Weather prediction using random forest machine learning model

R. Meenal[1], Prawin Angel Michael[2], D. Pamela[3], E. Rajasekaran[4]

[1, 2]Department of Electrical and Electronics Engineering, Karunya Institute of Technology and Sciences,
Coimbatore, India
[3]Department of Biomedical Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India
[4]Department of Science and Humanities, VSB Engineering College, Karur, India

ABSTRACT

The complex numerical climate models pose a big challenge for scientists in weather predictions, especially for tropical system. This paper is focused on presenting the importance of weather prediction using machine learning (ML) technique. Recently many researchers recommended that the machine learning models can produce sensible weather predictions in spite of having no precise knowledge of atmospheric physics. In this work, global solar radiation (GSR) in MJ/m2/day and wind speed in m/s is predicted for Tamil Nadu, India using a random forest ML model. The random forest ML model is validated with measured wind and solar radiation data collected from IMD, Pune. The prediction results based on the random forest ML model are compared with statistical regression models and SVM ML model. Overall, random forest machine learning model has minimum error values of 0.750 MSE and R2 score of 0.97. Compared to regression models and SVM ML model, the prediction results of random forest ML model are more accurate. Thus, this study neglects the need for an expensive measuring instrument in all potential locations to acquire the solar radiation and wind speed data.

Corresponding Author:

R. Meenal
Department of Electrical and Electronics Engineering
Karunya Institute of Technology and Sciences
Coimbatore-641114, Tamil Nadu, India
Email: meenasekar5@gmail.com

1. INTRODUCTION

Weather forecasts are very essential for a safe and efficient management of the power grid. Solar and Wind energy are the main renewable energy sources owing to its simplicity of access and surplus in major parts of the world. Accurate knowledge of solar and wind data is very essential for renewable energy based applications. The solar and wind data is not available in all the locations of the worls because of high cost and difficulties in measurement techniques. Because of the data scarcity, solar radiation and wind speed prediction plays an important role for effective and proper utilization of these renewable energy sources. This study is focused on neglecting the need of expensive measuring instruments to acquire the wind speed and solar radiation data. There are many approaches available in the literature for the prediction of wind speed and radiation including the physical methods, persistence method, statistical methods, spatial correlation methods, artificial intelligence methods, hybrid methods and so on [1]-[5]. Artificial intelligence based methods are widely used for the estimation of wind speed and solar radiation [6]-[10]. The physical models are complex and time consuming. In the climate change models, there are noticeable flaws pertaining to the

statistical analysis, which tend to increase the extent of future climate change. Many researchers recommended ensemble forecasts to improve forecast accuracy of the models [11]. Sandeep [12] recommended that there is a need for emerging techniques such as machine learning [13]-[17], deep learning [18] and IoT based weather monitoring system [19]-[20].

The major research gap in weather prediction is better understanding of atmospheric science. Compared to the European weather or the North American weather, in tropical systems, such as that over India, the systems are fast changing and hence the complex numerical models pose a big challenge for our Indian scientists. The Indian grid is presently the 5th biggest in the world. Despite having the best models, the best computers, the best observation system, best instruments, and best satellites and so on, India is behind in weather prediction just because of the inherent complexity of the tropical systems. In order to increase the renewable energy generation in India, forecasting is an unavoidable task for the country's electricity grid. Also renewable energy generations from wind and solar plants are equally posturing considerable technical difficulties of grid management. According to the wire report on environment, forecast capacity of the numerical models used by the IMD framed by the US National Centres for environmental prediction is still weak. Longer the period of forecast, the more uncertainty there is. It is difficult to understand the physics and precursor conditions of extreme causing weather pattern. So accurate forecast is difficult, no matter how much computing power we put in.

There are three broad categories of solar and wind forecasting models namely physical model, statistical [21] and computational models [22]-[23]. Artificial intelligence [24]-[26], machine learning [30]-[32] and deep learning methods, such as CNN and LSTM, are widely used in many applications like load forecasting, prediction of solar radiation, prediction of heating energy consumption [33] and image classification and so on. The deep learning neural networks produce good reasonable weather forecasts in spite of having no clear knowledge of atmospheric physics [27]. Incorporation of artificial intelligence; machine learning and deep learning methods in a climate model will give a faster and more accurate model. In this work monthly global solar radiation and monthly wind speed are estimated for Tamil Nadu using random forest machine learning algorithm with commonly available meteorological input parameters. Seasonal wind resource maps and solar resource maps were developed based on the estimated GSR and wind speed data

2.   DATA SET

The input parameters for building the random forest machine learning model are maximum temperature (T_{max}), minimum temperature (T_{min}), surface pressure in kPa, percentage relative humidity (RH), months and latitude and longitude. The measured data for this work namely the monthly mean maximum temperature, minimum temperature, daily GSR in MJ/m^2/day and wind speed in m/s for different locations of India were collected from India meteorological department, Pune. Table 1 and Figure 1 present the input and output parameters of the random forest machine learning model for the prediction of wind speed and global solar radiation. The most relevant input parameters are selected using correlation analysis. The selected influencing input parameters are month, latitude and maximum temperature. For 15 Indian locations hourly global solar radiation data is collected from IMD, Pune. Daily mean and monthly mean global solar radiation data is computed from hourly data set. Monthly mean data available for few Indian locations is utlised to train the solar radiation models to predict the solar radiation for other remote solar potential Indian locations where the solar data is unavailable.

The entire dataset is divided into two sections such as training set (installation sub-data set -70%) and testing set (validation sub-data set -30%). It is further processed with the help of solar radiation and wind energy models using python programming language executed in spyder IDE with random forest and SVM machine learning algorithm. Pandas in python have libraries that are particular to handling time series data. The scikit-learn python machine learning library 'sklearn.ensemble' provides an implementation of random forest for machine learning. This library holds the built in function for the random forest regressor model. After importing the model the training data must be fitted into the model for the machine to learn about the data and to build the model. Finally a prediction with the random forest regressor model testing data value is done for checking the predicted value with the actual value of testing data. The performance of the machine learning model is evaluated using the performance metrics R^2 and MSE. R^2 corresponds to the squared correlation between the recorded solar radiation and wind speed values and the predicted values by the model. The higher the R^2 value the model will be perfect. Mean square error (MSE) is the mean of the square of the errors. Simply put, the lower the value, the model is perfect. Figure 2 shows the scatter plots presenting the comparision between the recorded and predicted solar radiation data for the selected Indian locations.

Table 1. Input and output parameters

S.N		Input Parameters
		Measured input parameters
1	T_{max}	Max. Temperature in degree celcius
2	T_{min}	Min.Temperature in degree celcius
3	S	Surface Pressure kPa
4	RH	Percentage Relative humidity %
		Geographical input parameters
5		Month Number
6		Latitude °(N) of the location
7		Longitude °(E) of the location
		Output parameters
8		Monthly Global Solar Radiation in MJ/m^2/day
9		Monthly Wind Speed in m/s

3. METHODOLOGY AND RESULTS

3.1. Random forest algorithm

In this study, random forest machine learning algorithm is utilized to predict the wind speed and global solar radiation for Tamil Nadu. Leo Breiman (2001) developed the random forest (RF) ensemble algorithm which is based on decision tree predictors 28]. It is applied for both classification and regression applications. For the analysis of large dat sets, random forest algorithm is the most suitable algorithm. Due to high-prediction accuracy, this algorithm is most approachable and presents particulars on the importance of variables for classification and regression. Unlike ANN and SVM, the training procedure for random forest algorithm is simple. The main parameter to adjust is the number of trees. A quick training process is performed compared to artificial intelligence and support vector machine based models. As this algorithm involves rule-based approach, normalizing of data is not necessary.

3.2. Wind resource seasonal mapping using random forest algorithm

The wind speed prediction is very important for integration in to the electricity grid and also helps the power system operators decrease the possibility of unreliability of electricity supply. The commonly used meteorological inputs for the prediction of wind speed are relative humidity, temperature, wind direction, air pressure and so on. The wind speed data was predicted for some cities of Tamil Nadu such as Chennai, Coimbatore, Cuddalore, Dindigul, Erode, Madurai, Pondicherry, Salem, Thanjavur, Tirunelveli, Trichy, Tuticorn using random forest algorithm. Table 2 shows the predicted monthly wind speed values in m/s for the selected cities of Tamil Nadu state. Table 3 summarizes the performance metrics of random forest ML algorithm in predicting monthly wind speed in m/s. With the predicted wind speed data seasonal wind resource mapping was developed. Figure 3 shows the seasonal wind resource maps for the following Seasons such as winter, summer, south west monsoon and north east monsoon. January and February are winter season. March, April, May and June are the summer season. July, August and September are the month of South West Monsoon. October, November and December are the month of North East Monsoon.

3.3. Solar resource maps

The global solar radiation is predicted for Tamil Nadu cities using random forest machine learning model. Random forest machine learning model is compared with the conventional temperature based regression models. Table 4 summarizes the predicted global solar radiation data (MJ/m^2/day) of selected cities of Tamil Nadu state using the random forest machine learning algorithm. Table 5 presents the comparison of performance metrics between temperatures based regression model and random forest machine learning model. Overall, random forest machine learning model has minimum error values and the high correlation coefficient value. Hence, random forest model developed is utilized to predict the solar potential for Tamil Nadu. Figure 4 presents the Tamil Nadu district map with the predicted annual global solar radiation in MJ//m^2/day. Figure 5 shows the Tamil Nadu district solar resource map. From the graph, it is noticed that very good solar potential is available in Tamil Nadu.
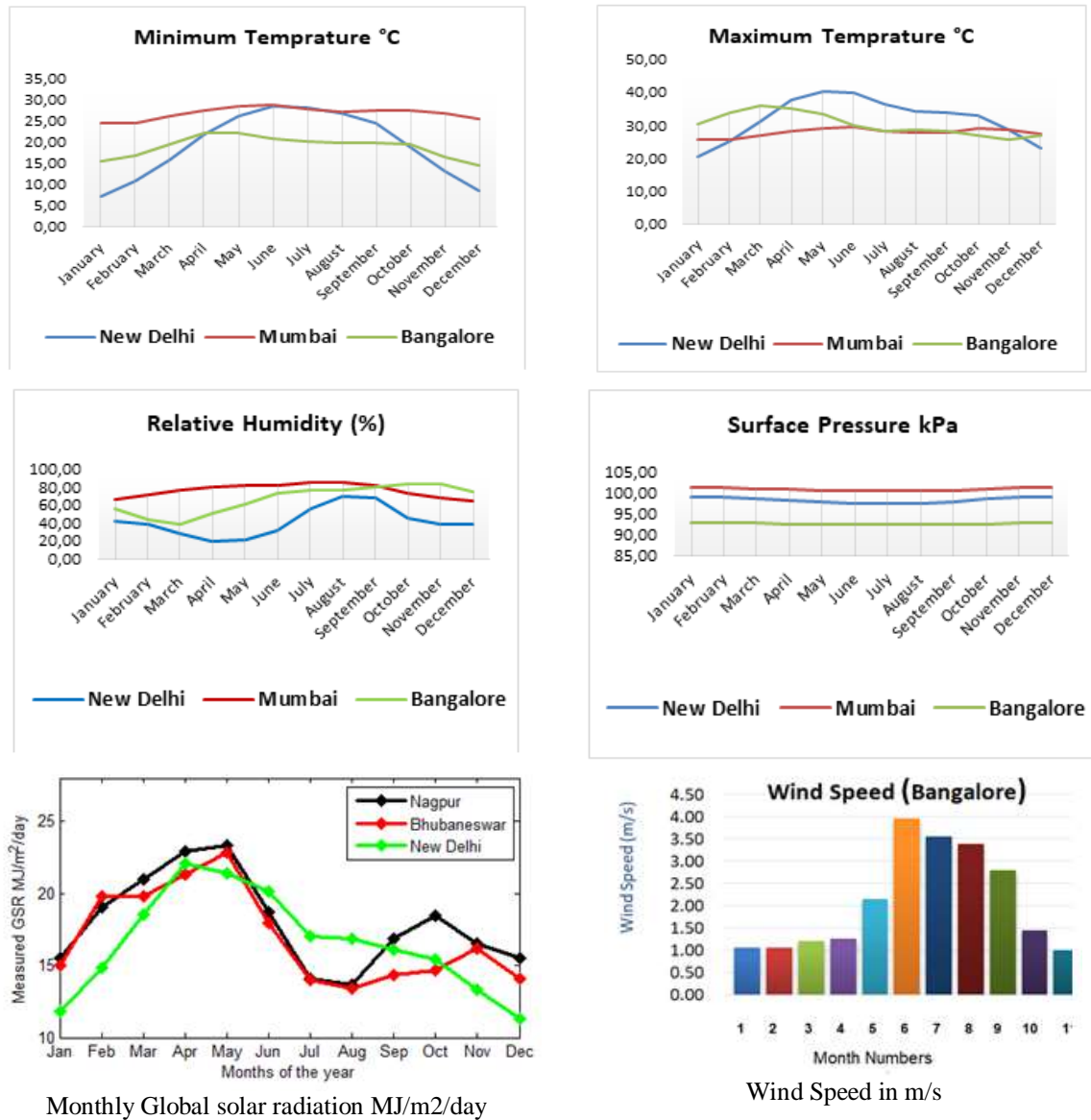
Figure 1. Input and output parameters collected from IMD, Pune for the selected Indian cities to build the SVM and random forest machine learning models
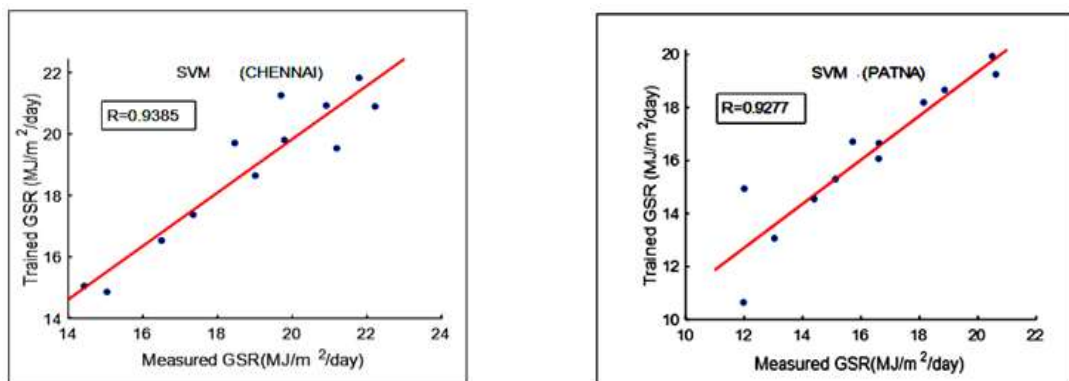

Figure 2. Scatter plots presenting the comparision between the recorded and the predicted solar radiation for the selected Indian Cities using SVM ML method

Table 2. Predicted monthly wind speed in m/s using random forest algorithm

Location	January	February	March	April	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
Chennai	5.61	4.97	4.63	4.75	5.39	5.79	5.31	4.86	4.33	4.60	5.44	5.45
Coimbatore	3.29	3.81	4.20	4.40	3.90	5.92	5.82	3.29	2.91	2.97	3.20	3.06
Cuddalore	5.39	5.02	4.59	4.81	5.71	6.37	6.30	5.92	5.21	4.62	5.63	6.55
Dindigul	4.56	3.91	4.91	3.45	4.48	5.58	6.29	6.44	6.37	5.03	4.92	5.28
Erode	5.40	4.82	3.30	4.26	5.70	6.36	6.44	6.20	5.13	4.84	5.39	5.73
Madurai	4.56	4.26	3.61	2.79	5.16	5.80	5.86	5.18	4.55	4.40	4.08	4.21
Pondycherry	5.40	4.64	3.11	4.07	5.88	6.45	6.59	6.30	5.53	4.86	5.33	5.63
Salem	5.48	4.78	3.39	4.55	5.94	6.59	6.60	6.40	5.50	4.79	5.30	6.00
Thanjavur	5.54	4.79	3.36	4.25	5.73	6.26	6.23	5.98	5.17	4.84	5.24	5.98
Tirunelveli	5.38	5.00	3.66	4.70	5.66	6.63	6.74	6.53	4.82	4.51	5.50	5.79
Trichy	5.16	4.92	3.36	3.57	5.59	6.41	6.73	6.41	5.53	5.17	5.43	5.58
Tuticorin	5.49	4.97	3.33	4.76	5.27	6.06	6.24	6.15	4.84	4.83	5.34	5.89
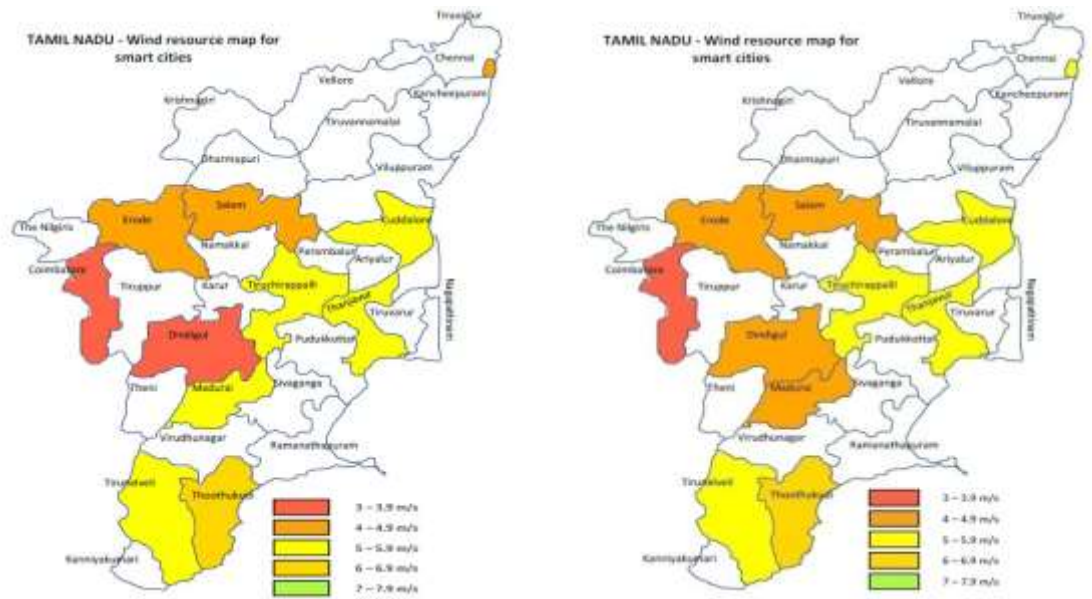
Seasonal map - Winter Season(January– Feb.)   Seasonal map - Summer Season(March– June)

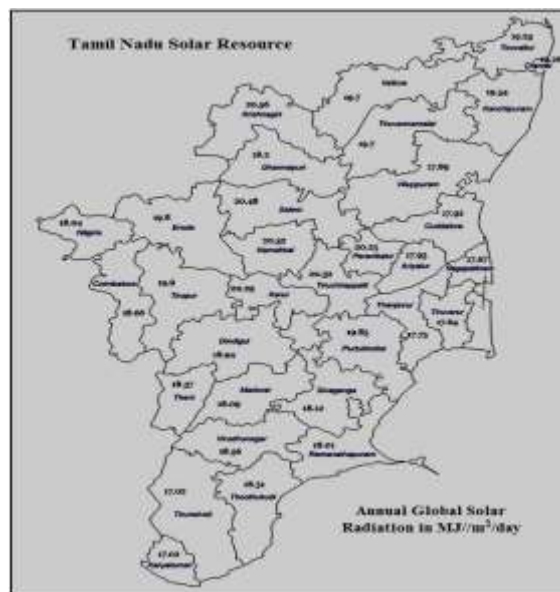Figure 3. Seasonal wind resource maps for selected cities of Tamil Nadu

Figure 4. Tamil nadu district map with the predicted annual GSR in MJ//m2/day
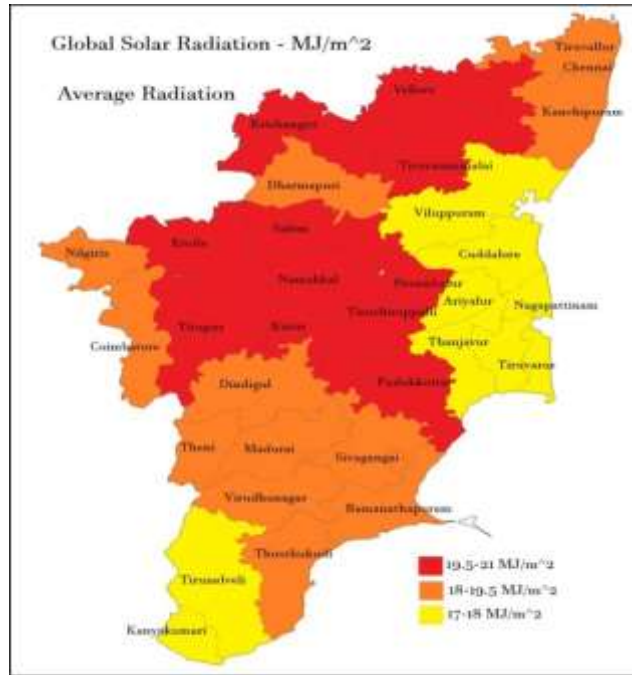
Figure 5.Tamil Nadu district solar resource map with the predicted annual GSR in MJ//m2/day

Table 3. Performance metrics of random forest ML algorithm in predicting monthly wind speed in m/s

Location	R^2 Score	Mean Squared Error
Chennai	0.93	1.03
Cuddalore	0.93	0.91
Tirunelveli	0.81	1.05
Coimbatore	0.63	0.94
Dindugul	0.75	0.89
Madurai	0.77	0.38
Pondycherry	0.61	0.93
Tutucorin	0.87	1.20

Table 4. Predicted global solar radiation data for Tamil Nadu, India using random forest machine learning algorithm

TamilNadu Cities/ Month	Chennai	Coimbatore	Cuddalore	Dindigul	Erode	Madurai	Salem	Thanjavur	Tirunelveli	Trichy
January	18.9	21.0	17.5	20.4	21.1	17.7	21.0	17.4	17.6	20.7
February	21.5	22.4	19.0	22.8	22.5	20.7	23.0	18.9	20.2	23.1
March	22.8	22.5	21.0	22.9	22.7	22.3	23.1	21.0	20.5	23.2
April	22.4	21.7	20.8	21.8	22.8	20.4	23.3	20.2	18.1	23.3
May	22.0	20.2	19.8	20.1	21.3	17.7	23.1	18.4	16.9	21.5
June	20.2	15.9	18.2	17.1	18.8	17.4	20.4	18.0	15.4	20.3
Jul	18.5	14.9	17.2	16.6	18.0	17.4	19.4	17.3	14.5	19.5
August	18.7	15.1	17.8	17.0	18.6	17.8	19.8	17.7	16.4	19.9
September	18.5	17.4	17.5	18.4	19.2	17.5	19.7	17.2	16.6	19.6
October	16.3	17.4	16.2	16.6	18.2	16.4	18.2	16.5	16.3	18.0
November	15.7	17.4	15.6	16.9	17.4	15.9	17.2	15.9	15.9	17.4
December	15.8	18.2	14.5	16.9	17.5	16.1	17.4	14.8	16.1	17.3
Annual GSR	18.9	18.6	17.9	18.9	19.8	18.1	20.5	17.7	17.0	20.3
Meteonorm Data		19.9	20.1	19.8	20.3	20.1	20.1	19.5	19.8	20.0
NASA Data		17.9	19.9	20.1	19.8	20.3	20.1	20.1	19.5	19.8

Table 5. Comparison of performance metrics between temperatures based solar radiation regression model and random forest machine learning model (Chennai)

Source	Name of the model	R^2 Score	Mean Squared Error
Chen *et al.* Model [29]	$\dfrac{H}{H_0} = a(T_{max} - T_{min})^{0.5} + b$	0.9348	1.757
	$\dfrac{H}{H_0} = a \ln(T_{max} - T_{min}) + b$	0.9343	1.750
Bristow and Campbell [29]	$\dfrac{H}{H_0} = a(1 - \exp(-b\,\Delta T)^c)$	0.9345	1.758
Current Studty	Support Vector Machine learning model	0.9385	0.867
Current Studty	Random Forest Machine learning model	0.9714	0.750

4. CONCLUSION

In this study, global solar radiation and wind speed is predicted for Tamil Nadu using random forest machine learning model. The random forest model is validated with measured data collected from IMD, Pune. Compared to regression models the randon forest machine learning model results are more accurate. Seasonal wind resource and annual solar resource mapping was developed for Tamil Nadu using the predicted wind speed and GSR values. Thus this study neglects the need of expensive measuring instrument to acquire the wind speed and solar radiation data. The research gap infered in climate science could be overcome by the machine learning techniques as discussed in this paper. Thus the random forest machine learning model produced good reasonable prediction results inspite of having no precise knowledge of atmospheric physics.

ACKNOWLEDGEMENTS

REFERENCES

[1] Wen-Yeau Chang, "A literature review of wind forecasting methods," *Journal of Power and Energy Engineering*, vol. 2, pp. 161-168, 2014, doi: 10.4236/jpee.2014.24023.

[2] Rahul Sharma and Diksha Singh,"A review of wind power and wind speed forecasting," *Journal of Engineering Research and Application*, vol. 8, pp. 1-9, 2018.

[3] Saurabh S. Soman, *et al.*, "A review of wind power and wind speed forecasting methods with different time horizons," *North American Power Symposium*, pp. 1-8, 26-28 Sep 2010, doi: 10.1109/NAPS.2010.5619586.

[4] R. Meenal, *et al.*, "Review on mathematical models for the prediction of global solar radiation," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 15, no. 1, pp. 54-59, 2019, doi: 10.11591/ijeecs.v15.i1.pp56-61.

[5] Sani Salisu, *et al.*, "A Wavelet Based Solar Radiation Prediction in Nigeria Using Adaptive Neuro-Fuzzy Approach," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 12, no. 3, pp. 907-915, December 2018, doi: 10.11591/ijeecs.v12.i3.pp907-915.

[6] MakarandA. *et al.*, "Wind speed prediction using statistical regression and neural network," *Journal of Earth System Science,* vol. 117, pp. 457-463, 2008, doi: 10.1007/s12040-008-0045-7.

[7] Salisu Muhammad Lawan, *et al.*, "Wind speed prediction in non-monitored areas based on topographic radial basis neural network (T-RBNN)," *IOP Conference Series: Earth and Environmental Science*, vol. 168, pp. 012012, 2018.

[8] Cabral Lima, *et al.*, "A model to forecast wind speed through singular spectrum analysis and artificial neural networks," *International Conference on Artificial Intelligence*, Las Vegas, pp. 235-240, 2017.

[9] R.Meenal, *et al.*, "Sensitivity analysis based artificial neural network approach for global solar radiation prediction in India," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 1, pp. 31-38, 2020, doi: 10.11591/ijeecs.v20.i1.pp31-38.

[10] R.Meenal and A.I.Selvakumar, "Review on artificial neural network based solar radiation prediction," *International Conference on Communication and Electronics System*, Coimbatore, pp. 302-305,15-16 Oct 2018, doi: 10.1109/CESYS.2017.8321285.

[11] Roberto Buizza, et.al, "A Comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems", *Monthly Weather Review,* vol. 133, No. 5, pp. 1076-1097, 2005, doi: 10.1175/MWR2905.1.

[12] SandeepPattnaik, "Weather forecasting in India: Recent developments," *Mausam,* vol. 70, no. 3, pp. 453-464, 2019.

[13] G.Santamaría-Bon, *et al.*, "Wind speed forecasting for wind farms: A method based on support vector regression," *Renewable Energy*, vol. 85, pp. 790-809, 2015, doi: 10.1016/j.renene.2015.07.004.

[14] Bhuiyan MdAbul Ehsan, et.al., "Advanced wind speed prediction using convective weather variables through machine learning," *Applied Computing and Geosciences*, vol. 1, pp. 100002, 2019, doi: 10.1016/j.acags.2019.100002.

[15] Jonathan A, *et al*., "Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data," *Journal of Advances in ModelingEarth Systems*, vol. 11, pp. 2680-2693, 2019, doi: 10.1029/2019MS001705.

[16] R.Akash, *et al*., "Machine learning based univariate models for long term wind speed forecasting," *International Conference on Inventive Computational Technologies, Coimbatore*, pp. 779-784, 26-28 Feb 2020, doi: 10.1109/ICICT48043.2020.9112534.

[17] R.Akash, *et al*., "Day-ahead wind power forecasting using machine learning algorithms," *Advances in Intelligent Systems and Computing,* vol. 1227, pp. 329-341, 2021, doi: 10.1007/978-981-15-6876-3_25.

[18] Chujie Tian, *et al*., "A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network," *Energies,* vol. 11, no. 12, pp. 3493, 2018, doi: 10.3390/en11123493.

[19] Atul Kulkarni and Debajyoti Mukhopadhyay, "Internet of Things Based Weather Forecast Monitoring System," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 9, no. 3, pp. 555-557, March 2018, doi: 10.11591/ijeecs.v9.i3.pp555-557.

[20] Dwi Susanto, *et al*., "Parallel Processing Implementation on Weather Monitoring System for Agriculture," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 6, no. 3, pp. 682-687, June 2017, doi: 10.11591/ijeecs.v6.i3.pp682-687.

[21] R.Meenal and A. Immanuel Selvakumar, "Temperature based radiation models for the estimation of global solar radiation at horizontal surface in India," *Indian Journal of Science and Technology*, vol. 9, no. 46, pp. 101922, 2016, doi: 10.17485/ijst/2016/v9i46/101922.

[22] R.Meenal, *et al*., "Solar mapping of India using support vector machine", *Journal of Physics: Conference Series,* vol. 1142, pp. 012010.

[23] R. Meenal and A. Immanuel Selvakumar, "Assessment of solar energy potential of smart cities of Tamil Nadu using machine learning with big data," *Advances in Big Data and Cloud Computing*, vol. 750, pp. 27-36, 2019, doi: 10.1007/978-981-13-1882-5_3.

[24] Amar Choudhary, *et al*., "Artificial neural networks based solar radiation estimation using backpropagation algorithm," *International Journal of Renewable Energy Research*, vol. 10, no. 4, pp. 1566-1575, 2020.

[25] Bao The Nguyen, *et al*., "Model to generate daily and hourly solar radiation sequences for tropical climates," *International Journal of Renewable Energy Research*, vol. 10, no.4, pp. 1708-1724, 2020.

[26] G.Soufiane, *et al*., "Solar power time series prediction usin g wavelet analysis," *International Journal of Renewable Energy Research*, vol. 10, no. 4, pp. 1764-1773, 2020.

[27] Fei Wang, *et al.,* "Wavelet decomposition and convolutional LSTM networks based improved deep learning model for solar irradiance forecasting," Applied *Science*, vol. 8, no. 1286, pp. 1-30, 2018, doi: 10.3390/app8081286.

[28] Breiman, L, "Random forests." *J. Mach. Learn.*, vol. 45, no. 1, pp.5-32 , 2001, doi.org/10.1023/A:1010933404324.

[29] R. Meenal and A. Immanuel Selvakumar, "Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters," *Renewable Energy*, vol. 121, pp. 324-343, 2018, doi: 10.1016/j.renene.2017.12.005.

[30] Waqas Khan, P., Byun, Y. C., Lee, S. J., and Park, N, "Machine learning based hybrid system for imputation and efficient energy demand forecasting. *Energies*, vol. 13, no. 11, p. 2681, 2020, doi: 10.3390/en13112681.

[31] Khan, P. W., " Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources, *Energies*, vol. 13, no. 18, 4870, doi: 10.3390/en13184870.

[32] Khan, P. W, and Byun, Y. C, "Genetic Algorithm Based Optimized Feature Engineering and Hybrid Machine Learning for Effective Energy Consumption Prediction," *IEEE Access*, 8, 196274-196286, doi: 10.1109/ACCESS.2020.3034101.

[33] Lee, S., "Deep Neural Network Approach for Prediction of Heating Energy Consumption in Old Houses," *Energies*, vol. 14, no. 1, p. 122, doi: 10.3390/en14010122.