author:

Deepa Somasundaram
Department of Electrical and Electronics Engineering, Panimalar Engineering College
Chennai, India
Email: dee_soms123@yahoo.co.in

1. INTRODUCTION

The global pursuit of sustainable energy sources in response to climate change and the depletion of conventional fossil fuels has intensified. Renewable energy systems, particularly those harnessing wind and solar power, offer a promising solution. Accurate forecasting of energy production in these systems is crucial, influenced by meteorological and environmental factors [1], [2]. This study delves into the application of machine learning techniques to predict system production in renewable energy systems, focusing on features such as wind speed, air pressure, sunshine, air temperature, radiation, and relative air humidity [3]-[5].

Recent years have seen a significant integration of machine learning in renewable energy research, driven by the potential to enhance the accuracy of energy production predictions. The research aims to contribute to this field by examining the predictive capabilities of machine learning algorithms. With a specific emphasis on optimizing the efficiency and reliability of renewable energy systems, the study addresses the broader transition towards sustainable energy practices [6]-[9].

This work emphasizes the critical importance of accurate system production predictions for effective energy planning and management. Governments, energy utilities, and private stakeholders heavily rely on precise forecasts for optimizing energy distribution and making informed decisions about energy investments [10]. The study's outcomes have the potential to impact both operational efficiency and economic viability of renewable energy projects, contributing to sustainable and resilient energy infrastructure [11]-[13].

Moreover, the paper delineates its specific research objectives, which encompass conducting a comparative analysis of various machine learning algorithms for predicting system production. Through the evaluation of the performance of models such as linear regression, decision tree, random forest, and support vector machine (SVM), the study seeks to offer insights into the most efficient methodologies for achieving accurate and dependable predictions. This exploration enhances the comprehensiveness of the research, offering a nuanced understanding of algorithmic strengths and weaknesses in the context of renewable energy forecasting [14]-[16]. Figure 1 shows the air pressure vs time. Figure 2 shows the sunshine vs time. Figure 3 shows the radiation vs time. Figure 4 shows the wind speed vs time. Figure 5 shows the correlation matrix of various factors. In summary, the introduction and literature review chapters set the stage for exploring machine learning applications in predicting system production in renewable energy systems. The research addresses a critical gap in knowledge and contributes to the broader discourse on sustainable energy by offering novel insights into the effectiveness of various algorithms in optimizing energy forecasting [17], [18]. Figure 6 shows the application of machine learning in renewable energy prediction.

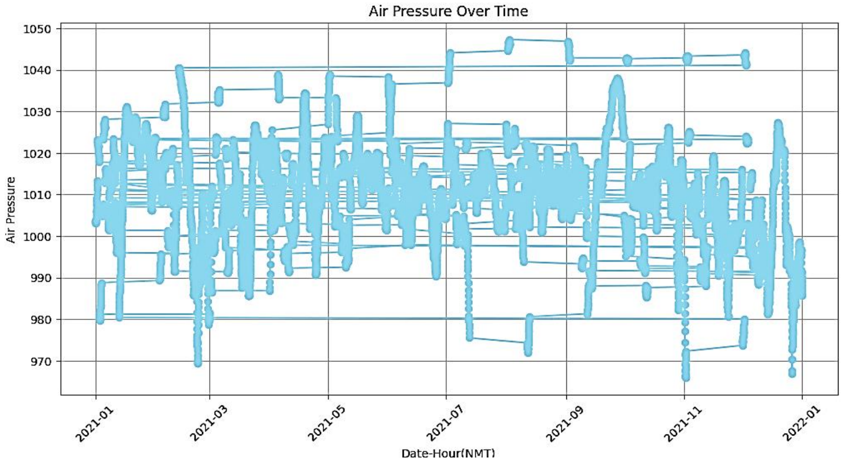

Figure 1. Air pressure vs time

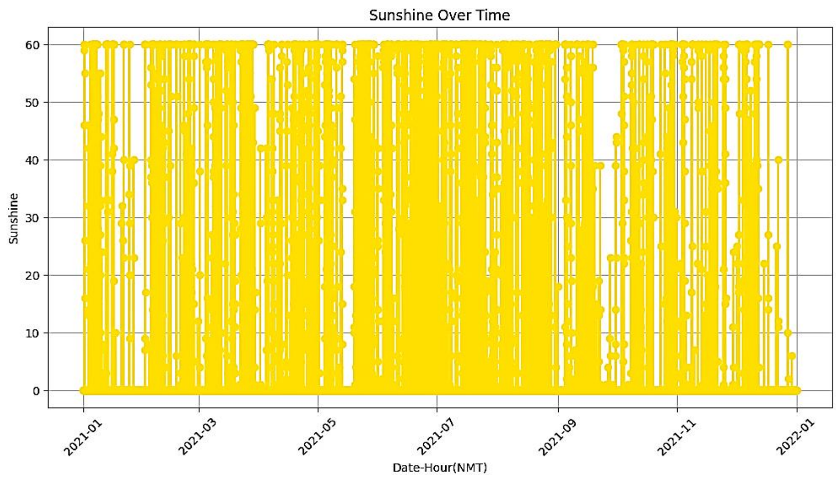

Figure 2. Sunshine vs time

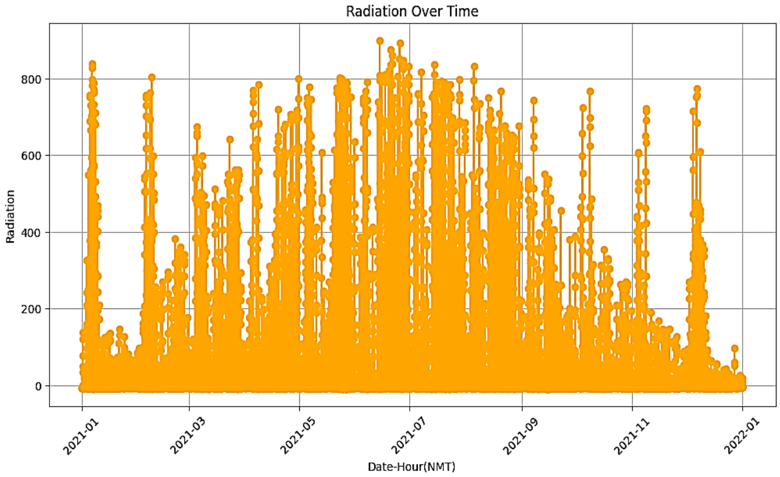

Figure 3. Radiation vs time

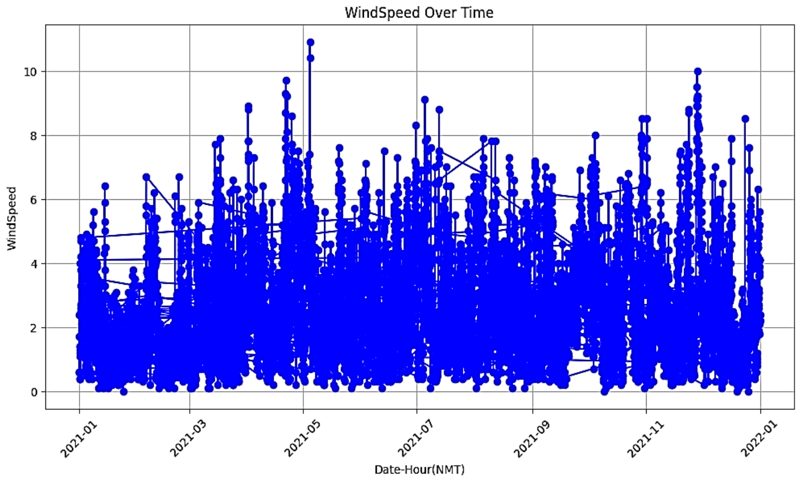

Figure 4. Wind speed vs time

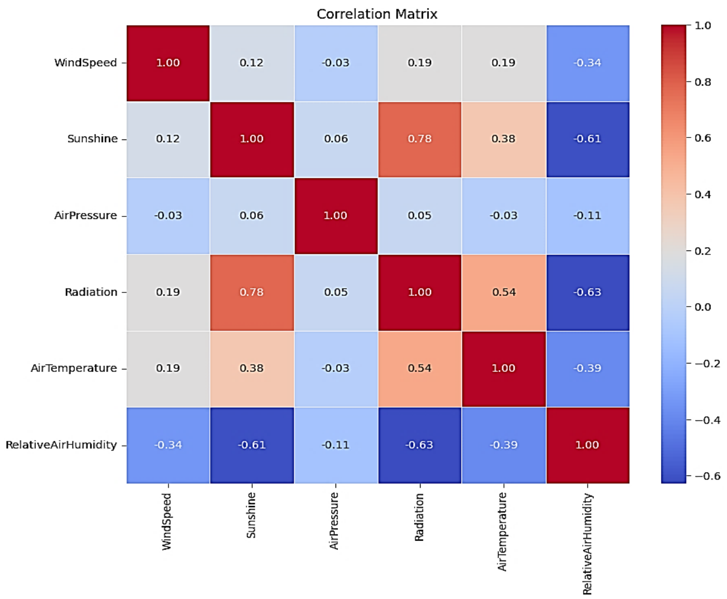

Figure 5. Correlation matrix of various factors

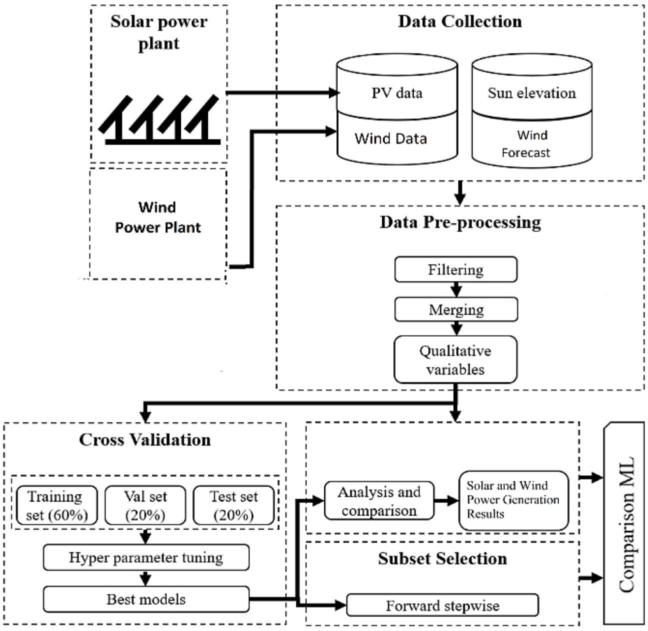

Figure 6. Application of machine learning in renewable energy prediction

2. METHODOLOGY

2.1. Data collection

The study utilizes a dataset comprising hourly observations of key environmental and meteorological factors along with corresponding system production data. The 'Date-Hour (NMT)' column, serving as the timestamp, is converted to data time format for temporal analysis. The chosen features for predicting system production comprise wind speed, sunshine, air pressure, radiation, air temperature, and relative air humidity [19], [20].

2.2. Data preprocessing

Before commencing model training, a sequence of preprocessing procedures is executed to uphold the quality and appropriateness of the dataset. These procedures incorporate, they are: i) Data cleaning: handling missing values, outliers, and any inconsistencies in the dataset; ii) Feature scaling: standardizing or normalizing numerical features to bring them to a similar scale; iii) Categorical encoding: if applicable, encoding categorical variables into numerical format for model compatibility [21], [22]; and iv) Train-test split: dividing the dataset into training and testing sets to assess model generalization.

2.3. Model selection

Four distinct machine learning algorithms are chosen for this study. They are: i) Linear regression: a straightforward and easily interpretable model that assumes a linear relationship between input features and system production; ii) Decision tree: a non-linear model that recursively splits the data based on feature thresholds to make predictions; iii) Random forest: an ensemble model that aggregates predictions from multiple decision trees, providing improved accuracy and robustness; and iv) Support vector machine (SVM): a model aiming to find a hyperplane that best separates the input space into regions corresponding to different output classes.

2.4. Model training

For each selected algorithm, the dataset is used to train the model. During the training phase, the model learns the relationships between the input features and the target variable (system production). The training process involves optimizing model parameters to minimize the difference between predicted and actual values.

2.5. Model testing and evaluation

The trained models undergo evaluation using a distinct testing dataset that hasn't been utilized during the training phase. Performance metrics such as mean squared error (MSE), mean absolute error

(MAE), and R-squared are calculated to gauge the accuracy and predictive capability of each model. These metrics offer valuable insights into the models' ability to generalize to new, unseen data [23].

2.5.1. Mean squared error

Mean squared error (MSE) stands as a commonly used metric in regression analysis, quantifying the average of the squared disparities between predicted and actual values. By accentuating larger errors, MSE offers a comprehensive evaluation of prediction accuracy. The formula for MSE entails the summation of squared errors divided by the number of observations, as in (1).

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{k}(yi - \widehat{y}i)^2 \tag{1}$$

2.5.2. R-squared error

R-squared (R^2), known as the coefficient of determination, quantifies the proportion of variance in the dependent variable elucidated by independent variables in a regression model. Serving as an indicator of model fit, it gauges how effectively the model aligns with the dataset. The R^2 is calculated as the ratio of the explained variance to the total variance, as in (2).

$$R^2 = 1 - \frac{\sum_{i=1}^{k}(yi - \widehat{y}i)^2}{\sum_{i=1}^{k}(yi - \overline{y}i)^2} \tag{2}$$

Where: n is the number of data points; y_i the actual value of the target variable for the i_{th} observation; $\widehat{y}i$ is the predicted value of the target variable for the i_{th} observation. The following measures assess the effectiveness of regression models, the following metrics were utilized for performance evaluation.

2.6. Hyper parameter tuning and visualization

To optimize the performance of each model, hyper parameter tuning may be conducted. This involves systematically adjusting the model's hyper parameters to find the configuration that results in the best predictive performance [24], [25]. The results are visualized through scatter plots, illustrating the relationship between actual and predicted system production for each model. This visual representation aids in the interpretation of model accuracy and potential areas of improvement.

3. RESULTS AND DISCUSSION

3.1. Model performance metrics

The predictive models underwent evaluation utilizing three crucial performance metrics: mean squared error (MSE), mean absolute error (MAE), and R-squared. These metrics offer insights into the accuracy and goodness of fit for each algorithm. Figure 7 illustrates the comparison between actual and predicted energy across different models.

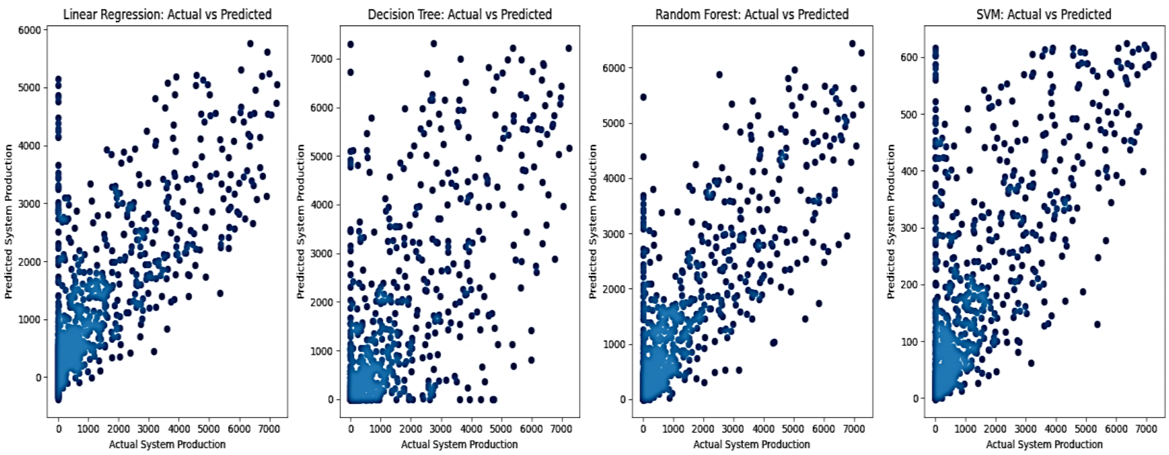

Figure 7. Actual versus predicted energy comparison for different models

The linear regression model yielded an MSE of 7.67e+05, an MAE of 461.70, and an R-squared of 0.61. These metrics indicate a moderate level of predictive accuracy, with the R-squared value suggesting that approximately 61.2% of the variance in the system production can be explained by the selected features. The decision tree model resulted in an MSE of 9.53e+05, an MAE of 409.32, and an R-squared of 0.52. While the decision tree performed reasonably well, it exhibited slightly lower predictive accuracy compared to linear regression, suggesting a need for more complex models. Figures 8(a)-8(c) shows the comparison of different Models. Table 1 shows the performance metrics comparison.

The random forest model demonstrated improved performance with an MSE of 5.54e+05, an MAE of 333.77, and an R-squared of 0.72. These metrics suggest that the ensemble nature of random forest effectively mitigated overfitting and enhanced predictive accuracy, making it a promising algorithm for system production prediction. The SVM model, despite being a powerful algorithm, displayed the lowest predictive accuracy among the models. It resulted in an MSE of 1.96e+06, an MAE of 582.10, and an R-squared of 0.01. The low R-squared value indicates poor fit to the data, suggesting that the linear SVM might not be well-suited for the complex patterns present in the dataset.

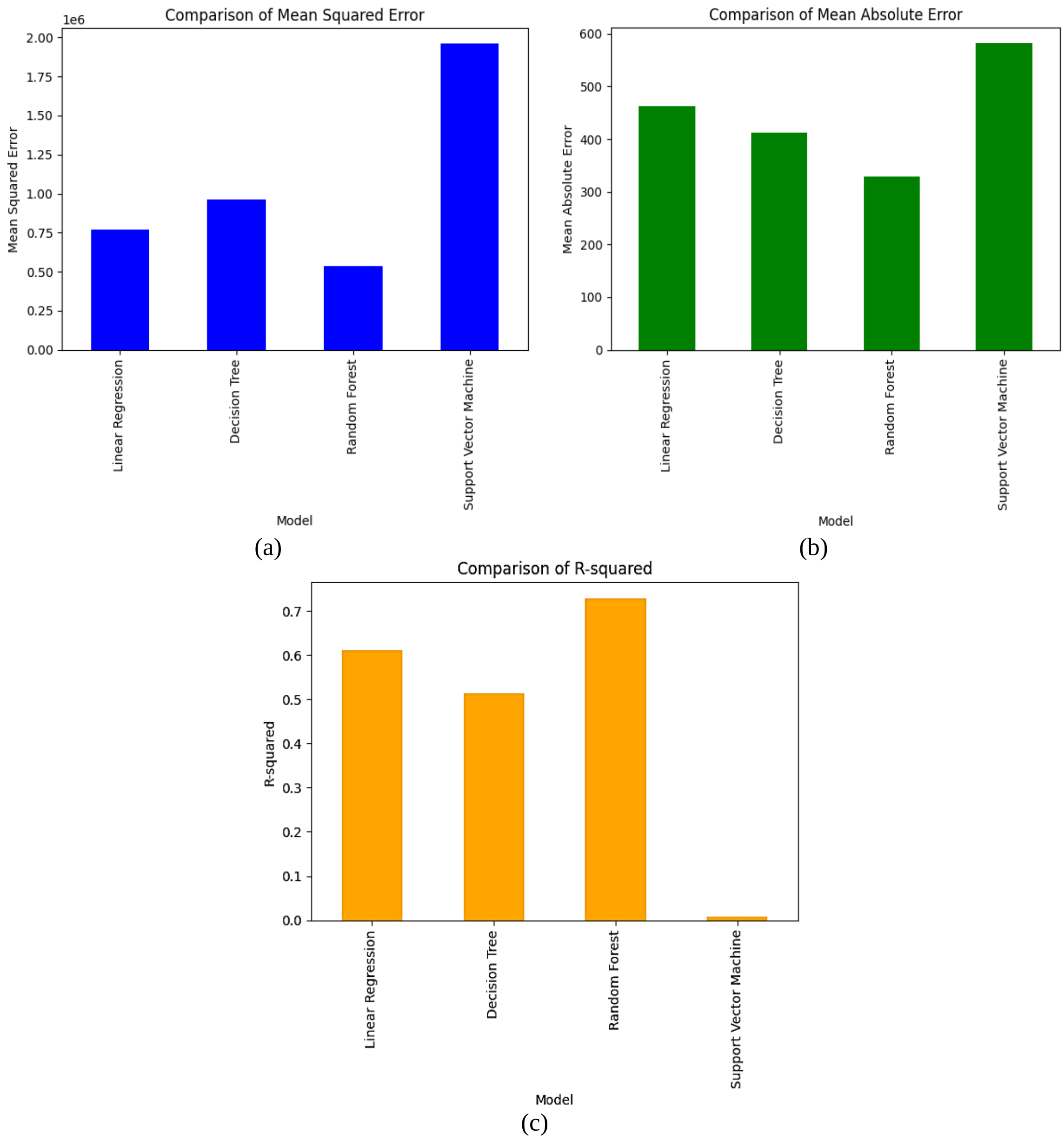

Figure 8. Comparison of different models: (a) comparison of mean squarred error, (b) mean absolute error, and (c) comparison of R-squared

Table 1. Performance metrics comparison

S. No	Model	Mean squared error	Mean absolute error	R-squared
1	Linear regression	7.6711*10^5	461.702	0.612264
2	Decision tree	9.63*10^5	412.920	0.513248
3	Random forest	5.373*10^5	328.109	0.728375
4	SVM	1.96*10^6	582.0973	0.008838

3.2. Scatter plot analysis

To visually assess the performance of each algorithm, scatter plots were generated comparing the actual system production values against the predicted values. The scatter plot for linear regression illustrates a moderate linear relationship between the actual and predicted system production values. However, some deviations from the ideal line suggest areas where the model may not capture more complex patterns in the data. The decision tree scatter plot shows a noticeable scatter, reflecting the model's tendency to capture non-linear relationships. However, the model might struggle with overfitting, as indicated by its lower R-squared value.

The scatter plot for random forest demonstrates a more cohesive alignment of points along the diagonal, indicating improved accuracy compared to the other models. The ensemble approach helps mitigate the over fitting observed in the decision tree. The SVM scatter plot reveals a dispersed pattern with no clear trend, aligning with the low R-squared value. This suggests that the linear SVM model struggles to capture the underlying patterns in the data adequately. In summary, the scatter plots provide valuable insights into the strengths and weaknesses of each algorithm, corroborating the quantitative metrics obtained. The random forest model emerges as the most promising for accurate system production prediction in this context. However, further investigations and fine-tuning of hyper parameters may be necessary for optimal performance.

4. CONCLUSION

In summary, our investigation into predicting system production in renewable energy systems has yielded valuable insights. The comparison of machine learning algorithms, including linear regression, decision tree, random forest, and support vector machine, revealed varying levels of predictive accuracy. Notably, the models demonstrated the importance of features such as wind speed, sunshine, and radiation in forecasting system production. The implications of our findings extend to the optimization of renewable energy systems, allowing for more informed decision-making in resource utilization and energy planning. Accurate predictions enable better grid management and contribute to the overall efficiency of sustainable energy production.

For practical applications, integrating the identified features into real-time monitoring systems can enhance the responsiveness of renewable energy installations. Additionally, our research highlights the need for continuous refinement of models and the exploration of advanced machine learning techniques for further improvements in prediction accuracy. Future research avenues could explore the integration of additional environmental variables, advancements in model interpretability, and the adaptation of machine learning algorithms to dynamic and evolving energy landscapes. Our work lays the foundation for ongoing efforts to enhance the reliability and efficiency of renewable energy systems through advanced predictive modeling.

REFERENCES

[1] J. Wang, J. Hu, K. Ma, and Y. Zhang, “A self-adaptive hybrid approach for wind speed forecasting,” *Renewable Energy*, vol. 78, pp. 374–385, Jun. 2015, doi: 10.1016/j.renene.2014.12.074.

[2] G. Nguyen *et al.*, “Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey,” *Artificial Intelligence Review*, vol. 52, no. 1, pp. 77–124, Jun. 2019, doi: 10.1007/s10462-018-09679-z.

[3] L. Tan, J. Han, and H. Zhang, “Ultra-short-term wind power prediction by salp swarm algorithm-based optimizing extreme learning machine,” *IEEE Access*, vol. 8, pp. 44470–44484, 2020, doi: 10.1109/ACCESS.2020.2978098.

[4] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, “Deep learning approach for intelligent intrusion detection system,” *IEEE Access*, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/ACCESS.2019.2895334.

[5] S. Wang, N. Zhang, L. Wu, and Y. Wang, “Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method,” *Renewable Energy*, vol. 94, pp. 629–636, Aug. 2016, doi: 10.1016/j.renene.2016.03.103.

[6] N. Priyadarshi, S. Padmanaban, M. S. Bhaskar, F. Blaabjerg, and J. B. Holm-Nielsen, “An improved hybrid PV-wind power system with MPPT for water pumping applications,” *International Transactions on Electrical Energy Systems*, vol. 30, no. 2, p. e12210, Feb. 2020, doi: 10.1002/2050-7038.12210.

[7] N. Lu and Y. Liu, “Application of support vector machine model in wind power prediction based on particle swarm optimization,” *Discrete & Continuous Dynamical Systems - S*, vol. 8, no. 6, pp. 1267–1276, 2015, doi: 10.3934/dcdss.2015.8.1267.

[8] I. C. Gunadin *et al.*, "Forecasting voltage collapse when large-scale wind turbines penetrated to power systems using optimally pruned extreme learning machines (OPELM) - case study: electric power system South Sulawesi-Indonesia," *Przegląd Elektrotechniczny*, vol. 98, no. 5, p. 80, May 2022, doi: 10.15199/48.2022.05.15.
[9] E. H. M. Ndiaye, A. Ndiaye, M. Faye, and S. Gueye, "Intelligent control of a photovoltaic generator for charging and discharging battery using adaptive neuro-fuzzy inference system," *International Journal of Photoenergy*, vol. 2020, pp. 1–15, Mar. 2020, doi: 10.1155/2020/8649868.
[10] Y. García-Vera, R. Dufo-López, and J. Bernal-Agustín, "Optimization of isolated hybrid microgrids with renewable energy based on different battery models and technologies," *Energies*, vol. 13, no. 3, p. 581, Jan. 2020, doi: 10.3390/en13030581.
[11] Y. Wu *et al.*, "Large scale incremental learning," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 374–382. doi: 10.1109/CVPR.2019.00046.
[12] A. Mosavi, S. Shamshirband, E. Salwana, K. Chau, and J. H. M. Tah, "Prediction of multi-inputs bubble column reactor using a novel hybrid model of computational fluid dynamics and machine learning," *Engineering Applications of Computational Fluid Mechanics*, vol. 13, no. 1, pp. 482–492, Jan. 2019, doi: 10.1080/19942060.2019.1613448.
[13] Z. Wang, K. Liu, J. Li, Y. Zhu, and Y. Zhang, "Various frameworks and libraries of machine learning and deep learning: a survey," *Archives of Computational Methods in Engineering*, vol. 31, no. 1, pp. 1–24, Jan. 2024, doi: 10.1007/s11831-018-09312-w.
[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
[15] J. R. Saura, B. R. Herraez, and A. Reyes-Menendez, "Comparing a traditional approach for financial brand communication analysis with a big data analytics technique," *IEEE Access*, vol. 7, pp. 37100–37108, 2019, doi: 10.1109/ACCESS.2019.2905301.
[16] M. N. Akhter *et al.*, "An hour-ahead PV power forecasting method based on an RNN-LSTM model for three different PV plants," *Energies*, vol. 15, no. 6, p. 2243, Mar. 2022, doi: 10.3390/en15062243.
[17] A. Awasthi *et al.*, "Review on sun tracking technology in solar PV system," *Energy Reports*, vol. 6, pp. 392–405, Nov. 2020, doi: 10.1016/j.egyr.2020.02.004.
[18] N. AL-Rousan, N. A. Mat Isa, and M. K. Mat Desa, "Efficient single and dual axis solar tracking system controllers based on adaptive neural fuzzy inference system," *Journal of King Saud University - Engineering Sciences*, vol. 32, no. 7, pp. 459–469, Nov. 2020, doi: 10.1016/j.jksues.2020.04.004.
[19] M. Kharrich *et al.*, "Optimal design of an isolated hybrid microgrid for enhanced deployment of renewable energy sources in Saudi Arabia," *Sustainability*, vol. 13, no. 9, p. 4708, Apr. 2021, doi: 10.3390/su13094708.
[20] J. A. Carballo, J. Bonilla, L. Roca, and M. Berenguel, "New low-cost solar tracking system based on open source hardware for educational purposes," *Solar Energy*, vol. 174, pp. 826–836, Nov. 2018, doi: 10.1016/j.solener.2018.09.064.
[21] J. A. Carballo, J. Bonilla, M. Berenguel, J. Fernández-Reche, and G. García, "New approach for solar tracking systems based on computer vision, low cost hardware and deep learning," *Renewable Energy*, vol. 133, pp. 1158–1166, Apr. 2019, doi: 10.1016/j.renene.2018.08.101.
[22] X. Sun *et al.*, "Optimal estimation of the PEM fuel cells applying deep belief network optimized by improved archimedes optimization algorithm," *Energy*, vol. 237, 2021, doi: 10.1016/j.energy.2021.121532.
[23] Aly, H.H. "A novel approach for harmonic tidal currents constitutions forecasting using hybrid intelligent models based on clustering methodologies," *Renewable Energy*, vol. 147, pp. 1554-1564, 2020, doi: 10.1016/j.renene.2019.09.107.
[24] X. LI and Q. ZHU, "Tax forecasting based on Adaboost algorithm and BP neural network," *Journal of Computer Applications*, vol. 32, no. 12, pp. 3558–3560, May 2013, doi: 10.3724/SP.J.1087.2012.03558.
[25] M. Aqib, R. Mehmoo, A. Alzahrani, I. Katib, A. Albeshri, and S. M. Altowaijri, "Smarter traffic prediction using big data, in-memory computing, deep learning and GPUs," *Sensors*, vol. 19, no. 9, p. 2206, 2019.

BIOGRAPHIES OF AUTHORS

Deepa Somasundaram received her B.E. from K.S.R. college of Technology, affiliated to Periyar University, in 2003, M.E. from Annamalai University in 2005. She completed her Ph.D. degree form Sathyabama university in 2013. Presently, she is working as an professor in the Department of EEE at Panimalar Engineering College, Chennai. She has published a more than 30 papers in International and National Journals. Her area of interest is power system and optimization technique. She has more than 15 years of experience in teaching field. She can be contacted at email: dee_soms123@yahoo.co.in.

R. Muthukumar obtained his Bachelor's degree in Electrical and Electronics Engineering at Coimbatore Institute of Technology, Coimbatore, M.E. in Power Systems Engineering at Government College of Technology, Coimbatore, India. He obtained Ph.D. in Anna University, Chennai in the year 2014. He is working as an Associate Professor in the Department of Electrical and Electronics Engineering at Erode Sengunthar Engineering College. He is in the field of power systems engineering. He has been in the teaching profession for more than 22 years. He has published a number of papers in National and International Journals. His main area of interest includes soft computing and optimization techniques. He can be contacted at email: muthukumarr2004@gmail.com.

N. Rajavinu received the B.E. degree in Electrical and Electronics Engineering from Noorul Islam College of Engineering, India, in 2006 and M.E. degree in Power Electronics and Drives from St. Peter's University, India, in 2014. From 2007 to 2012, she served as a Lecturer in Sakthi Engineering college, India. In July 2014, she joined as assistant professor in S.A. Engineering College, India. Her current research interests include power electronics and its applications. She can be contacted at email: rajavinun@saec.ac.in.

Kalaivani Ramaiyan obtained her B.E. degree in electrical and electronics engineering in the year 2000 from National Institute of Technology, Trichy, M.E. degree in power systems in the year 2002 from National Institute of Technology, Trichy. She obtained her Ph.D. degree in the year 2013 from Anna University, Chennai. She is currently working as professor in Rajalakshmi Engineering College in the Department of Electrical and Electronics Engineering. Her current research areas are smart grid, renewable energy, and electric vehicle. She can be contacted at email: kalaivani.r@rajalakshmi.edu.in.

P. Kavitha working as an associate professor in the Department of Electrical and Electronics Engineering R.M.K. Engineering College, has about 27 years of teaching experience. She received her B.E. degree in Electrical and Electronics Engineering with First class and M.E. degree in Control and Instrumentation with distinction from Anna University, Chennai. She has published 17 research papers in refereed international journals and various international conferences. Her areas of research include power controllers, machine design, and renewable energy systems. She can be contacted at email: pka.eie@rmkec.ac.in.