Survey on Forecasting Renewable Energy Generation(Solar,Wind) using ML & DL techniques

Suresh Kumar S[1], Vikashini S[2], Vijay Kumar V[3]
[1]PROFESSOR, Department of Computer Science and Engineering
[2,3]UG Scholar, Department of Artificial Intelligence and Data Science(AI&DS)
Rajalakshmi Engineering College(REC), Chennai, Tamil Nadu, India.
221801062@rajalakshmi.edu.in

***Abstract*—Renewable Energy plays a vital role in our daily life by saving the environment as it reduces the usage of non-renewable resources for energy production which in turn brings down carbon emission. To increase renewable energy generation, an accurate predictive model is required. Traditional forecasting models have some drawbacks, which includes inefficiency in acknowledging the nonlinear relationship and the seasonality pattern over the period of time. The following constraints can be removed by applying the algorithms of ML and DL. Both of the methodologies help us predict renewable energy generation with accuracy. This survey is going to introduce a systematic review and detailed view of the researches made on renewable energy prediction, focusing on the prediction of solar and wind energies, using ML and DL techniques like SVM, RF, LR, MLP, DT, Gradient Boost, LSTM, and hybrid models.**

Keywords— Renewable Energy, Wind, Solar, Predictive model, Machine learning, Deep Learning, hybrid models, Seasonality

I. INTRODUCTION

In recent years , the demand for renewable energy sources are increasing particularly for solar and wind power, which are leading contributors to renewable energy production. However, in the earlier days, we were mostly dependent on the fuel from non-renewable resources, which was essentially fossil fuels in the form of coal, natural gas, and petroleum, for our energy requirements. Such resources are scarce and lead to extreme consequences on air quality, water sources, and outcome in greenhouse gas emissions that degrade our environment. These problems have been countered by another eco-friendly alternative that is generating power from renewable sources of energy which can be replenished in particular. These include solar and wind energy. The statistics show an exponential growth in both solar and wind energy production from 2010 to 2023. As shown in the bar chart in Fig.1 [data source: OurWorldInData], solar energy production globally increased from just 30 TWh in 2010 to 1,500 TWh by 2023. Wind energy also has a steady growth, rising from 340 TWh in 2010 to over 2,000 TWh in 2023. This growth indicates that our necessity for renewable energy is highly increasing.

However, the variable nature of these resources poses unique challenges to managing as well as operating a power grid and an energy market. These challenges can be overcome by the precise predictions of renewable energy generation that can enable a full switch to clean energy. We began with Statistical techniques like ARIMA, Seasonal ARIMA(SARIMA) for forecasting. These methodologies were established already for the analysis and especially for forecasting the time series data. However, these methods were not efficient as they have envisioned where it was applied on complex and nonlinear behavior of renewable energy systems where these traditional techniques believe in a linear relationship among variables. We then found that these limitations may be dealt with by machine learning and deep learning techniques, so we had focused on those techniques to make an accurate forecasting.

II. OVERVIEW OF ENERGY FORECASTING

Energy forecasting is essential for managing power supply and demand particularly with the growing integration of green energy resources that are continuously variable and hard to predict . We provide a detailed view of techniques that are used for forecasting renewable energy generation instead of traditional methods these days in this section.

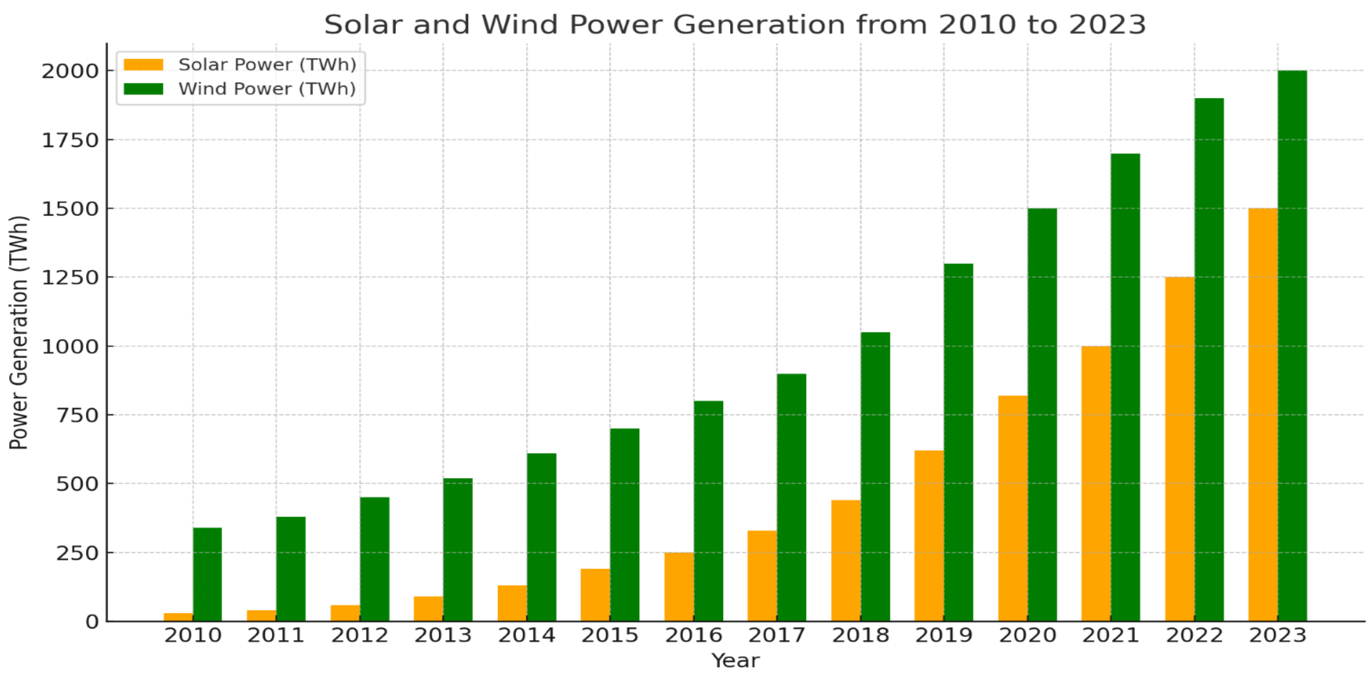
A. Machine Learning(ML)

Linear Regression Technique (LR)

It is one of the machine learning techniques that depends on predictions, particularly in conditions when the dependent and independent variables have a linear relationship among them. finds the best-fit straight line through the points in data so that future trends can be anticipated based on historical trends. Although it is a very primitive technique, it has acted as a benchmark against which more sophisticated forecasting techniques are measured even until today; however, fails while dealing with nonlinear or complex data produced by dynamic systems such as renewable energy generation since it assumes a linear as well as time-invariant relationship among the variables. Applications to renewable energy include simple forecasting models, where, for example, temperature, wind speed, or sun irradiation is directly proportional to the rate of energy production. Such models are useful but very limited in their application to real-world events. The general form of LR is given by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon$$

- The dependent variable is y (e.g., energy generated).
- independent variables are x1,x2,x3,…,xn (e.g., temperature, wind speed, solar radiation).
- The intercept is β0 (bias term)..
- β1,β2,…,βn the coefficients corresponding to the independent variables.
- ϵ denotes the error term.



The value of the residual sum of squares (RSS) should be optimized , mathematically expressed as follows.

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Where ŷ is an estimation of every observation included in the sample

Decision Tree(DT)

DT is mostly used ML method for forecasting, particularly due to their efficiency in handling intricate and non-linear relationships between variables. In renewable energy forecasting, DTs can capture relationships and patterns between environmental conditions effectively which act as a key factor for predicting the power genes, with each node representing a decision based on an input feature (e.g., wind speed, temperature). A Decision Tree works by dividing the data into nodes and branches. The tree continues to split the data until it reaches the leaf nodes, where the final forecast is made. In renewable energy forecasting, DT can model how variables like wind speed, solar radiation, and temperature interact to influence energy output, which often shows nonlinear behaviors.

Decision Trees predict energy output by calculating the average target value at each leaf node. The prediction ŷ is given by

$$\hat{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

It is in the form, where ŷ denotes the forecasted value, yi represents actual values, finally n represents count of node's data points. The algorithm minimizes the value of Mean Square Error. It heps in determining the splits.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

By choosing splits , the tree will learn from the data and improve accuracy. Decision Trees have excellent capability for capturing non-linear relationships and handle complex data sets really well. However, they would tend to overfit the data if it is taken too deep, otherwise, capture the noise rather than the trend.

Support Vector Machine(SVM)

Support Vector Machine is an suitable machine learning algorithm for handling regression and also classification. In renewable energy forecasting, SVM effectively captures linear and nonlinear relationships between conditions (e.g., wind speed, solar radiation). For more complex patterns, this algorithm uses kernel function that helps in separating the data more effectively by creating a linear boundary.

In regression (SVR), the SVM finds such a function f(x) that stays within a margin ϵ from the actual data points, trying simultaneously to minimize any errors beyond this margin. This is done by balancing two things. One is to keep the boundary as flat as possible and the second one is to minimize errors that are controlled by the parameter C and also SVM is a very strong model but it may not work properly with extremely complex datasets unless the correct parameters are used.

$$\min \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{n} \xi_i \right)$$

Gradient

Gradient Boosting is a method in ensemble learning with stronger models based on aggregations of several much simpler models, which are normally decision trees. Each new model corrects the mistakes done by the preceding ones and tries to improve the overall precision. This is excellent for predicting renewable energy phenomena in capturing patterns of variables such as wind speed and temperature.

The basic principle of Gradient Boosting is to iteratively add trees, which could minimize MSE. At each step, model gets corrected for mistakes of the previous trees, and if every update at each step is represented by

$$\hat{y}_t = \hat{y}_{t-1} + \eta \cdot h_t(x)$$

Where $\hat{y}_t$ is the updated prediction, $h_t(x)$ denotes current prediction from the current tree, and η denotes rate of learning controlling the contribution from each tree. Although very powerful and effective, Gradient Boosting is slow and likely to overfit unless careful tuning is applied.

B.Deep Learning Methods

Multi-Layer Perceptron (MLP)

MLP is one of feedforward ANN's which contains several layers of neurons. It is suitable for handling problems on regression and classification including renewable energy forecasting, where it can handle complex relationships among conditions like temperature, wind speed, and solar radiation.

It processes the input data by passing them through neurons of multiple layers. Neurons present in the model compute weighted sums and then use any non-linear activation function (such as ReLU or sigmoid) to produce a prediction. The goal is to increase accuracy, with backpropagation used to adjust the neuron weights during training. The output of an MLP is given as

$$y = f(W \cdot X + b)$$

In this case, X denotes input data here, W denotes the matrix of weights, b denotes bias term, and f denotes the activation function. While MLPs are highly adaptable and can learn intricate patterns from complex datasets, for effective training they typically need vast amounts of data. Additionally, without careful training and regularization, they can be prone to overfitting.

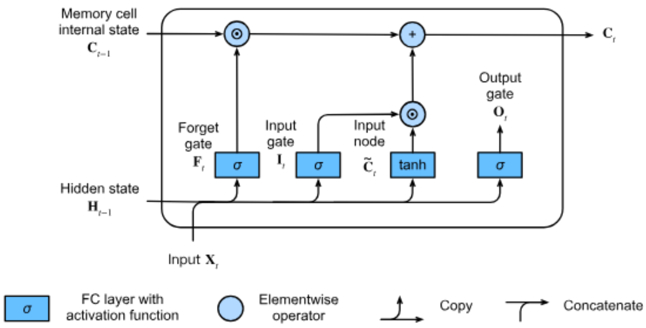
Long Short-Term Memory (LSTM)

It can efficiently handle time-series data, making it particularly useful for forecasting in renewable energy applications. LSTMs are designed to capture long-term relationships by using a set of gates given at Fig.2. These mechanisms manage how information moves through the network, allowing it to retain essential details.

1. Forget Gate: It determines which details that come from the previous cell state needed to be removed. It looks at the present input and last hidden state to produce a value between 0 and 1. Near to 0 means forget the information, while near to 1 means keep it. This helps the LSTM forget irrelevant data from past sequences, preventing information overload.

2. Input Gate: It decides the new details that are needed to be added to the cell state. This evaluates and makes decisions on input parts that are important enough to be remembered. This gate helps in updating the memory of the LSTM with new, relevant data while ignoring less significant inputs.

3. Output Gate: It manages which of the cell state's parts are forwarded as output and what continues to the next step. By analyzing the input and hidden state, it determines which information should be transferred to the following layer or time step. This selective process enables the LSTM to retain and pass on only the most relevant information, making it particularly useful for prediction.



The core of an LSTM cell is the following set of equations

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

$$h_t = o_t * \tanh(C_t)$$

Where f_t denotes forget gate, i_t denotes input gate and C_t denotes cell state, and h_t represents hidden state. The gates allow the neural network to regulate which information should be remembered or forgotten, which makes it highly efficient in finding long-term patterns such as predicting energy generation over time. While LSTMs are powerful, they can be challenging to train.

III. RELATED WORKS

[1] In the study named Renewable Energy Prediction through Machine Learning Algorithms , the authors applied various ML techniques such as SVMs, Linear Regression (LR),Random Forest and Neural Networks (NN) to predict solar and wind energy generation.They collected data from a solar plant and meteorological stations in Aguascalientes, Mexico. The dataset included speed of wind, direction of wind, humidity temperature and solar irradiance. These were used as input features in different ML models to predict energy generation. The evaluation was based on errors such as MAE,MSE.

In this study Random Forest algorithm outperformed the others, achieving a 99.5% accuracy, with the lowest error values. The exact error rates were not disclosed but the precision was significantly high in comparison to other models like SVM and NNM.The Random Forest model showed the best capability in predicting renewable energy generation compared to other algorithms.Conclusion is that Random Forest among all other techniques considered could provide high precision in renewable energy forecasting, reducing the reliance on fossil fuels and improving efficiency in energy systems.

[2] The paper titled Application of Support Vector Machine Models for Forecasting Solar and Wind Energy Resources: A Review is a comprehensive review whose focus is the usage of SVM for predicting renewable energy particularly solar and wind. The review analyzed a variety of SVM models, including hybrid models that combine SVM with Particle Swarm Optimization (PSO), Genetic Algorithms (GA).The paper reviews numerous studies that employed SVM for forecasting solar irradiance, wind speed, and power generation. It critically evaluated the efficiency in many SVM kernels like Radial Basis Function (RBF) and Polynomial and also hybrid SVM models for improving prediction accuracy.

The SVM models consistently outperformed traditional empirical models, with error rates significantly lower than other machine learning methods. For example, a hybrid SVM model with PSO optimization achieved a Root MSE of 0.74 for wind energy predictions. The study also compared results across solar and wind predictions, where hybrid SVM models outshone standalone SVM with better accuracy and lower errors, often by 10-15%.The paper concluded that SVM-based models are highly effective for renewable energy forecasting, especially when combined with optimization techniques like PSO and GA. These hybrid models reduce prediction errors and handle complex relationships efficiently, which makes them apt to make both short- and long-term forecasts. Future work should explore more hybridization with other AI techniques to enhance performance further.

[3] In the paper titled Optimized Long Short-Term Memory with Rough Set for Sustainable Forecasting Renewable Energy Generation the authors proposed a hybrid DL model that uses both LSTM networks and Rough Set-based Nutcracker Optimization Algorithm (NOA) for renewable energy forecasting.The model was designed to handle the complexity of energy data. For feature selection, The authors applied rough set theory to optimize input data before passing it into the LSTM model for time-series forecasting. The optimization algorithm (NOA) further improved the feature selection process by reducing redundancy and noise in the data.

The model got a better outcome, whose Root Mean Square Error (RMSE)value is 4.2113, MAEvalue is 2.835 and R^2 score is 0.96, and . These metrics indicate highly accurate predictions, significantly outperforming traditional statistical models like ARIMA and Holt-Winters. The proposed LSTM model also showed improved performance when compared to simpler ML models. The integration of LSTM with the Rough Set-based NOA proved to be a powerful approach for renewable energy forecasting.The results suggest that deep learning models, particularly LSTM with optimized feature selection, can provide highly accurate predictions and could be applied to various renewable energy sources, thus contributing to better energy management.

[4] A Comparative Study on Solar Power Forecasting Using Ensemble Learning explored techniques for predicting solar power. Techniques such as SVR, LR,Gradient Boosting, RF and LSTM were evaluated for their effectiveness in predicting solar irradiance. The dataset used recorded solar power every five minutes over four days. Preprocessing steps such as scaling and outlier removal were applied before training the models. The results showed that SVR and Gradient Boosting were the best-performing models, with GB showing more stable outputs while SVR had slight output variability. While LSTM performed well, the study concluded that ML techniques, particularly SVR and GB, were more suitable for solar power forecasting, as DL models struggled with seasonal input variations.

[5] In the work titled Renewable Energy Power Generation Forecasting Using Deep Learning Method, LSTM neural network to predict solar power generation. The study utilized various weather-related inputs and other meteorological data, to improve forecasting accuracy. Their primary focus was on addressing the complexity of energy data associated with renewable sources, such as photovoltaic (PV) systems, which are often influenced by fluctuating weather conditions. By incorporating LSTM, a deep learning model well-suited for sequential data, the study demonstrated a marked improvement in handling large, nonlinear datasets over more traditional machine learning models. The LSTM model excelled at capturing temporal dependencies within the data, leading to significant gains in accuracy. The authors noted that the error rates achieved with LSTM were substantially lower than those of competing methods, reinforcing the model's utility for efficient energy management, especially within household-level smart grids. This application could help optimize energy usage and improve grid reliability by better forecasting renewable energy availability

[6] In the study Predicting Power Generation from Renewable Energy Sources Using ANN uses of Artificial Neural Networks (ANN) for predicting energy output from photovoltaic systems. The researchers leveraged a variety of meteorological data, to train the ANN model. The inherent stochastic nature of renewable energy sources, which often results in inconsistent power generation, was a primary challenge addressed in the study. To overcome this, the authors integrated an Adaptive Neuro-Fuzzy Inference System (ANFIS) into MATLAB, allowing the model to adapt to fluctuations in the data more effectively. This hybrid model was particularly successful in reducing the forecasting error to as low as 1.5%, a notable achievement that provided highly reliable day-ahead energy predictions for grid operators. By improving the precision of forecasts, the study contributed to the overall stability and balance of energy grids, making the approach valuable for integrating renewable sources into existing energy networks.

[7]In Deep Learning Approach for Wind Energy Time-Series Forecasting with Environmental Factors, the authors introduced a Temporal Convolutional Network (TCN) as the primary model for forecasting wind energy based on time-series data and factors like temperature, wind speed, and humidity. They compared TCN with other models like LSTM, GRU, and Neural Basis Expansion Analysis (N-Beats). Using data from power stations across Europe, the TCN showed the best results in both accuracy and speed, especially for long-term forecasts, achieving an sMAPE of 6.18, MSE of 21.9, and RMSE of 4.69. The model's ability to handle complex spatial and temporal dependencies with dilated convolutions and residual blocks made it more effective than traditional models like LSTM and GRU. The study concluded that TCN forecasts wind energy with best results and encourages its use for other renewable energy forecasting tasks

IV CONCLUSION

In summary, this survey highlights how machine learning (ML) and deep learning (DL) techniques have greatly improved accuracy in the prediction of power generated , mainly for solar and wind power. Traditional models like ARIMA and SARIMA are not good to handle the complex nonlinear nature of data, but advanced models like RF, SVM and LSTM networks excel at capturing these intricate patterns. For instance, studies show Random Forest models achieving accuracy as high as 99.5%, and hybrid models using optimization techniques further boosting performance. LSTM models, especially effective for analyzing time-series data, stood out for long-term predictions. Overall, ML and DL provide more accurate and reliable methods for forecasting renewable energy, helping to drive the shift towards cleaner energy and better grid management.

V REFERENCES

[1] L. F. J. Alvarez, S. R. González, A. D. López, D. A. H. Delgado, R. Espinosa, and S. Gutiérrez, "Renewable Energy Prediction through Machine Learning Algorithms," 2020 IEEE Xplore, pp. 1–6, 2020. doi: 10.1109/978-1-7281-9365-6/20.

[2] A. Zendehboudi, M. A. Baseer, and R. Saidur, "Application of Support Vector Machine Models for Forecasting Solar and Wind Energy Resources: A Review," Journal of Cleaner Production, vol. 199, pp. 272–285, 2018. doi: 10.1016/j.jclepro.2018.07.164.

[3] G. I. Sayed, E. I. Abd El-Latif, A. E. Hassanien, and V. Snasel, "Optimized Long Short-Term Memory with Rough Set for Sustainable Forecasting Renewable Energy Generation," Energy Reports, vol. 11, pp. 6208–6222, 2024. doi: 10.1016/j.egyr.2024.05.072.

[4]A. Khan, R. Bhatnagar, V. Masrani, and V. B. Lobo, "A Comparative Study on Solar Power Forecasting using Ensemble Learning," in Proceedings of the 4th International Conference on Trends in Electronics and Informatics (ICOEI 2020), pp. 224–230, 2020, doi: 10.1109/ICOEI.2020.9143010.

[5] D.A. Widodo, N. Iksan, and E. D. Udayanthi, "Renewable Energy Power Generation Forecasting Using Deep Learning Method," IOP Conference Series: Earth and Environmental Science, vol. 700, no. 1, p. 012026, 2021, doi: 10.1088/1755-1315/700/1/012026.

[6] O. Rubanenko, M. Cosovic, S. L. Gundebommu, and V. Lesko, "Predicting Power Generation from Renewable Energy Sources Using ANN," in 20th International Symposium INFOTEH-JAHORINA, 2021, doi: 10.1109/INFOTEH51037.2021.9400527.

[7] U. M. Iqbal, J. Ramesh, and A. R. Al-Ali, “Deep Learning Approach for Wind Energy Time-Series Forecasting with Environmental Factors,” in Proc. American University of Sharjah, Sharjah, UAE, Sep. 2023, pp. 1–15.