International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 265-272 www.ijaem.net ISSN: 2395-5252

A Review on Statistical Models for Estimating Renewable Energy Production

Eti Jain[1] Prof. A.K. Sharma[2]
Department of Electrical Engineering[1,2]
Jabalpur Engineering College, Jabalpur, India

Date of Submission: 10-12-2024 Date of Acceptance: 20-12-2024

ABSTRACT: The transition towards renewable energy is no longer just an option but a necessity in addressing the growing challenges of climate change, environmental degradation, and energy security. Fossil fuels, which currently dominate global energy consumption, are finite and contribute significantly to greenhouse gas emissions. Shifting to renewable energy sources such as solar, wind, hydropower, and biomass offers a sustainable path forward, benefiting both the environment and global economies. Despite its benefits, the migration to renewable energy faces challenges, including the need for substantial initial investments, technological limitations, and resistance from entrenched fossil fuel industries. Governments, private sectors, and international organizations must work together to create policies and incentives that encourage renewable energy adoption. Public awareness campaigns, subsidies, and research funding are critical in overcoming these barriers. To develop a sustainable renewable energy infrastructure, we need to forecast the amount of renewable energy production in future and meet energy demands in future through renewable sources. This paper presents a comprehensive review of statistical models for forecasting renewable energy along with associated challenges that the sector faces.

Keywords: Energy Demand, Renewable Energy, Statistical Models to total primary energy supply (TPES), Regression Models.

I. INTRODUCTION

One of the most pressing reasons to migrate to renewable energy is its potential to combat climate change [1]. The burning of fossil fuels releases massive amounts of carbon dioxide and other greenhouse gases into the atmosphere, driving global warming. In contrast, renewable energy sources produce little to no emissions during operation. For instance, solar panels and wind turbines generate clean energy without polluting the air or water [2].

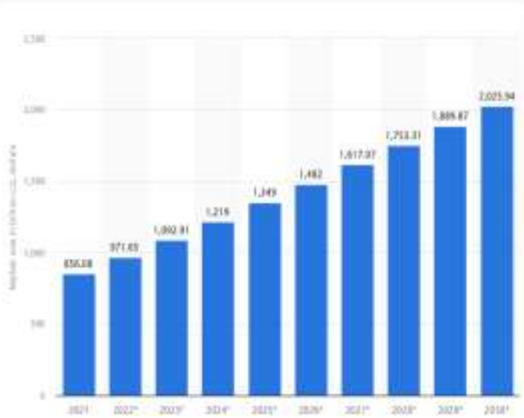

Fig.1 Global Renewable Energy Market (Source: https://www.statista.com/statistics/1094309/renewable-energy-market-size-global/)

Figure 1 depicts the Global Renewable Energy Market. By transitioning to renewables, we can significantly reduce our carbon footprint and slow the pace of environmental destruction. Other factors include:

Economic Benefits of Renewable Energy: Renewable energy also presents substantial economic opportunities [3]. As technology advances, the cost of generating renewable energy has dropped dramatically, making it increasingly competitive with traditional fossil fuels. Investments in renewable infrastructure create jobs in manufacturing, installation, and maintenance, stimulating economic growth. Furthermore, by reducing dependence on imported fossil fuels, countries can enhance their energy security and achieve greater economic stability [4].

Energy Security: Renewable energy sources are abundant and widely distributed, making them less susceptible to geopolitical

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 265-272 www.ijaem.net ISSN: 2395-5252

tensions and supply chain disruptions. Solar and wind energy, for instance, are available almost everywhere, enabling nations to generate their own energy rather than relying on volatile international markets. By investing in renewable energy, countries can secure a stable and reliable energy supply for future generations [5].

Health and Social Advantages: The migration to renewable energy also improves public health by reducing air and water pollution caused by burning fossil fuels. Cleaner air leads to fewer respiratory and cardiovascular illnesses, decreasing healthcare costs and improving quality of life. Additionally, renewable energy systems often require less water for operation than traditional energy sources, alleviating stress on water resources in drought-prone regions [6].

Technological Advancements and Scalability: Innovations in renewable energy technologies, such as advanced solar panels, efficient wind turbines, and energy storage systems, have made renewables more scalable and accessible. Energy storage solutions, like batteries, enable consistent power supply even when the sun isn't shining or the wind isn't blowing. Grid modernization and smart technologies are also paving the way for seamless integration of renewables into existing energy systems, making the transition more feasible [7].

II. FORECASTING RENEWABLE ENERGY

To forecast future renewable energy trends, it is necessary to model it in terms of a time series model as:

$$\textbf{Energy Generation} = \mathbf{f}(\textbf{time}, \textbf{other govering variables}) \quad \textbf{(1)}$$

As the global population continues to grow and industrialization accelerates, the demand for energy is expected to rise significantly. At the same time, the environmental consequences of relying on fossil fuels have become increasingly evident. This creates a pressing need to transition toward renewable energy sources [8]. Renewable energy offers a sustainable, reliable, and environmentally friendly solution to meet future energy demands while addressing critical issues like climate change and energy security. Forecasting the role of renewable energy in the future reveals its importance in shaping a sustainable world [9].

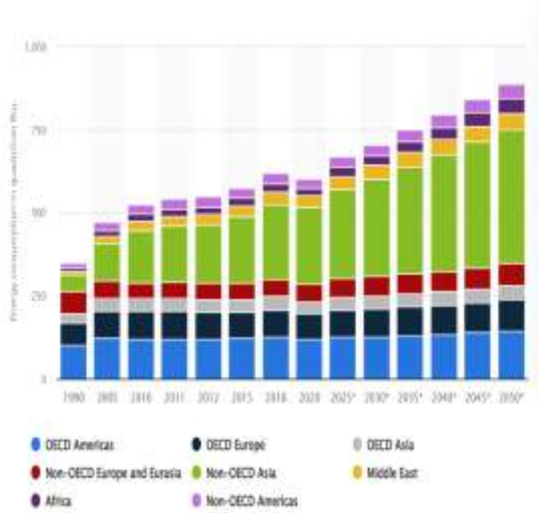

Fig.2 Future Energy Forecast (Source: https://www.statista.com/statistics/268297/world-energy-consumption-outlook/)

Figure 2 depicts the Energy consumption worldwide from 1990 to 2020, with a forecast until 2050, by region*(in quadrillion British thermal units [10].

While statistical models have been employed thus far, yet machine learning algorithms are proving to be more powerful in this regard in terms of forecasting accuracy. Accurate energy forecasting is crucial for maintaining the balance between energy supply and demand. Overestimating demand can lead to unnecessary energy production and wasted resources, while underestimating demand can cause energy shortages and grid instability. By leveraging ML models, utility companies and grid operators can make more precise predictions, minimizing these risks and optimizing energy production and distribution [12].

Machine learning (ML) models have emerged as powerful tools for analyzing complex data and generating reliable forecasts. These models can process vast amounts of historical data, capture intricate patterns, and adapt to dynamic systems. By leveraging machine learning, energy providers and policymakers can optimize renewable energy generation, distribution, and consumption [13].

Forecasting renewable energy demand is essential for integrating renewable sources like solar and wind into power grids. Unlike traditional energy sources, renewable energy is often variable and intermittent due to weather conditions and

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 265-272 www.ijaem.net ISSN: 2395-5252

seasonal changes. Accurate forecasts help predict fluctuations in energy consumption, align production with demand, and prevent issues like overgeneration or power shortages. Machine learning models are particularly well-suited for this task, as they can analyze both historical energy usage data and external factors like weather patterns, economic trends, and population growth.

III. EXISTINGSTATISTICAL MODELS

The statistical machine learning models used for forecasting are presented in brevity in this section [14]:

Support Vector Machine (SVM):

Before the advent of deep learning, traditional machine learning models such as Support Vector Machines (SVM), Decision Trees, Random Forests, and K-Nearest Neighbors (KNN) were widely used for satellite object detection. These models typically relied on handcrafted features, such as texture, edges, and spectral indices, to distinguish between different objects.

The SVM classifies based on the hyperplane.
The selection of the hyperplane H is done on the basis of the maximum value or separation in the Euclidean distance d given by:

$$\mathbf{d} = \sqrt{\mathbf{x}_1^2 + \cdots \dots \dots \mathbf{x}_n^2} \qquad \textbf{(2)}$$

Here,
x represents the separation of a sample space variables or features of the data vector,
n is the total number of such variables
d is the Euclidean distance

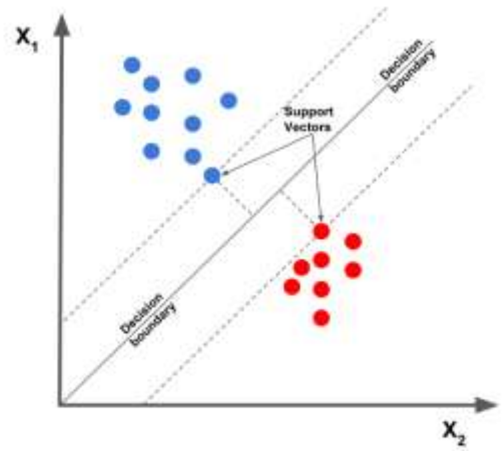

Fig.3 The SVM Model

Figure 3 depicts the SVM Model.

The (n-1) dimensional hyperplane classifies the data into categories based on the maximum separation. For a classification into one of **‘m’** categories, the hyperplane lies at the maximum separation of the data vector **‘X’.** The categorization of a new sample **‘z’** is done based on the inequality:

$$\mathbf{d_x^z = Min(d_{C1}^z, d_{C2}^z \dots d_{C2=m}^z)} \qquad \textbf{(3)}$$

Here,
d_x^z is the minimum separation of a new data sample from **‘m’** separate categories
$d_{C1}^z, d_{C2}^z \dots d_{C2=m}^z$ are the Euclidean distances of the new data sample **‘z’** from m separate data categories.

For instance, SVMs are effective for binary classification tasks, such as distinguishing between urban and rural areas, while Random Forests are used for multi-class classification problems, such as land cover mapping. However, these models struggle with complex patterns in high-resolution imagery and require extensive feature engineering, which limits their scalability and accuracy

ARIMA:

In an autoregressive integrated moving average model commonly known as the ARIMA model assumes that the future value of a variable can be linearly modelled as a function previous samples of the variables and errors of prediction.

$$\mathbf{y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_p y_{t-p} + \cdots \dots \dots \theta_q \varepsilon_{t-q}} \quad \textbf{(4)}$$

Here,
y_t is the value of the output variable at time ‘t’
ε is the prediction error
θ and φ are called the model parameters
p and q are called the orders of the model

One of ARIMA's key strengths lies in its ability to handle both stationary and non-stationary data. While the ARIMA model assumes the input time series is stationary (i.e., its statistical properties like mean and variance remain constant over time), it incorporates differencing techniques to convert non-stationary data into a stationary format. This makes it highly adaptable for real-world datasets that often exhibit trends or seasonality.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 265-272 www.ijaem.net ISSN: 2395-5252

Neural Networks [15]:

Owing to the need of non-linearity in the separation of data classes, one of the most powerful classifiers which have become popular is the artificial neural network (ANN). The neural networks are capable to implement non-linear classification along with steep learning rates. The neural network tries to emulate the human brain's functioning based on the fact that it can process parallel data streams and can learn and adapt as the data changes. This is done through the updates in the weights and activation functions.

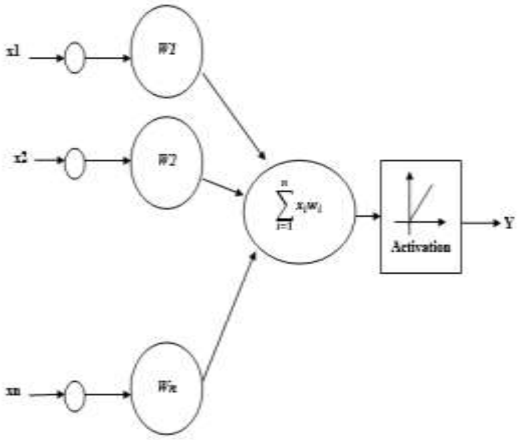

Fig.4 The ANN Model

Figure 4 depicts the ANN model.
The input-output relation of a CNN is given by:

$$\mathbf{y = f(\sum_{i=1}^{n} x_i w_i + b)} \quad \mathbf{(5)}$$

Here,
x denote the parallel inputs
y represents the output
w represents the bias
f represents the activation function

The neural network is a connection of such artificial neurons which are connected or stacked with each other as layers. The neural networks can be used for both regression and classification problems based on the type of data that is fed to them. Typically the neural networks have 3 major conceptual layers which are the input layer, hidden layer and output layer. The parallel inputs are fed to the input layer whose output is fed to the hidden layer. The hidden layer is responsible for analysing the data, and the output of the hidden layer goes to the output layer. The number of hidden layers depends on the nature of the dataset and problem under consideration. If the neural network has multiple hidden layers, then such a neural network is termed as a deep neural network. The training algorithm for such a deep neural network is often termed as deep learning which is a subset of machine learning. Typically, the multiple hidden layers are responsible for computation of different levels of features of the data.

Long Short Term Memory (LSTM):

The LSTM networks are a specialized type of recurrent neural network (RNN) designed to process and predict data sequences by learning long-term dependencies. Unlike traditional RNNs, which suffer from vanishing or exploding gradient problems during training, LSTMs incorporate a unique architecture with gates and memory cells that help retain important information over long periods.

The LSTM primarily has 3 gates:
1) Input gate: This gate collects the presents inputs and also considers the past outputs as the inputs.
2) Output gate: This gate combines all cell states and produces the output.
3) Forget gate: This is an extremely important feature of the LSTM which received a cell state value governing the amount of data to be remembered and forgotten.

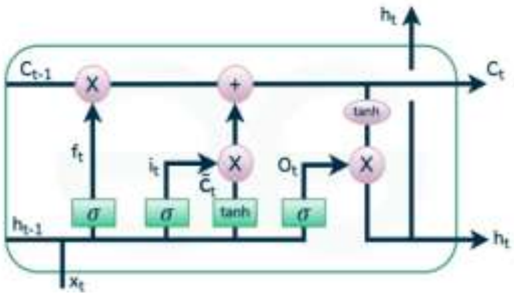

Fig.5 The LSTM Model

Figure 5 depicts the LSTM model.

The relation to forget by the forget gate is given by:

$$\mathbf{f = \sigma(W_f[h_{t-1}, x_t] + b_i)} \quad \mathbf{(6)}$$

Here,
f denotes forget gate activation
w_f are forget gate weights.
h_{t-1} Denotes Hidden state from the previous time step
x_t is present input.
b_i is the bias

The advantages of LSM are:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 265-272 www.ijaem.net ISSN: 2395-5252

Capturing Long-Term Dependencies: LSTMs maintain long-term memory using the cell state, unlike traditional RNNs.
Mitigating Vanishing/Exploding Gradients: Gates help regulate gradient flow, enabling stable training over long sequences.
Versatility: Useful for several time series prediction problems.

However, the major challenge happens to be the problem of overfitting.

Convolutional Neural Networks (CNNs): The family of CNNs are the backbone of modern satellite object detection. CNNs automatically learn hierarchical features from raw images, eliminating the need for manual feature extraction. The Convolutional Neural Networks (CNNs) can automatically extract hierarchical characteristics from images, they have become the mainstay for image classification applications. These neural networks perform exceptionally well in applications like picture identification because they are specifically made for processing organised grid data.

Convolutional, pooling, and fully linked layers are among the layers that make up a CNN's architecture. Convolutional layers identify patterns in the input image by applying filters, hence identifying local features. By reducing spatial dimensions, pooling layers preserve significant information. High-level features are integrated for categorization in fully connected layers.

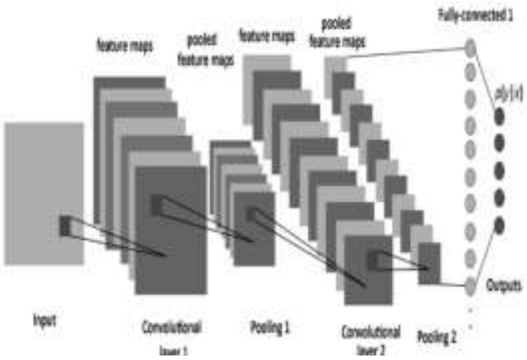

Fig.6 The CNN Model

Figure 6 depicts the CNN model.

The convolution operation is given by:

$$\boldsymbol{x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau} \qquad \textbf{(7)}$$

Here,
x(t) is the input
h(t) is the system under consideration.
y is the output
*is the convolution operation in continuous domain
For a discrete or digital counterpart of the data sequence, the convolution is computed using:

$$\boldsymbol{y(n) = \sum_{-\infty}^{\infty} x(k)h(n-k)} \qquad \textbf{(8)}$$

Here
x(n) is the input
h(n) is the system under consideration.
y is the output
*is the convolution operation in discrete domain

In this approach, the back propagation based neural network model has been used. A backpropagation neural network for traffic speed forecasting typically consists of an input layer, one or more hidden layers, and an output layer. The number of nodes in the input layer corresponds to the features used for prediction, The hidden layers contain nodes that learn and capture the intricate patterns within the data, while the output layer provides the predicted traffic speed. The training of a backpropagation neural network involves the iterative application of the backpropagation algorithm. During the training process, historical data is used to feed the network, and the algorithm calculates the error between the predicted and actual energy demands. This error is then propagated backward through the network, adjusting the weights and biases of the connections to minimize the prediction error. This iterative process continues until the network converges to a state where the error is minimized.Successful backpropagation neural network models for traffic speed forecasting can be integrated into energy management systems.

IV. PREVIOUS WORK

A summary of noteworthy contribution in the domain is presented here:

Aziz et al.[16] proposed a hybrid machine learning technique to precisely predict the energy level of natural resources. The hybrid machine learning is the combination of Multilayer Perceptron (MLP), Support Vector Regression (SVR) and CatBoost algorithm that increases the performance and predictability of renewable energy consumption. The proposed system dataset's results are evaluated at the train and test levels, and the results are then compared to other current approaches. The end results reveal that the proposed hybrid machine learning technique has a high prediction level when compared to others, as well as a lower cost rate and improved overall system performance.

Mostafa et al. [17] proposed a framework was developed for the potential implementation of big data analytics for smart grids and renewable energy power utilities. A five-step approach is proposed for predicting the smart grid stability by using five different machine learning methods. Data from a decentralized smart grid data system consisting of 60,000 instances and 12 attributes was used to predict the stability of the system through three different machine learning methods. The results of fitting the penalized linear regression model show an accuracy of 96% for the model implemented using 70% of the data as a training set. Using the random forest tree model has shown 84% accuracy, and the decision tree model has shown 78% accuracy. Both the convolutional neural network model and the gradient boosted decision tree model yielded 87% for the classification model.

Mahmud et al. [18] proposed a machine learning-based PV power generation forecasting for both the short and long-term. The study has chosen Alice Springs, one of the geographically solar energy-rich areas in Australia, and considered various environmental parameters. Different machine learning algorithms, including Linear Regression, Polynomial Regression, Decision Tree Regression, Support Vector Regression, Random Forest Regression, Long Short-Term Memory, and Multilayer Perceptron Regression, are considered in the study. Various comparative performance analysis is conducted for both normal and uncertain cases and found that Random Forest Regression performed better for our dataset. The impact of data normalization on forecasting performance is also analyzed using multiple performance metrics. The study may help the grid operators to choose an appropriate PV power forecasting algorithm and plan the time-ahead generation volatility.

Goncalves et al. [19] proposed that data exchange between multiple renewable energy power plant owners can lead to an improvement in forecast skill thanks to the spatio-temporal dependencies in time series data. However, owing to business competitive factors, these different owners might be unwilling to share their data. In order to tackle this privacy issue, this paper formulates a novel privacy-preserving framework that combines data transformation techniques with the alternating direction method of multipliers. This approach allows not only to estimate the model in a distributed fashion but also to protect data privacy, coefficients and covariance matrix. Besides, asynchronous communication between peers is addressed in the model fitting, and two different collaborative schemes are considered: centralized and peer-to-peer. The results for a solar energy dataset show that the proposed method is robust to privacy breaches and communication failures, and delivers a forecast skill comparable to a model without privacy protection.

Corrizo et al. [20] proposed that increasing presence of renewable energy plants has created new challenges such as grid integration, load balancing and energy trading, making it fundamental to provide effective prediction models. Recent approaches in the literature have shown that exploiting spatio-temporal autocorrelation in data coming from multiple plants can lead to better predictions. Although tensor models and techniques are suitable to deal with spatio-temporal data, they have received little attention in the energy domain. In this paper, we propose a new method based on the Tucker tensor decomposition, capable of extracting a new feature space for the learning task. For evaluation purposes, we have investigated the performance of predictive clustering trees with the new feature space, compared to the original feature space, in three renewable energy datasets. The results are favorable for the proposed method, also when compared with state-of-the-art algorithms.

Typically, training algorithms try to attain low error rate metrics, which are defined next [21]:
The mean square error or mse given by:

$$mse = \frac{\sum_{i=1}^{n} e_i^2}{n} \quad (9)$$

The final computation of the performance metric is the mean absolute percentage error given by:

$$\mathbf{MAPE} = \frac{100}{M}\sum_{i=1}^{N}\frac{E-E_i|}{i} \quad (10)$$

The accuracy of prediction is computed as:

$$\mathbf{Ac} = 100 - \frac{100}{M}\sum_{i=1}^{N}\frac{E-E_i|}{i}\ \% \qquad (11)$$

Here,
n is the number of errors
i is the iteration number
E is the actual value
E_i is the predicted value

V. CONCLUSION

It can be concluded that a quick transition to renewable energy sources is essential in light of the growing population, increased energy consumption, and severe climate change. So that

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 12 Dec. 2024, pp: 265-272 www.ijaem.net ISSN: 2395-5252

supply and demand are balanced, it is necessary to have a precise forecast of both the future energy demands and the capacity of renewable energy sources. Machine learning models are revolutionizing the way renewable energy demand is forecasted. From simple linear regression to advanced neural networks like LSTMs, these models offer unparalleled accuracy and flexibility in predicting energy needs. By leveraging these tools, energy providers and policymakers can optimize renewable energy systems, improve grid stability, and plan for a sustainable future. As machine learning technologies continue to evolve, they will play an even more significant role in the global transition to clean energy. This paper presents the necessity to forecast renewable energy trends along with existing baseline models which would allow future research in developing accurate forecasting models for the domain.

REFERENCES:

[1] W. Ahmed et al., "Machine Learning Based Energy Management Model for Smart Grid and Renewable Energy Districts," in IEEE Access, vol. 8, pp. 185059-185078, 2020.

[2] G. Li, S. Xie, B. Wang, J. Xin, Y. Li and S. Du, "Photovoltaic Power Forecasting With a Hybrid Deep Learning Approach," in IEEE Access, vol. 8, pp. 175871-175880, 2020,.

[3] 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 ANDESCON, Quito, Ecuador, 2020, pp. 1-6/

[4] M. A. Muñoz, J. M. Morales and S. Pineda, "Feature-Driven Improvement of Renewable Energy Forecasting and Trading," in IEEE Transactions on Power Systems, vol. 35, no. 5, pp. 3753-3763, Sept. 2020.

[5] N. Shabbir, R. AhmadiAhangar, L. Kütt, M. N. Iqbal and A. Rosin, "Forecasting Short Term Wind Energy Generation using Machine Learning," 2019 IEEE 60th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON), Riga, Latvia, 2019, pp. 1-4.

[6] S. Shamshirband, T. Rabczuk and K. -W. Chau, "A Survey of Deep Learning Techniques: Application in Wind and Solar Energy Resources," in IEEE Access, vol. 7, pp. 164650-164666, 2019.

[7] P Sharmila, J Baskaran, C Nayanatara, "A hybrid technique of machine learning and data analytics for optimized distribution of renewable energy resources targeting smart energy management", Procedia in Computer Science, Elsevier 2019, vol.165, pp. 278-284.

[8] J Del Ser, D Casillas-Perez, L Cornejo-Bueno, L Prieto-Godino, J Sanz-Justo, "Randomization-based machine learning in renewable energy prediction problems: critical literature review, new results and perspectives", Applied Soft Computing, Elsevier 2022, vol.118, pp. 108526.

[9] IE Livieris, "An advanced active set L-BFGS algorithm for training weight-constrained neural networks", Neural Computing and Applications, Springer 2020, vol.32, pp. 6669–6684.

[10] M Huisman, A Plaat, JN van Rijn, "Stateless neural meta-learning using second-order gradients", Machine Learning, Springer 2022, vol.111, pp.3227–3244.

[11] R. Xin, S. Pu, A. Nedić and U. A. Khan, "A General Framework for Decentralized Optimization With First-Order Methods," in Proceedings of the IEEE, 2020, vol. 108, no. 11, pp. 1869-1889.

[12] M. Liu, L. Chen, X. Du, L. Jin and M. Shang, "Activated Gradients for Deep Neural Networks," in IEEE Transactions on Neural Networks and Learning Systems, 2023, vol. 34, no. 4, pp. 2156-2168.

[13] 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 ANDESCON, Quito, Ecuador, 2020, pp. 1-6.

[14] G. M. Khan, J. Ali and S. A. Mahmud, "Wind power forecasting — An application of machine learning in renewable energy," 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China, 2014, pp. 1130-113.

[15] RM Abd El-Aziz, "Renewable power source energy consumption by hybrid machine learning model", Alexandria Engineering Journal, Elsevier, 2022, vol.61, no.12, pp. 9447-9455

[16] J. R. Andrade and R. J. Bessa, "Improving Renewable Energy Forecasting With a Grid of Numerical Weather Predictions," in IEEE Transactions on Sustainable Energy, 2017, vol. 8, no. 4, pp. 1571-1580

[17] N Mostafa, HSM Ramadan, O Elfarouk, "Renewable energy management in smart

grids by using big data analytics and machine learning”, Machine Learning with Applications, Elsevier, 2023, vol.9, 100363.
[18] K. Mahmud, S. Azam, A. Karim, S. Zobaed, B. Shanmugam and D. Mathur, "Machine Learning Based PV Power Generation Forecasting in Alice Springs," in IEEE Access, vol. 9, pp. 46117-46128, 2021.
[19] C. Gonçalves, R. J. Bessa and P. Pinson, "Privacy-Preserving Distributed Learning for Renewable Energy Forecasting," in IEEE Transactions on Sustainable Energy, vol. 12, no. 3, pp. 1777-1787, July 2021.
[20] R Corizzo, M Ceci, H Fanaee-T, J Gama, “Multi-aspect renewable energy forecasting”, Information Sciences, Elsevier 2021, vol.546, pp.701-722.
[21] M. Xia, H. Shao, X. Ma and C. W. de Silva, "A Stacked GRU-RNN-Based Approach for Predicting Renewable Energy and Electricity Load for Smart Grid Operation," in IEEE Transactions on Industrial Informatics, vol. 17, no. 10, pp. 7050-7059