Journal of Recent Trends in Computer Science and Engineering (JRTCSE)
Vol. 14, No.1, January - February 2026, PP. 16-25.
ISSN: 2322-0872
https://jrtcse.com
DOI: https://doi.org/10.70589/JRTCSE.2026.14.1.3

Deep Learning Driven Renewable Energy Forecasting Using Distributed Cloud Computing and Large Scale Weather Data

Ashvini S. Kolate,
MSc. Computer Science, PO Nahata College Bhusawal, Maharashtra 425201.

Vikas M. Somvanshi,
Lecturer in Department of Computer Engineering, SSVPS College Of Engineering, Dhule, Maharashtra 424005

Abstract

Renewable energy sources such as solar and wind are expanding rapidly; however, their output fluctuates with weather conditions. This variability makes accurate forecasting essential for power grid management and energy planning. Traditional forecasting methods often struggle to model complex weather data effectively. Recent advances in deep learning and cloud computing enable more accurate and scalable forecasting approaches. This paper reviews deep learning methods for renewable energy forecasting, with particular emphasis on systems that integrate distributed cloud computing and large-scale weather datasets. We organize the existing literature according to model type, infrastructure design, and data source. We then compare recent studies in terms of accuracy, scalability, and computational efficiency. Finally, we highlight key challenges and research gaps and propose future directions for developing smarter and more sustainable forecasting systems.

Keywords: Renewable Energy Forecasting; Deep Learning; Solar Power Prediction; Wind Power Prediction; Distributed Cloud Computing; Big Data Analytics; Transformer Models; LSTM; Smart Grid

Citation: Ashvini S. Kolate & Vikas M. Somvanshi. (2026). Deep learning driven renewable energy forecasting using distributed cloud computing and large scale weather data. Journal of Recent Trends in Computer Science and Engineering, 14(1), 16–25.
https://doi.org/10.70589/JRTCSE.2026.14.1.3

1. Introduction

The increasing global demand for sustainable energy has led to significant deployment of solar and wind power systems. However, renewable energy generation are inherently intermittent and highly depend on meteorological conditions. Accurate forecasting plays a critical role for grid stability, economic dispatch, demand response management, and energy market operation [1]. Weather variability introduces nonlinear and stochastic patterns that traditional linear model fails to capture effective [2]. Moreover, the availability of high-resolution meteorological data from satellites, sensor, and numerical weather prediction system has resulted in large-scale dataset that requires advanced processing framework [3].

Deep learning has emerged as a powerful tool for modeling complex temporal dependencies in time-series data [4]. Meanwhile, cloud computing platforms enable distributed storage, scalable training, and real-time deployment of forecasting models [5]. Despite these advances, research remains fragmented across model development and infrastructure scalability. This review systematically analyzes deep learning approaches for renewable energy forecasting, emphasizing distributed cloud computing and big data integration.

2. Renewable Energy Forecasting Background

2.1 Solar Energy Forecasting

Solar photovoltaic (PV) forecasting depends on parameters such as solar irradiance, temperature, humidity, and cloud cover [6]. Forecasting horizons include ultra-short-term (minutes), short-term (hours to days), and long-term (weeks to months). Recent research shows that deep learning methods significantly outperform classical regression and ARIMA-based models in capturing nonlinear solar patterns [7]. These approaches learn complex relationships between meteorological variables and PV output without requiring extensive manual feature engineering.

However, solar forecasting remains challenging due to rapid cloud movement, sudden weather changes, and inconsistent sensor data. Integrating high-resolution weather datasets, including satellite imagery and numerical weather predictions, can improve forecast accuracy but also increases data processing demands.

2.2 Wind Energy Forecasting

Wind energy prediction is more complex because turbulence and atmospheric instability make it more harder to model. Machine learning and deep learning techniques was applied to wind speed and wind power forecasting, and they often shows improved accuracy compare to traditional physical model [8]. These models can capture nonlinear relationship between wind speed, wind direction, atmospheric pressure and power output, which is very difficult to be modeled by classical approach. However, wind forecasting are sensitive to local terrain and micro–climate effect, causing large variations between different region. Also, wind data often contain many noise and missing value, so it need robust pre–processing and data cleaning method. Deep learning architecture like recurrent network and hybrid model has show promising result, especially for short–term forecasting where to capture rapid change are essential for keep grid stability.

2.3 Forecasting Horizons

Short-term forecasting asre essential for operational grid management, whereas long-term forecasting support planning and investment decision. Short-term forecasts enable utility to balance supply and demands, schedule reserve, and perform economic dispatch, while long-term forecast inform capacity expansions and strategic plan. Deep neural network is particularly effective for short-term forecasting due to its ability to learns temporal dependency and adapt to changing weather pattern. However, long-term forecasting requires broader spatial and temporal contexts and must account for seasonal and climate variabilities, making it more challenge. The increasing available of large-scale weather dataset and advance deep learning model provide opportunity to improve long-term prediction, but it also demand scalable computing framework and careful model validations to ensured reliable.

3. Deep Learning Models for Renewable Forecasting

3.1 Recurrent Neural Networks (RNN)

RNNs were among the earliest deep learning approaches used for time-series prediction. However, they suffer from vanishing gradient problems when modeling long sequences [4]. This limitation reduces their effectiveness in capturing long-term dependencies in renewable energy data, especially for longer forecasting horizons. Although RNNs can model sequential data, their performance declines as the length of input sequences increases. Despite these challenges, RNNs laid the foundation for more advanced recurrent architectures and continue to be used in simpler forecasting tasks or as baseline models in comparative studies.

3.2 Long Short-Term Memory (LSTM)

LSTM networks address RNN limitations through memory cells and gating mechanisms. They are widely used for solar and wind forecasting due to their strong temporal modeling capabilities [10]. LSTMs retain information over long periods, enabling improved prediction accuracy. However, they require substantial computational resources and long training times, particularly when trained on large-scale datasets, which motivates the use of distributed cloud computing

3.3 Gated Recurrent Units (GRU)

GRU models simplify LSTM architecture while maintaining comparable performance and lower computational cost [11]. By combining the input and forget gates into a single update gate, GRU reduces complexity and accelerates training. GRU models have shown strong results in wind power prediction and other renewable forecasting tasks, making them suitable for applications where computational efficiency is crucial. However, they may be less effective than LSTM in capturing extremely long-term dependencies, depending on the dataset and forecasting horizon.

3.4 CNN-LSTM Hybrid Models

CNN–LSTM architectures extract spatial features through convolutional layers and capture temporal dependencies using LSTM layers. These models are particularly effective when satellite-based or grid-structured weather data are incorporated [12]. The CNN component identifies spatial patterns, such as cloud movement and wind field structures, whereas the LSTM component models their temporal evolution. Hybrid architectures have demonstrated superior predictive accuracy compared with single-model approaches, especially when integrating multi-source data, including satellite imagery and meteorological measurements. However, they typically require large datasets and substantial computational resources, making distributed cloud-based training frameworks essential for practical and scalable deployment.

3.5 Transformer-Based Models

Transformers leverage self-attention mechanisms to capture global dependencies in long sequences. Recent studies indicate superior performance for long-horizon renewable forecasting, though computational cost is significant [13]. Transformers can attend to all time steps simultaneously, enabling them to model long-range interactions and complex temporal patterns. While they often outperform recurrent architectures for long sequences, their high memory and processing requirements present challenges for training and real-time deployment. Distributed cloud computing and scalable frameworks are therefore crucial to manage the computational demands of transformer-based forecasting systems.

4. Cloud and Distributed Computing for Renewable Forecasting

4.1 Big Data Challenges

Renewable energy forecasting systems process massive and diverse datasets. Researchers collect historical power generation records over long periods. They also gather high-resolution weather datasets from monitoring stations and numerical prediction models. Satellite imagery provides spatial information about cloud movement and atmospheric behavior. IoT sensors installed at renewable plants continuously stream operational data.

These datasets differ in format, scale, frequency, and spatial resolution. Researchers must clean, align, and preprocess them before model training. Missing values, noise, and inconsistent sampling create additional complexity. Systems must also store and retrieve large volumes of data efficiently. Real-time forecasting requires low-latency pipelines that can process continuous data streams without delays. As dataset size increases, computational demand rises sharply. These challenges force researchers to adopt scalable storage and processing frameworks [3].

4.2 Distributed Training Frameworks

Researchers utilize cloud computing platforms to address increasing computational demands. Cloud providers offer elastic resources that dynamically scale according to workload requirements. Deep learning tasks can be distributed across multiple GPUs or computing nodes to accelerate training. Frameworks such as Apache Spark enable parallel data processing and support large-scale model training. GPU clusters further enhance computational efficiency for complex architectures, including LSTM and transformer-based models.

Distributed learning significantly reduces training time and allows efficient processing of large-scale meteorological and power generation datasets. Cloud environments also provide automated resource allocation, monitoring, and fault tolerance, thereby improving system reliability and operational flexibility. As model complexity and dataset size continue to grow, distributed training becomes a necessity rather than an option [5].

4.3 Scalability Considerations

Distributed systems improve performance, but they introduce new challenges. Nodes must communicate frequently during model training. This communication creates overhead and slows down synchronization. Network latency can further delay parameter updates across distributed clusters. When systems operate across geographical regions, latency becomes more significant.

Cloud deployment also introduces cost considerations. High-performance GPU instances and large storage volumes increase operational expenses. Researchers must balance prediction accuracy with infrastructure cost. Energy consumption represents another critical issue. Large-scale data centers consume substantial electricity during deep learning training. This raises sustainability concerns and contradicts the environmental goals of renewable energy research [14]. Therefore, developers must design scalable yet energy-efficient forecasting systems.

5. Taxonomy of Existing Research

Figure 1 shows a heatmap that summarizes the research coverage across model types and infrastructure categories. The heatmap highlights common research areas, such as LSTM models on cloud platforms, and it also reveals under-explored areas, such as transformer models on distributed big data frameworks with multi-source data. This visualization helps to identify research gaps and guides future work.

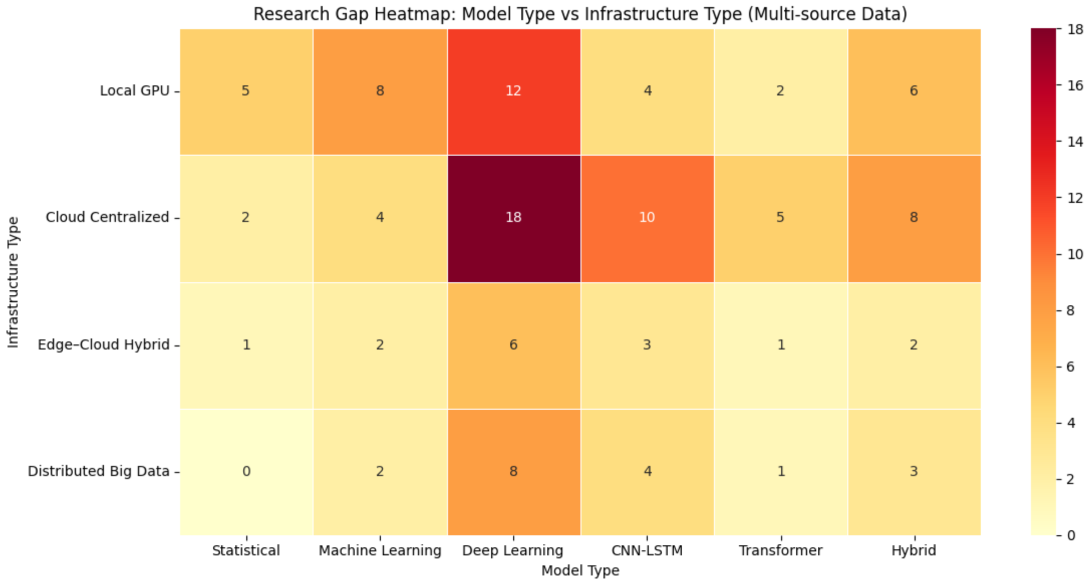

Figure 1. Heatmap of research distribution across model types and deployment infrastructures

5.1 Based on Model Type

Researchers classify renewable energy forecasting models into several main types. Statistical models like ARIMA and SARIMA represent early approaches, and they work well for simple and linear time-series. However, these models struggle with nonlinear renewable energy patterns [2]. Machine learning models such as Support Vector Machines and Random Forest improve prediction accuracy, but they need manual feature engineering and domain expertise [8]. Deep learning models have transformed forecasting research, and architectures like LSTM and GRU capture long-term temporal patterns automatically. CNN models extract spatial features from grid-based or satellite data, and transformer models use self-attention to learn global dependencies in long sequences. These models usually achieve higher accuracy than traditional methods [10][13]. Researchers also use hybrid models that combine statistical and deep learning approaches, such as ARIMA-LSTM or CNN-LSTM, to improve robustness and performance [12].

5.2 Based on Infrastructure

Infrastructure choice affects forecasting performance and scalability. Early studies used local GPU systems, which limited dataset size and model complexity. As cloud computing became more accessible, researchers adopted cloud platforms for training and deployment, and these platforms offer scalable storage and computing power. Recent studies explore edge–cloud systems, where edge devices handle initial data processing and the cloud performs heavy training tasks. This approach reduces latency and improves response time. Researchers also use distributed big data frameworks such as Hadoop and Apache Spark to process large datasets and parallelize training tasks. Infrastructure design directly impacts cost, scalability, and energy efficiency [5].

5.3 Based on Data Source

The choice of data sources strongly affects forecasting accuracy. Some studies use only historical generation data, which captures time-series patterns but ignores weather influences. Weather-based models include parameters like temperature, wind speed, humidity, and solar irradiance, and these models usually perform better. Satellite-based approaches add spatial information about cloud cover and atmospheric conditions, improving short-term solar forecasts. More advanced systems use multi-source data fusion, combining weather data, satellite images, and IoT sensor streams into a single model. This integration captures complex spatiotemporal patterns and improves generalization [12].

6. Comparative Analysis of Recent Studies

Recent comparative studies demonstrate the evolution of renewable forecasting techniques from statistical models toward deep learning and distributed cloud implementations. ARIMA-based approaches reported moderate accuracy but were constrained by linear assumptions [7]. LSTM models achieved substantial improvements in prediction accuracy, although training time remained significant [10]. GRU-based systems reduced computational complexity but exhibited limitations in cross-regional generalization [11]. CNN-LSTM hybrid models enhanced performance through spatial-temporal feature extraction, particularly when satellite data were incorporated [12]. Transformer-based models deployed on distributed cloud platforms achieved the lowest mean absolute percentage error among contemporary approaches but incurred high computational and infrastructure costs [13]. These findings highlight the trade-off between accuracy and computational efficiency across different architectures.

Table 1. Comparison of Forecasting Approaches and Computing Platforms

Ref	**Year**	**Model**	**Dataset**	**Infrastructure**	**Accuracy Metric**	**Limitations**
[7]	2019	ARIMA	Solar PV	Local	RMSE 6.5%	Linear assumption
[10]	2020	LSTM	Solar + Weather	GPU	RMSE 4.1%	High training time
[11]	2021	GRU	Wind	Local	MAE 3.8%	Limited generalization
[12]	2022	CNN-LSTM	Satellite + Weather	Cloud	MAPE 3.2%	Data intensive

[13]	2023	Transformer	Multi-source	Distributed Cloud	MAPE 2.7%	High computation cost

Table 1 summarizes recent studies on renewable energy forecasting and compares their models, datasets, infrastructure, and performance. The table shows how forecasting methods have shifted from traditional statistical models to deep learning and distributed cloud systems. ARIMA-based models deliver moderate accuracy but rely on linear assumptions, which limits their effectiveness [7]. LSTM models improve prediction accuracy significantly, yet they require long training times and high computing power [10]. GRU models reduce computational complexity but they often struggle to generalize across different regions [11]. CNN-LSTM hybrids improve performance by extracting spatial and temporal features, especially when satellite data is included, but they need large amounts of data and processing resources [12]. Transformer-based models on distributed cloud platforms show the best accuracy among the listed studies, but they incur high computation and infrastructure costs [13]. Overall, the table highlights the trade-off between accuracy and computational efficiency in different forecasting architectures.

7. Challenges Identified in Literature

The literature highlights several ongoing challenges in renewable energy forecasting. Data quality remains a major concern, as imbalanced datasets and missing sensor values can negatively affect model accuracy and reliability [6, 9]. Gaps in data collection often introduce bias and reduce predictive performance. Weather uncertainty and extreme events also pose significant difficulties. Sudden changes in atmospheric conditions are inherently complex and can lead to substantial forecasting errors [1]. In addition, advanced forecasting models, particularly deep learning approaches, require considerable computational resources, which may limit their practicality in real-time applications [14].

Another important issue is the limited evaluation of system scalability. While many studies focus primarily on improving prediction accuracy, fewer assess computational efficiency, training time, or resource consumption in large-scale or cloud-based environments [5, 14]. The absence of standardized benchmark datasets further complicates fair comparison and reproducibility across different studies [9]. Finally, generalization remains a challenge. Forecasting models developed for one region may not perform equally well in other geographical areas characterized by different climatic and environmental conditions [11]. Addressing these challenges is essential for developing more robust, scalable, and reliable renewable energy forecasting systems.

8. Research Gaps

This review identifies several research gaps that need further investigation. First, researchers rarely integrate large-scale weather datasets with distributed cloud training. Most studies focus either on model accuracy or on cloud infrastructure, but they do not combine both aspects. Second, the community lacks consistent benchmarking for scalability and computational efficiency. Few studies report metrics such as training time, cost, and energy usage, which prevents fair comparison of distributed forecasting systems. Third, transformer-based models remain under-explored in renewable energy forecasting. Transformers show strong performance in long-sequence prediction, but researchers have not fully tested them in solar and wind forecasting tasks, especially in distributed environments. Fourth, explainable AI methods are largely missing from current forecasting systems. Grid operators need interpretable models to trust predictions and to make informed decisions. Finally, researchers have placed little focus on energy-efficient and green AI model design [15]. Training large models consumes significant energy, which contradicts the sustainability goals of renewable energy research.

9. Future Research Directions

Future research should explore federated learning for geographically distributed renewable plants. Federated learning can help multiple locations train models together without sharing raw data, which improves privacy and reduces data transfer. Researchers should also develop edge–cloud collaborative forecasting systems, where edge devices handle initial data processing and the cloud performs heavy training tasks. This approach can reduce latency and improve real-time performance. Another important direction is the use of explainable AI for grid operators. Interpretable models can help operators understand predictions and make better decisions during weather uncertainty. Researchers should also design energy-efficient deep learning architectures that reduce computational cost without sacrificing accuracy. Finally, green cloud computing strategies should be integrated into forecasting systems to minimize energy consumption and align model training with sustainability goals.

10. Conclusion

Deep learning has significantly improved renewable energy forecasting by capturing complex nonlinear patterns in weather-dependent generation data. This review demonstrates that models such as LSTM, GRU, CNN-LSTM, and transformer-based architectures predict solar and wind power more accurately than traditional statistical methods. These models learn directly from large datasets and reduce reliance on manual feature engineering, enabling better adaptation to sudden weather changes and more reliable grid operation.

Distributed cloud computing plays a critical role in scaling deep learning-based forecasting systems. Cloud platforms provide flexible storage, powerful GPU clusters, and tools for parallel training. Researchers can process large weather datasets and train complex models more efficiently than before. This scalability is essential because weather data originate from multiple sources, including satellites, sensors, and numerical weather prediction systems.

However, several important challenges remain. Many datasets contain missing values and noise, which reduce model accuracy and reliability [6]. Extreme weather events remain difficult to predict, leading to large forecasting errors [1]. Furthermore, deep learning models require substantial computational resources and long training times, especially when using transformer architectures and multi-source data [14].

The literature also indicates that researchers often prioritize accuracy without adequately evaluating scalability, cost, and energy efficiency [5]. Cloud-based training consumes considerable energy, potentially contradicting sustainability objectives [14]. In addition, the lack of standardized benchmark datasets complicates fair comparisons across studies [9]. Many models also struggle to generalize across regions with different climatic and geographical characteristics [11].

Overall, renewable energy forecasting has made significant progress, but integrated solutions that combine deep learning accuracy with cloud scalability and energy efficiency are still needed. Future research should focus on developing benchmark datasets, designing energy-efficient models, and evaluating scalability metrics such as training time and cost. Addressing these gaps will enable the development of forecasting systems that are accurate, scalable, and sustainable, thereby supporting reliable grid operations and accelerating renewable energy integration.

References

Benti, N. E., Chaka, M. D., & Semie, A. G. (2023). Forecasting renewable energy generation with machine learning and deep learning: Current advances and future prospects. *Sustainability*, *15*(9), 7087.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Zhang, L., Zhu, S., Su, S., Chen, X., Yang, Y., & Zhou, B. (2024). Wind power prediction method based on cloud computing and data privacy protection. *Journal of Cloud Computing*, *13*(1), 137.

Heaton, Jeff. "Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016, 800 pp, isbn: 0262035618." *Genetic programming and evolvable machines* 19.1 (2018): 305-307.

Benti, Natei Ermias, Mesfin Diro Chaka, and Addisu Gezahegn Semie. "Forecasting renewable energy generation with machine learning and deep learning: Current advances and future prospects." *Sustainability* 15.9 (2023): 7087.

Chen, Chao-Rong, and Unit Three Kartini. "K-nearest neighbor neural network models for very short-term global solar irradiance forecasting based on meteorological data." *Energies* 10.2 (2017): 186.

Díaz-Robles, L. A., Ortega, J. C., Fu, J. S., Reed, G. D., Chow, J. C., Watson, J. G., & Moncada-Herrera, J. A. (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment*, *42*(35), 8331-8340.

Karaman, Ömer Ali. "Prediction of wind power with machine learning models." *Applied Sciences* 13.20 (2023): 11455.

Xue, Y., Chen, Y., Zhao, J., Liu, X., Li, K., Qiu, W., & Gang, Y. (2015). A review on short-term and ultra-short-term wind power prediction. *Dianli Xitong Zidonghua/Automation of Electric Power Systems*, *39*(6), 1-12.

Patil, E. J. N., Bhavsar, M. P. N., Wagh, M. M. N., & Patil, E. P. S. (2026). Strategic Renewable Energy Pathways for India: Achieving Vision 2047 through Innovation, Policy, and Social Empowerment. innovations, 13(01)..

Zhang, Shuo, Emma Robinson, and Malabika Basu. "Wind power forecasting based on a novel gated recurrent neural network model." *Wind Energy and Engineering Research* 1 (2024): 100004.

Lim, S. C., Huh, J. H., Hong, S. H., Park, C. Y., & Kim, J. C. (2022). Solar power forecasting using CNN-LSTM hybrid model. *Energies*, *15*(21), 8233.

Huang, S., Yan, C., & Qu, Y. (2023). Deep learning model-transformer based wind power forecasting approach. *Frontiers in Energy Research*, *10*, 1055683.

Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, *367*(6481), 984-986.

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green ai. *Communications of the ACM*, *63*(12), 54-63.