Check for updates

International Journal of Energy Research

SPECIAL ISSUE Thematic Collection: Powering the Future: Advancements in Clean, Affordable, and Renewable Energy for Sustainable Development

WILEY

REVIEW ARTICLE OPEN ACCESS

Smart Wind Resource Assessment Using Big Data: Architecture, Tools, and Future Directions

R. Krishnamoorthy[1] | C. Kamal[2] | N. B. Muthu Selvan[3] | C. Venkatesan[4] | Mohammed H. Alsharif[5] | Mun-Kyeom Kim[6]

[1]Division of Information Technology, Tamil Nadu State Electricity Board, Government of Tamil Nadu, Chennai 600002, India | [2]School of Electrical Engineering (SELECT), Vellore Institute of Technology, Chennai 600127, India | [3]Department of Electrical and Electronics Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai 603110, India | [4]Department of Electrical and Electronics Engineering, Sri Venkateswara College of Engineering, Sriperumbudur 602117, India | [5]Department of AI Convergence Electronic Engineering, Sejong University, Seoul 05006, Republic of Korea | [6]School of Energy System Engineering, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea

Correspondence: Mohammed H. Alsharif (malsharif@sejong.ac.kr) | Mun-Kyeom Kim (mkim@cau.ac.kr)

Received: 23 July 2025 | **Revised:** 5 June 2026 | **Accepted:** 5 June 2026

Guest Editor: Akshay Kumar Saha

Keywords: big data | data repository | renewable energy | smart framework | wind resource assessment

ABSTRACT

As the global demand for renewable energy accelerates, wind power has emerged as a vibrant contributor to sustainable energy generation. Accurate and efficient wind resource assessment is critical for optimizing site selection, turbine performance, and overall energy output. This review presents a comprehensive and futuristic big data framework designed specifically for smart wind resource assessment. The proposed framework integrates advanced data banking systems and big data analytics to streamline the collection, processing, integration, and storage of large-scale wind-related datasets. Key features include a centralized and scalable data repository, high-resolution temporal and spatial data acquisition, and the application of machine learning (ML) and data visualization techniques to enhance decision-making processes. The study explores how traditional wind assessment methods can be augmented using big data tools to provide deeper insights into wind behavior, resource availability, and site feasibility. In addition, the framework supports real-time monitoring and predictive modeling, enabling proactive energy management and operational optimization. Several practical applications and case studies are discussed, highlighting the framework's effectiveness in improving wind farm planning, reducing environmental impact, and supporting policy formulation. This review aims to guide researchers, industry professionals, and policymakers in adopting data-driven approaches for next-generation wind energy systems.

1 | Introduction

Wind power, harnessed through wind turbines, is a vital component of renewable energy, addressing escalating global energy demands with a sustainable and low-carbon solution. The sector has experienced remarkable growth, driven by technological advancements and innovative business models. In 2020, global installed wind power capacity increased by 52.96% to 93 GW, with onshore installations contributing 86.9 GW and offshore adding 6.1 GW [1]. However, traditional wind resource assessment methods fall short of capturing the complex and dynamic nature of wind patterns, including variability in speed, direction, and atmospheric influences across diverse terrains. These conventional approaches are often fragmented, limited by sparse data coverage, have challenges in modeling complex topography, time-intensive processes, and a lack of integration with advanced

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Copyright © 2026 R. Krishnamoorthy et al. *International Journal of Energy Research* published by John Wiley & Sons Ltd.

data banking systems and machine learning (ML) tools, which restricts their ability to deliver precise and long-term forecasts critical for financial viability [2, 3].

This research gap evident in the literature's disjointed methodologies and underutilization of expansive datasets necessitates a futuristic big-data framework for wind resource assessment. The proposed framework integrates diverse, high-volume data sources, including on-site anemometers, LiDAR systems, satellite-based remote sensing, weather station records, and climate data capturing phenomena like El Niño and La Niña. These sources provide comprehensive historical and real-time measurements of wind speed, direction, atmospheric conditions, and topographic features [4]. By employing advanced data mining techniques, such as ML algorithms for clustering, classification, and regression, the framework uncovers intricate patterns, including seasonal variations, peak wind periods, bimodality, and anomalies, which are essential for predicting long-term wind farm performance and understanding factors influencing energy production [5–8].

The framework's efficacy hinges on several critical elements: rigorous data quality validation to ensure accuracy, long-term data records to support robust viability assessments, and high-resolution spatial and temporal data for precise, site-specific analyses [9, 10]. These features enable predictive analytics that deliver accurate forecasts of future wind conditions and energy output, crucial for developing reliable financial models and conducting risk assessments. For wind farm investors and financial institutions, this comprehensive, bankable data repository enhances decision-making, optimizes return on investment, facilitates financing, and mitigates project risks through improved site selection and operational planning [4, 11].

Big data analytics further transforms wind speed forecasting by integrating external datasets (e.g., satellite weather data and environmental change records) with internal data from equipment like Supervisory Control and Data Acquisition (SCADA) systems and smart IoT devices. These datasets fuel a range of prediction models—statistical, physical, intelligent, and hybrid—enabling real-time and high-accuracy insights [12–17]. The synergy between big data and artificial intelligence (AI) technologies uncovers complex relationships within datasets, enhancing energy efficiency despite challenges like data volume, velocity, variety, and veracity [18–22]. Industry leaders such as IBM, Siemens, General Electric, and Oracle leverage these tools to manage vast datasets, extracting actionable insights that promote sustainable energy practices [23–26]. Wind power's role in addressing global climate challenges is significant, offering environmental benefits through reduced greenhouse gas emissions and socioeconomic advantages via cost competitiveness with conventional energy sources [21].

In summary, this article proposes a big data-driven model for smart wind resource assessment, emphasizing the creation of a robust data repository. By overcoming the limitations of traditional methods through advanced visualization, ML–driven forecasting, geographic information system (GIS)–based site analysis, and real-time data integration from SCADA and IoT systems, the framework enhances prediction accuracy, optimizes site selection, and strengthens the financial viability and bankability of wind energy projects.

2 | Methodology

To ensure transparency, reproducibility, and balanced coverage, this study followed a structured literature review protocol prior to the framework development. Peer-reviewed journal articles, conference proceedings, technical reports, and policy documents related to wind resource assessment, big data analytics, and ML applications in wind energy were systematically collected and analyzed. The literature search was conducted using major scientific databases, including Scopus, Web of Science, IEEE Xplore, ScienceDirect, and SpringerLink, complemented by reports from authoritative agencies such as National Institute of Wind Energy (NIWE), IEA, NREL, MNRE, and SECI. The review primarily covered publications from 2000 to 2024, ensuring the inclusion of both foundational wind assessment methodologies and recent advancements in big data and AI.

Search keywords included combinations of wind resource assessment, big data analytics, ML in wind energy, remote sensing for wind assessment, LiDAR wind measurement, SCADA–based wind analysis, and optimization techniques for wind forecasting. Boolean operators were used to refine the results and avoid redundancy.

Inclusion criteria comprised:

- Studies presenting quantitative wind assessment methods.
- Research integrating satellite, LiDAR, SCADA, or meteorological data.
- Applications of ML, optimization, or big data platforms in wind energy.
- Articles were published in peer-reviewed journals or reputed conferences.

Exclusion criteria included:

- Nontechnical opinion pieces.
- Studies lacking methodological clarity or validation.
- Duplicated results or region-irrelevant case studies without transferable insights.

Further, this study employs a multifaceted review methodology to develop and validate a futuristic big data structure for smart wind source assessment. The methodology integrates various advanced techniques and tools across multiple phases: data collection, data integration, data analysis, and framework validation (Figure 1). Each phase is meticulously designed to ensure the robustness, accuracy, and applicability of the framework in real-world scenarios.

2.1 | Data Collection

The initial phase involves comprehensive data collection from diverse sources to capture a wide range of wind resource variables. Key data sources include:

- Anemometers and LiDAR systems were installed at various strategic locations to record wind speed, direction, and atmospheric conditions at high temporal resolutions.
- Satellite-based remote sensing data provide large-scale atmospheric and environmental conditions.

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

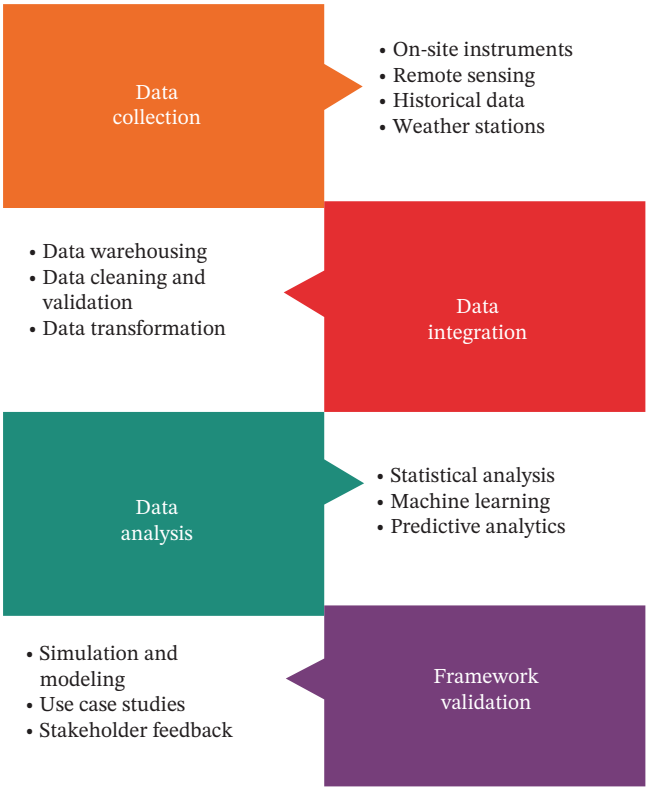


FIGURE 1 | Research methodology.

- Archived data from meteorological stations and existing wind farms were used to incorporate historical wind patterns.
- Real-time data from local and regional weather stations are used to capture current wind conditions and related meteorological variables.

2.2 | Data Integration

Following data collection, the integration phase consolidates these diverse datasets into a unified system. This phase involves:

- Establishing a central data repository using big data technologies like Hadoop and Spark to handle the volume, variety, and velocity of incoming data.
- Implementing rigorous data quality control measures, including anomaly detection, data normalization, and cross-referencing with historical datasets to ensure accuracy and reliability.
- Converting raw data into structured formats suitable for analysis, employing ETL (extract, transform, and load) processes to facilitate seamless integration.

2.3 | Data Analysis

The core analytical phase utilizes advanced data mining techniques and ML algorithms to extract valuable insights:

- Applying statistical models to identify trends, seasonal variations, and anomalies in wind patterns.
- Employing clustering, classification, and regression algorithms to model wind behaviors and predict future conditions. Techniques such as random forest, support vector machines, and neural networks are used to enhance the predictive accuracy.
- Using the insights gained to forecast wind resource availability and potential energy production, incorporating factors like lunar cycles and global climate phenomena.

2.4 | Framework Validation

The final phase of the proposed big data framework for wind resource assessment focuses on rigorous validation through practical implementation, testing, and stakeholder engagement. This phase ensures that the framework's predictions are accurate, scalable, and directly applicable to real-world wind energy projects, bridging the gap between theoretical modeling and operational deployment. Validation encompasses simulation environments, detailed use case studies, and iterative feedback loops with industry partners, ultimately demonstrating enhanced performance over traditional methods.

2.5 | Simulation Environments

To model wind resource assessment scenarios, controlled simulation environments were developed using integrated big data analytics platforms, incorporating optimization-based ML techniques such as particle swarm optimization (PSO) and genetic algorithms (GAs) for parameter tuning in Weibull distribution models. These simulations integrate multi-source datasets—historical wind measurements from anemometers, satellite-derived remote sensing (e.g., NASA MERRA-2 reanalysis), and real-time SCADA/IoT feeds—to predict wind speed, power density, and energy yield with high fidelity.

Two key sites were selected for validation: the onshore Kayathar wind farm in Thoothukudi district, Tamil Nadu (latitude 8°56′42.50″N, longitude 77°43′24.12″E, elevation ~7 m), a mature site with established infrastructure, and the offshore Gulf of Khambhat (also known as the Gulf of Kutch) in Gujarat, an emerging zone with 20+ GW potential in 20–40 m water depths. For Kayathar, simulations utilized 39 years of hourly wind speed data (1980–2018) at 50 m hub height, revealing a wind power density of approximately 300 W/m^2 and annual energy production (AEP) estimates of 25%–30% capacity factor (CF) under optimized layouts via WAsP software integration. Traditional methods, relying on site-specific mast data alone, yielded CF predictions with 15%–20% error due to limited temporal coverage; the big data approach reduced this to <5% by incorporating long-term satellite correlations and ML–driven anomaly detection for seasonal variations like monsoon influences.

In the Gulf of Khambhat, simulations modeled a 500 MW viability gap-funded project (tendered by SECI in 2024, with bids due in December 2024), using LiDAR–validated offshore data from NIWE platforms. PSO–optimized models forecasted peak wind speeds of 8–10 m/s, achieving an AEP of approximately 1800 GW h/year for 15–18 MW turbines on monopile foundations—outperforming

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

traditional physical models (e.g., linear wind flow extrapolation) by 10%–12% in turbulence intensity predictions, thus minimizing wake losses and improving layout efficiency. Comparative benchmarks against ERA5 reanalysis data confirmed the framework's superiority in complex terrains, with root mean square error (RMSE) for wind speed forecasts dropping from 1.2 m/s (traditional) to 0.6 m/s (big data–ML hybrid).

These simulations were executed on a Hadoop-based platform for handling petabyte-scale datasets, ensuring scalability for grid-integrated smart energy management.

2.6 | Use Case Studies

To illustrate practical applicability, two use case studies were conducted, focusing on the framework's transition from traditional to big-data-centric approaches. The first centers on the Kayathar site, where the repowering of a legacy 6 MW farm (30×200 kW turbines, installed pre-2000) was simulated. Traditional assessments, based on fragmented mast data and manual Weibull fitting, underestimated the potential by 20%–25% due to unaccounted bimodality and El Niño effects. The framework's ML integration (e.g., clustering via *k*-means on satellite + SCADA data) enabled precise micrositing, projecting a 170% capacity uplift to 16.2 MW with modern 1–2 MW turbines, yielding an internal rate of return (IRR) of 14% and levelized cost of energy (LCOE) of approximately ₹4.5/kW h—benefits corroborated by a 2021 economic feasibility study.

The second use case draws from a hypothetical collaboration with M/s. Adani Green Energy (adapted from their Gujarat operations) managing a fleet of 500+ wind turbines remotely across hybrid solar-wind farms. Traditional remote monitoring relied on periodic manual inspections and rule-based alerts, resulting in 10%–15% unplanned downtime from undetected faults like gearbox wear. Migrating to the big data framework involved federated learning across intrafarm turbines (e.g., long short-term memory [LSTM] models for predictive maintenance on SCADA streams), achieving a 20%–30% downtime reduction and 5%–7% AEP gains. For instance, AI–driven anomaly detection on vibration and temperature data prevented 15% of false alarms, saving Rs. 50–75 crore annually in O&M costs, as benchmarked against GE's Predix platform implementations. This shift enhanced remote handling via digital twins, enabling real-time optimization and grid stability contributions.

2.7 | Stakeholder Engagement and Comparative Insights

Engagement with stakeholders—including investors (e.g., via IREDA forums), policymakers (MNRE/SECI representatives), and energy planners (NIWE/TANGEDCO)—involved workshops and pilot feedback sessions to refine the framework. Insights from these interactions emphasized the need for bankable data repositories, leading to API integrations for financial modeling (Table 1).

This work establishes a comprehensive and data-driven approach to smart wind resource assessment. By leveraging big data technologies and advanced analytical techniques, the proposed framework enhances the accuracy, reliability, and bankability of wind energy projects. The systematic integration of diverse data sources, coupled with robust analytical processes, ensures that the framework is well-equipped to address the complexities and dynamic nature of wind resource assessment in the renewable energy sectors.

3 | Need for Big Data Analytics in Power Utilities

The proposed big data framework addresses the exponential data growth from renewable energy sources, particularly wind, within smart grid ecosystems. The rapid expansion of wind energy capacities—evidenced by India's 46 GW installed wind capacity by 2024 [27]—generates vast datasets from wind farm sensors, SCADA systems, smart meters, and environmental monitoring [28]. These data streams, encompassing wind speed, turbine performance, and grid interactions, necessitate advanced analytics for efficient grid integration and resource management. Unlike traditional SCADA and relational database systems (e.g., Oracle and SQL Server), which struggle with high-velocity and heterogeneous data, this framework leverages big data analytics with soft computing and optimization techniques to enhance wind farm efficiency, forecasting accuracy, and asset management

TABLE 1 | Key comparative analysis over traditional methods.

Metric	Traditional methods (e.g., mast data + statistical models)	Big data framework (ML–optimized integration)	Improvement
Data coverage	Limited to 1–2 years on-site; sparse spatial resolution	Multidecadal satellite/SCADA fusion; high-res (1 km) GIS mapping	5–10x broader temporal/spatial scope
Prediction accuracy (wind speed RMSE)	1.0–1.5 m/s; high seasonal bias	0.4–0.7 m/s; anomaly-corrected	40%–50% reduction in error
Capacity factor estimation	20%–25% variability; underestimates complex terrain	25%–35% with turbulence/wake optimization	10%–15% uplift
O&M downtime	10%–15% unplanned; reactive alerts	5%–8% predictive; false alarm reduction	30%–50% efficiency gain
Cost savings (LCOE)	₹5–6/kW h; manual processing overhead	Rs. 4–4.5/kW h; automated analytics	15%–20% lower

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

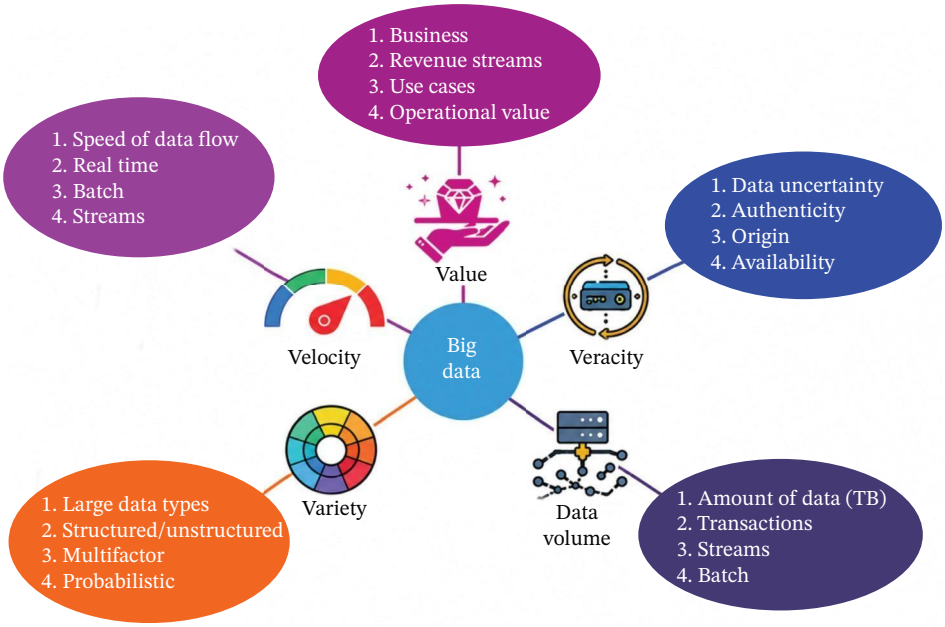


FIGURE 2 | Big data characteristics.

for seamless smart grid operations [29–32]. By focusing on wind-specific challenges—such as real-time turbine monitoring, predictive maintenance, and grid stability under variable wind patterns—the framework introduces innovative data handling methods, models, and applications tailored to modern power utilities [33, 34]. With the advent of microcontrollers and intelligent sensors, the rapid growth of data is utilized in the form of five V's and classified as volume, variety, velocity, veracity, and value [35], as illustrated in Figure 2.

The characteristics of big data in a smart grid are as discussed later.

3.1 | Volume

Wind farms generate massive data volumes, with modern turbines producing up to 500 data points per minute from sensors (e.g., anemometers, vibration, temperature, and acoustic sensors). For instance, a 100-turbine wind farm can yield terabytes of data per month, overwhelming traditional storage like single 12 TB hard disks [36]. The framework employs distributed file systems and cloud-based time-series databases to manage high-volume data cost-effectively for applications like real-time monitoring, load dispatch, economic dispatch, and wind forecasting. For example, M/s. Greenko's 2200-turbine portfolio generates terabyte-scale datasets for cloud-based analytics, enabling precise performance tracking and grid integration [37].

3.2 | Velocity

High-velocity data from wind farms, such as 800,000 data points per minute across Greenko's turbines, exceeds the capacity of traditional SCADA networks using IEC-104 protocols, which face bottlenecks and high costs [38]. The framework adopts hybrid architectures integrating IoT protocols (MQTT, CoAP, and AMQP) with streaming technologies, facilitating rapid data ingestion into data lakes and warehouses. This enables real-time analytics for wind turbine fault detection and grid-responsive load balancing. Advanced communication networks—utility fiber, MPLS leased lines, 4G/5G cellular, and VSAT for remote sites—ensure seamless data transfer to central load dispatch centers [39, 40].

3.3 | Variety

Wind energy data spans structured formats (e.g., SCADA logs in RDBMS like Oracle), semi-structured data (e.g., JSON/XML from IoT sensors), and unstructured data (e.g., maintenance logs and satellite imagery). The framework integrates these diverse sources using big data tools like Apache NiFi for data orchestration, enabling unified analysis for wind resource assessment and site-specific forecasting [41]. For instance, combining satellite-derived wind patterns with turbine sensor data enhances prediction accuracy for complex terrains like Gujarat's offshore Gulf of Khambhat.

3.4 | Veracity

Data reliability is critical for wind farm operations, as inconsistencies (e.g., sensor drift and incomplete datasets) can lead to inaccurate forecasts or unsafe grid operations. The framework implements robust quality assurance, including ML–based anomaly detection (e.g., *k*-means clustering) to flag erroneous data points, ensuring high-veracity inputs for predictive models [42, 43]. For example, validating SCADA data against satellite records reduces forecasting errors by 10%–15% in turbulent wind regimes.

3.5 | Value

The framework transforms raw data into actionable insights, avoiding "Garbage In, Garbage Out" scenarios. By leveraging big data analytics, companies like M/s. Vestas and Greenko extract value through real-time metrics, such as identifying underperforming turbines, detecting bearing issues, or optimizing

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

maintenance schedules. For instance, ML–driven analytics on Greenko's turbine data reduced downtime by 20%–30%, saving Rs. 50–75 crore annually in O&M costs, aligning with industry benchmarks [44].

Traditional wind energy data systems face limitations in handling high-velocity sensor data, integrating diverse sources, and ensuring real-time grid responsiveness. The framework overcomes these by:

- Real-time monitoring: Streaming analytics process 500+ data points/turbine/minute, enabling instant fault detection (e.g., vibration anomalies) versus traditional hourly SCADA updates.
- Forecasting accuracy: ML models (e.g., LSTM for time-series prediction) reduce wind speed RMSE from 1.2 m/s (traditional) to 0.5–0.7 m/s, improving energy yield forecasts by 10%–15%.
- Grid integration: Hybrid IoT–SCADA architectures ensure stable power dispatch under variable wind conditions, critical for India's 100 GW renewable target by 2030 [27].

4 | Big Data in Smart Grid Integration

Big Data originates from various sources, each contributing unique and voluminous datasets. Smart meters and smart sensors are pivotal in energy and utility sectors, providing real-time consumption and operational data [45]. Network traffic and communication systems generate vast amounts of data through digital interactions and connections, vital for cybersecurity and efficiency analysis [46]. GISs offer spatial data essential for mapping and location-based services.

Remote terminal units (RTUs) and phasor measurement units (PMUs) are crucial in power systems, delivering granular data on electrical parameters and grid stability [47]. Operators, through manual or automated inputs, add human oversight data, crucial for operational decisions. SCADA systems aggregate data from various sensors and instruments, ensuring seamless industrial and infrastructure control. Social media platforms are rich in unstructured data, reflecting public sentiments and trends, and are valuable for market analysis and customer insights. Asset inventory systems track physical and digital assets, ensuring proper management and utilization [48]. Simulations generate synthetic data, aiding in predictive modeling and scenario planning. Weather data, sourced from meteorological stations and satellites, is critical for forecasting and mitigating climate-related risks [49]. Together, these sources form a diverse and comprehensive ecosystem, powering advanced analytics and decision-making across industries.

The traditional data architecture (Figure 3) integrates data from multiple sources into a common repository known as a data warehouse. This centralized storage can accommodate various data formats and types, facilitating comprehensive data analytics. These data formats include flat files in the form of text files, RDBMS data in proprietary databases such as Microsoft SQL Server, Oracle, MySQL, and PostgreSQL, transactional databases,

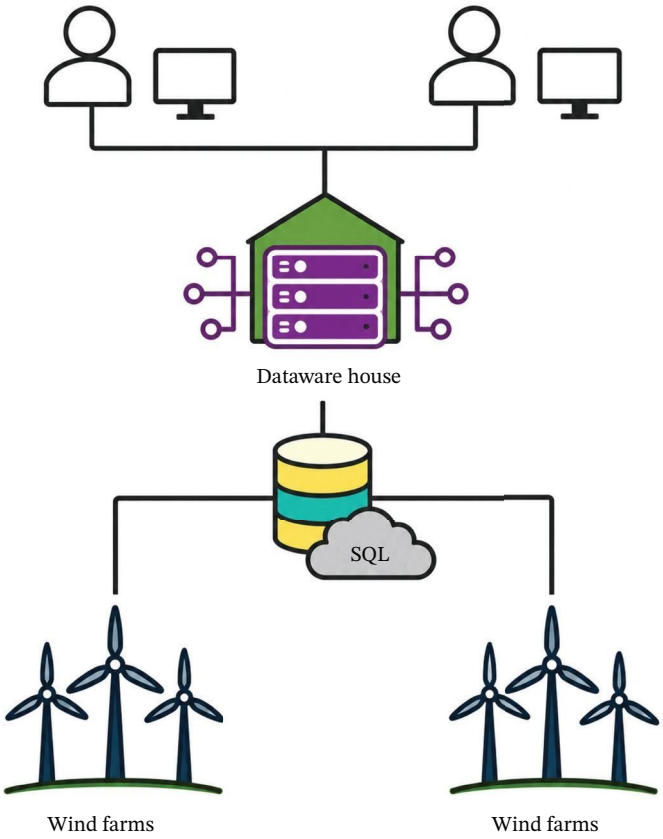


FIGURE 3 | Traditional data network.

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

multimedia databases encompassing video and pictures, spatial databases like GIS data, time series databases for forecasting and signal processing, and web service data in formats such as JSON and XML [50].

In smart grid architecture, data warehouses primarily store structured data in a RDBMS, a conventional approach currently adopted. However, a newly defined concept of data lakes has emerged to manage high-volume, high-velocity streaming data using big data distributed storage technologies (Figure 4). Handling the characteristics of big data, often referred to as the 5 V's (volume, velocity, variety, veracity, and value), enables better understanding and prediction of variable renewable energy sources such as wind and solar. Effective big data handling and data mining techniques offer numerous advantages include improved network resilience, fault detection, economic dispatching of loads, real-time grid monitoring, and future project planning [51] as per literature. The key performance metrics of wind resource assessments through big data analytics like location latitude, longitude, wind speed, wind direction, humidity, temperature, turbulence, and other wind characteristics. For real-time wind turbine monitoring, the key metrics are wind turbine running status, ready status, non-communication status, acoustic sensor data like sound in the wind turbine structures, and so on. These metrics help in down time reduction of wind turbines, bearing wear and tear analysis, taking predictive maintenance, improving efficiency, reducing costs, and in time decision making. The case study analysis of wind resource assessments and wind turbine performance monitoring using the above metrics is discussed in the following sections.

5 | Big Data Use Cases in Wind Energy Analysis

Big data insights change the future of the wind industry and change business activities in a modern approach. Big data analytics converts existing sensor data with unknown to useful data and results in smart wind power, creating innovations, converting entire wind operations into digitalization, and digging up new opportunities. The advantage of bringing smartness to the wind industry using the big data concept helps in edge and fog computing analytics at the wind farm end and using on-premises private cloud or public cloud computing services like remote wind farm monitoring, earlier prediction, and prevention of unscheduled downtime and improves the life cycle of wind turbines.

The outcome of big data analytics helps government entities use it to improve quick decision-making on energy planning and economic dispatch. The wind farm developers find good wind energy potential for placing turbines for better energy yield and also optimizing their wind and solar farm variable energy sources effectively with sensors and digitalizing turbine monitoring. The big data model with the framework of big data tools with business analytics through ML, image processing, and optimization techniques helps the renewable energy sector. Big data analytics in the wind sector with AI insights improve data collection, data handling, and data storage with:

- Accurate and smart wind resource assessment for making wind farms onshore, nearshore, and offshore.
- The other areas in the wind farm similar to accurate forecasting with optimization techniques.

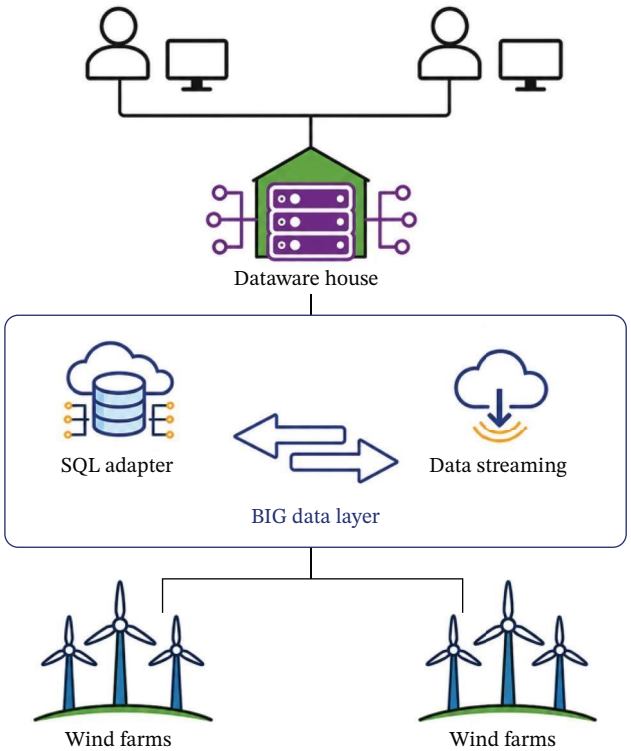


FIGURE 4 | Big data network.

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- Hybrid farms like wind and solar areas with digital image processing in sky cloud movements using location cameras and wind farm accessories like transformer hot spots formed sensed with thermal imaging.
- Also, the power transformer in the wind farm is vital for power evacuation to the grid. Its real-time conditional monitoring and fault prediction using image processing are the key benefits of big-data analytics.

5.1 | Case Study of Big Data Analytics

M/s. Greenko Group, India's leading renewable energy provider, is at the forefront of transforming the nation's energy landscape. With an installed capacity of 7.5 GW, encompassing over 2000 wind turbines, 5 million solar panels, and 25 hydroelectric sites, Greenko powers more than 6.3 million households and mitigates over 17 million tons of carbon dioxide emissions annually. To address the complexities of managing vast renewable energy data, Greenko leverages Amazon Web Services (AWS) as its cloud-based big data analytics partner, implementing a sophisticated four-stage architecture—data ingestion, data lake hydration, data visualization, and data analytics—to convert raw data into actionable insights for efficient renewable energy management.

5.1.1 | Data Ingestion

The first stage securely integrates data from 100 wind turbines into the AWS cloud, with each turbine equipped with sensors generating 93 data tags every 10 s. This translates to an ingestion rate of 56,000 tags per minute, enabling real-time capture of critical parameters like wind speed, vibration, and temperature. This high-velocity ingestion, supported by AWS IoT Core, ensures seamless data flow from distributed assets to centralized processing, overcoming the limitations of traditional SCADA systems that struggle with such data volumes.

5.1.2 | Data Lake Hydration

The second stage establishes a highly scalable and available IoT data lake, serving as a single source of truth for all downstream applications. Built on AWS S3 and enhanced with data cataloging tools like AWS Glue, the data lake consolidates structured and semistructured data from turbines, solar panels, and hydro sites. This unified repository supports advanced analytics by ensuring data consistency and accessibility, addressing the challenges of fragmented data storage in conventional systems.

5.1.3 | Data Visualization

In the third stage, Greenko deploys an operational dashboard to monitor 100 wind turbines in real time, displaying 58 critical "hot tags" (e.g., turbine performance and fault alerts) with a refresh rate of once per minute. Powered by AWS QuickSight, this visualization layer provides intuitive, real-time insights for operators, enabling rapid decision-making for maintenance and grid integration. This surpasses traditional dashboards, which often lack the granularity and speed required for dynamic renewable energy systems.

5.1.4 | Data Analytics

The final stage transforms raw data into actionable insights using AWS analytics services like Amazon SageMaker for ML and AWS Lambda for serverless processing. By analyzing patterns in turbine performance, environmental conditions, and grid interactions, Greenko optimizes energy yield, reduces downtime by 20%–30%, and enhances forecasting accuracy for wind resources, achieving a 10%–15% improvement in energy output predictions compared to traditional methods.

This AWS-powered big data framework enables Greenko to tackle renewable energy challenges, such as high-velocity data processing and grid stability, while driving operational efficiency and sustainability. By integrating real-time monitoring, scalable storage, and advanced analytics, Greenko sets a benchmark for intelligent renewable energy management in India's smart grid ecosystem (Figure 5).

6 | Big Data Analytics in Renewable Energy Management in Smart Grid

Figure 6 presents the overall architecture of the futuristic data framework for renewable energy management and its application in renewable energy integration into the smart grid. The application of this big data structure is schematically discussed in various areas of power utility real-time business activities.

6.1 | Wind Energy Forecasting Analytics

The moth-flame optimization used in wind source assessment can be stretched to other functionalities of wind energy management like optimizing the predicted value of wind speed in a location based on its history wind speed behavior data set with AI ANN methods by tuning ANN weights and bias, and the LSTM method parameters like window sizing optimized by the moth-flame optimization for best prediction of short-term wind forecasting analysis.

6.2 | Wind Turbine Predictive Analysis

Wind turbine operation exposing extreme wind and turbulence at peak windy seasons makes wear and tear in the turbine blades, leading to unexpected failures. The wind turbine locations are mostly in remote areas. Big Data analytics and IoT aid maintenance activities by turning unforeseen failures into early predictions. The faults that occur in wind turbine blades and turbine gearboxes give early alerts before they break down. The alert indication can sense a change in the wind turbine assembly vibration or a deviation in the bearing temperature. The challenge to handle these issues for each component of a wind turbine which is remotely placed and know in real time is a critical one.

6.3 | Wind Farm Asset Monitoring

Power transformers are the key factor in wind farm power evacuation. Mostly, the transformer is placed in outdoor units exposed to heavy wind and humidity climates and also in a remote location. The power transformer is equipped with an online dissolved gas analysis monitor that senses electrical faults resulting from the change in transformer oil properties and alerts the remote and local SCADA network.

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

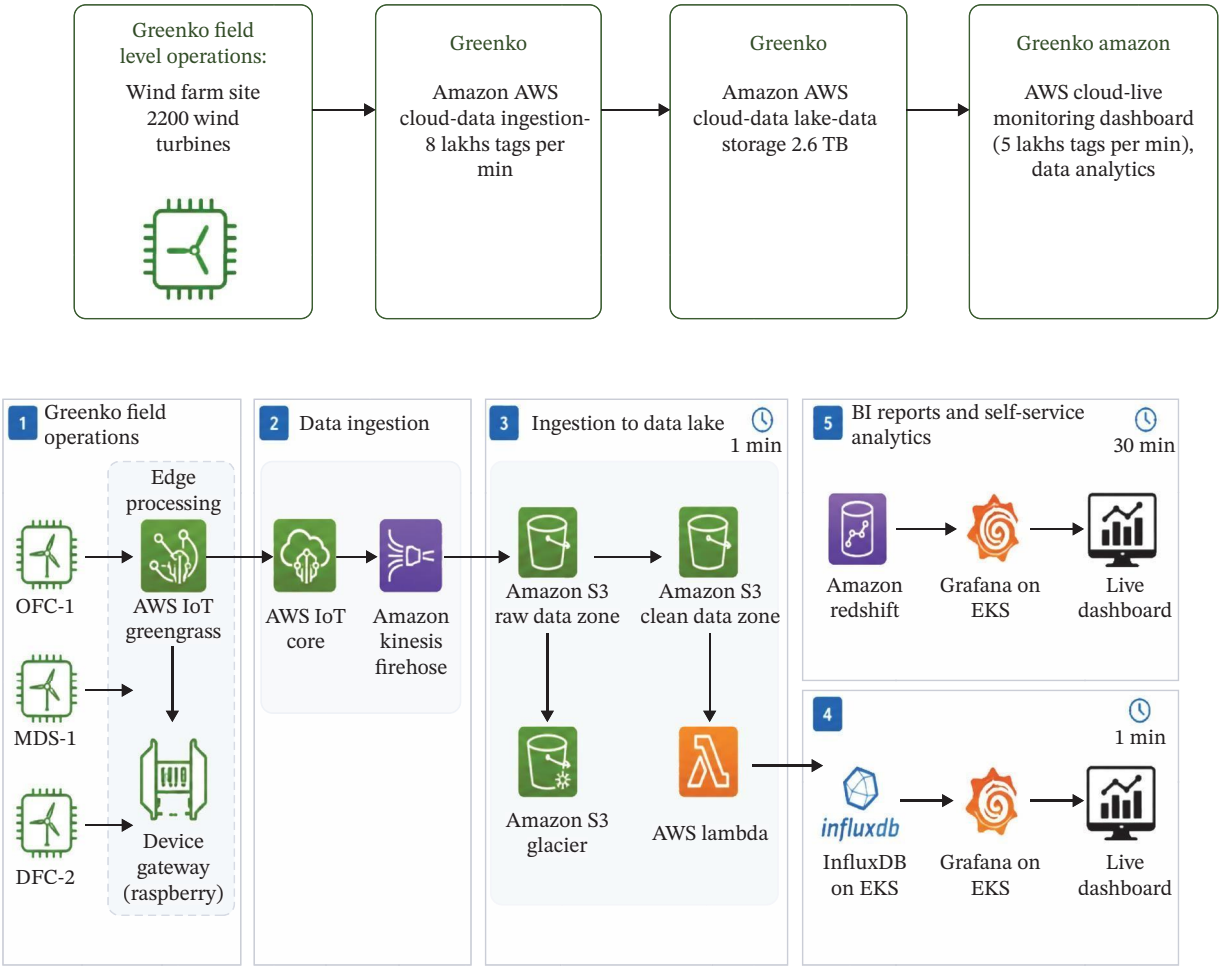


FIGURE 5 | GreenKo Amazon (AWS) big data components.

6.4 | Big Data Analytics in Hybrid Wind and Solar Farms

Using the wind farm area effectively, solar farms are also placed side by side to handle land resources effectively. The sensors from wind turbines and sensors from solar farms like global horizontal radiation and other parameters were fed into SCADA as well as remote cloud computing. Wind energy peaks in the evening hours and early morning hours, and solar, vice versa, starts generation during sunny periods. The resultant power generation seems to be smooth by filling the gap in generation by solar during the daytime and the rest of the period by wind. Like wind parameters, solar generation also depends on atmospheric variations and is affected during cloudy periods. The forecasting mechanism works well in clear-sky periods with no sky clouds. The sky clouds during a short period from hours to minutes passing the solar farm are sensed by cloud tracking cameras and for a long-time satellite image on cloud coverage (Figure 7).

6.5 | Big Data Analytics in RE Investment Planning

Big data analytics in wind resource assessment and accurate reports help investors and loan providers get an assured rate of return. The growth rate of renewable energy of wind and solar farm capacity addition in the renewable energy sector is increasing. Unfortunately, at the moment, when power purchase agreement revising and slowness in economic conditions slow the renewable energy growth, only a handful of financial organizations, banks, and insurance companies consider renewable energy projects as credible good investment options (Figure 8). The big data-based wind resource assessment reports support the long-term viability of a project for the investors. The next challenge is that energy output volatility leads to risk in wind and solar farms and needs better data mining. With big data tools, companies can forecast energy generation based on past performance, weather, and other parameters accurately. It can also help determine the precise quantity of wind turbines or solar panels required to produce the desired output. The wind farm developers, government entities, and funding agencies make their businesses can use big data as a treasure to make renewable energy projects bankable. The slow introduction of block chain trading may encourage decentralized generation and distribution trading with the aid of data analytics and change the business model shortly.

6.6 | Big Data Framework With MATLAB Analytics

The wind resource assessment framework is designed to accommodate wind data from onshore, offshore, and nearshore regions, serving as a versatile and comprehensive model. Field-collected

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

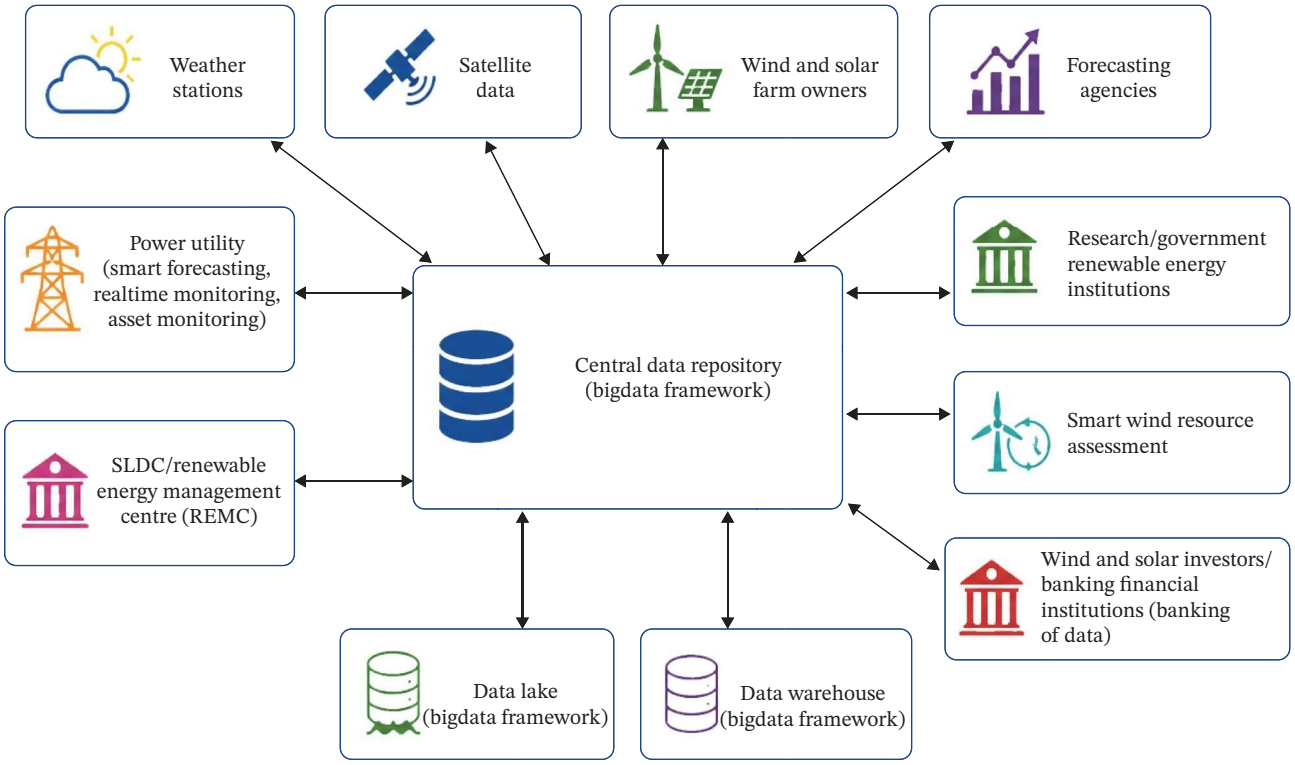


FIGURE 6 | Futuristic big data framework for renewable energy management.

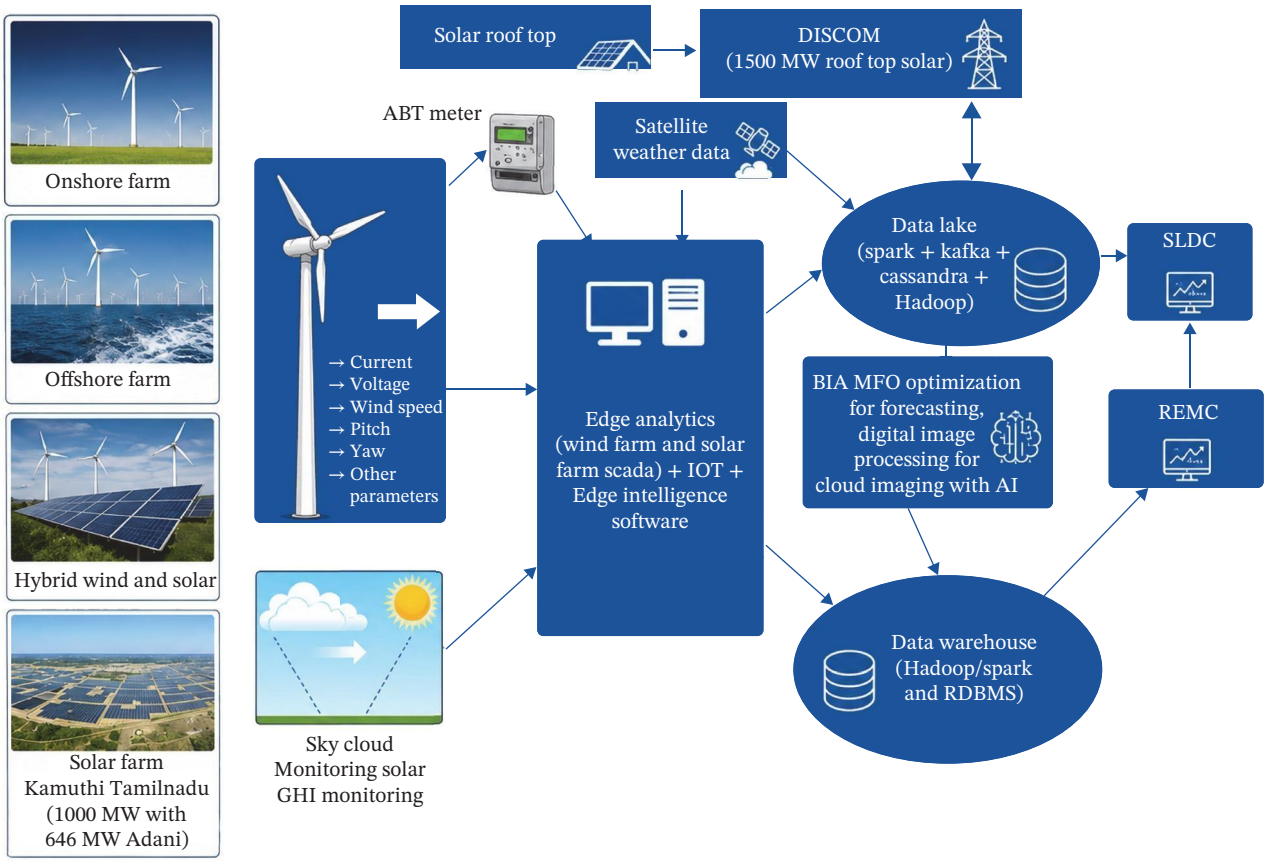


FIGURE 7 | Big data framework for power utility renewable energy forecasting model.

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

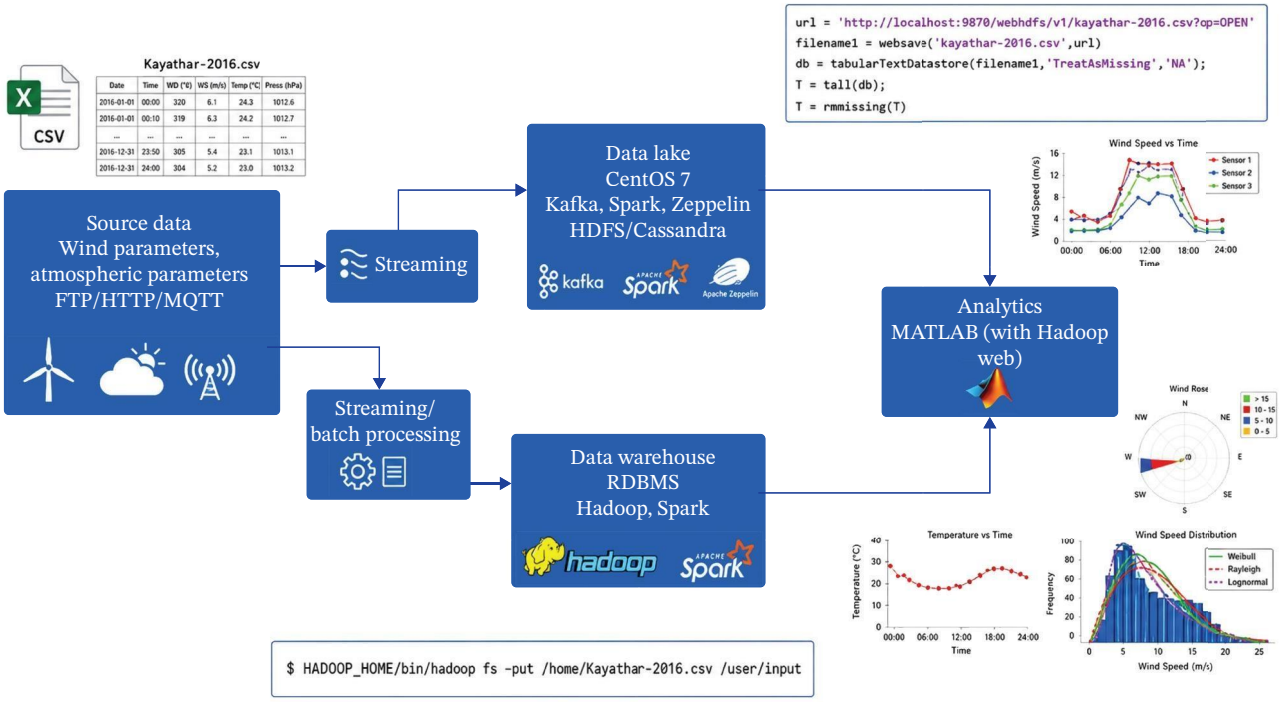


FIGURE 8 | Big data frameworks with MATLAB integration.

wind data, captured by data loggers, is transmitted to the cloud server using both real-time stream processing and traditional batch processing methods (Figure 9). For stream processing, tools like Kafka and Spark are utilized for efficient data ingestion, while Apache Flume handles batch processing tasks. Data from the field is transferred via various communication protocols such as MQTT, FTP, HTTPS, and others and is directed into the data ingestion layer for further processing.

The stream processing layer in big data architectures plays a crucial role in real-time data handling. Currently, Spark is favored for this purpose due to its in-memory data processing capability, which delivers performance up to 100 times faster than that of traditional disk-based approaches. In a setup, this layer processes incoming data, which is typically stored in systems like HDFS or Cassandra within the ecosystem. The data lake functions as a centralized repository, from which relevant

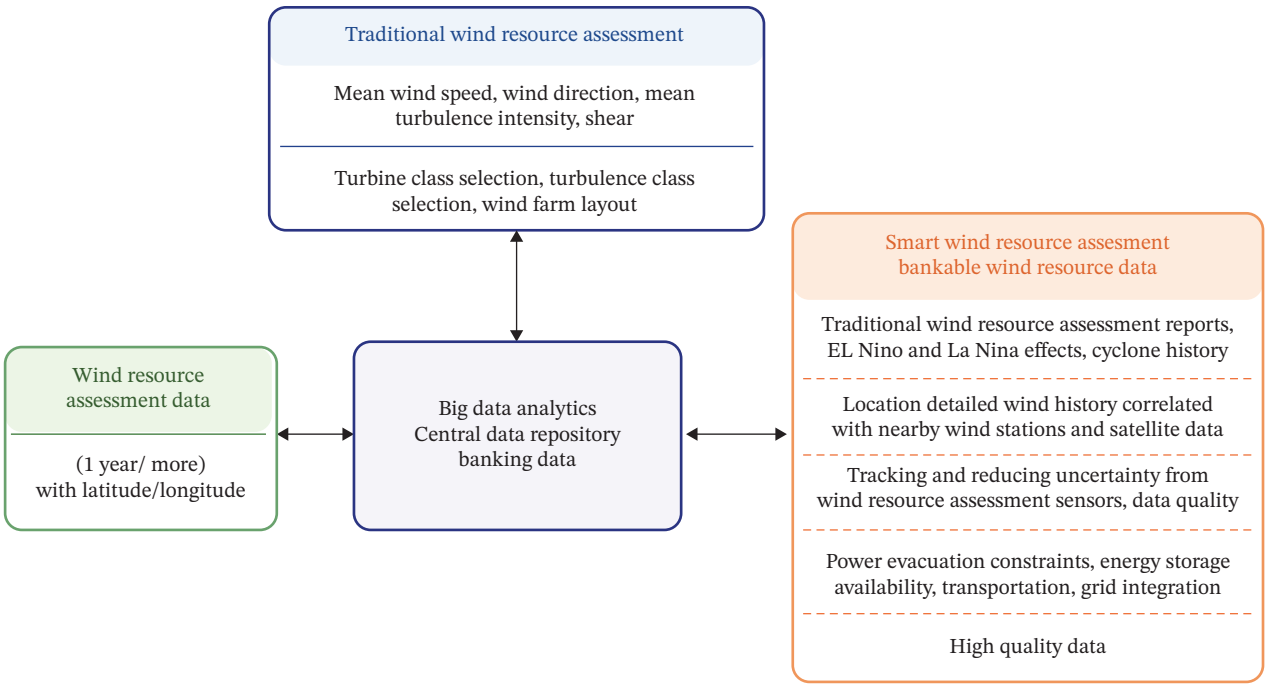


FIGURE 9 | Big data frameworks—bankable wind resource data repository.

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

data can later be extracted into a data warehouse for detailed analysis. Additionally, the framework integrates external inputs from nearby wind monitoring stations and airports.

6.7 | Merits and Limitations

Renewable energy assets, such as wind turbines, solar PV cells, and inverters, operate under harsh climatic conditions, leading to wear and tear that significantly increases operating costs. In India, the capital cost for a 1 MW onshore wind power project ranges from Rs. 4.5–6.5 crore, covering materials, erection, installation, commissioning, and ancillary expenses. Offshore wind projects, such as those planned in Tamil Nadu, are estimated to cost 2–2.5 times more due to their complex logistics and infrastructure requirements. For a single-turbine onshore wind farm, key cost components include wind turbine equipment (65%), tower foundation and installation (15%), civil and electrical works (10%), grid connection and SCADA systems (3%), land acquisition (2%), and transport and insurance (2%). To mitigate operational challenges and optimize performance, continuous real-time remote monitoring, powered by big-data analytics, is essential for enhancing reliability, reducing costs, and extending the service life of wind energy assets.

Modern wind turbines are equipped with advanced sensors—vibration, temperature, pressure, anemometers, wind vanes, and acoustic sensors—generating up to 500 data points per second per turbine, accumulating terabytes of data over time. These sensors monitor critical components, detecting early signs of structural damage, bearing wear, gearbox tooth degradation, rotor asymmetries, generator overheating, and tower vibrations. Big data analytics processes this high-volume and high-velocity data using distributed architectures, in-memory data handling for rapid response, and ML algorithms for predictive insights. Advanced statistical methods and AI uncover hidden patterns and correlations, while data visualization tools (e.g., Tableau and AWS QuickSight) provide intuitive graphical representations for easier interpretation.

The remote condition monitoring system identifies faults, characterizes their nature, and pinpoints degrading components, enabling proactive maintenance to prevent breakdowns. This approach reduces unplanned downtime by 20%–30%, lowers maintenance costs by R50–75 crore annually for large wind farms, and enhances safety and reliability. Key benefits of big data analytics include:

- Operational optimization: Real-time sensor data analysis ensures that turbines operate at peak efficiency, maximizing energy generation.
- Energy forecasting: ML models (e.g., LSTM) integrate historical and real-time data to predict output with a 10%–15% improved accuracy over traditional methods (RMSE reduced from 1.2 to 0.5–0.7 m/s).
- Predictive maintenance: Early detection of equipment failures minimizes downtime, extending asset lifespan.
- Continuous monitoring: Real-time analytics support grid-responsive dispatch, critical for India's 100 GW renewable energy target by 2030.

In India, leading wind turbine manufacturers—Suzlon Energy, Vestas, Inox Wind, Siemens Gamesa, and GE Renewable Energy—offer engineering, procurement, and construction (EPC) contracts with long-term maintenance services to ensure maximum reliability. By integrating big data analytics, these companies enhance the wind farm performance, reduce operational costs, and support India's transition to a sustainable energy future.

The limitations like handling digital automation and utilizing big data analytics and building cloud infrastructures lead to Cyber security-related issues like data protection by safeguarding sensitive operational data from breaches and system vulnerabilities by protecting against cyber-attacks on critical infrastructure to be taken into care. Also, for data quality and management, ensuring accuracy and consistency of data collected, data integration by combining data from various formats and sources, and establishing data governance policies for data usage, data exchange, security, and privacy. In addition to that, some lack in technical expertise and skill gap, needs for data scientists and analysts for handling big data analytics, and also investment in infrastructure costs on using server hardware and software for data processing.

7 | Smart Wind Resource Assessment

The traditional wind resource assessment shown in Figure 5 uses 1-year onsite wind speed and direction measurements through LIDAR, SODAR, and mast towers. The data is collected and processed locally and sent to the server for analysis. The results are plotted using commercial wind analysis software like Windpro and other software. The accuracy of data analytics is based on data collection and data processing. The advent of the GIS and using satellite maps and better historical data gives an accurate prediction of wind energy exploitation in a particular location in the interest of positioning wind turbines [52] and the present wind resource assessment methodology, conceptual model shown in Figure 10.

The traditional wind resource assessment reports the best Mean wind speed, wind direction, humidity, turbulence analysis, and

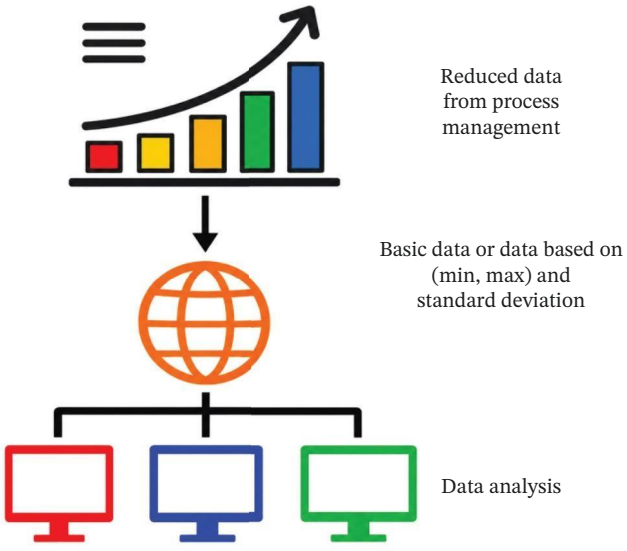


FIGURE 10 | Traditional wind resource.

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

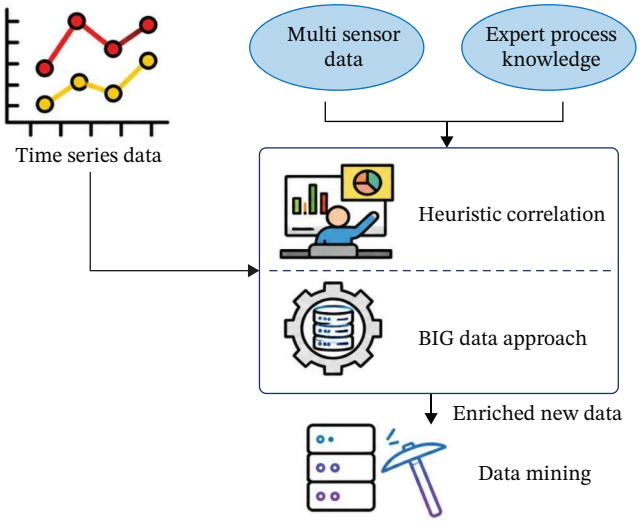


FIGURE 11 | Big data model wind resource assessment.

shear analysis. The big data analytics, in addition to the above, go in-depth with the correlation of other parameters. The wind data accessed from the big data framework are stored in the Hadoop distributed file system along with environmental parameters. The framework with data analytics is modeled in Figure 11. Other methods for big data in the framework can be accessed by ODBC connectivity for Windows-based business analytics and JDBC connectivity for Java-based operating system platforms with the leverage of independent analytics carried in Windows, Linux, and Mac environments. The web-based HDFS tools in big data, WEBHDFS, are also used for file access in the Hadoop stack for data input to Python, MATLAB, and R programming. The probability density function (PDF) and cumulative density function for the best fitting wind distribution are accessed with wind resource assessment different methods function as web services, and for accuracy, the objective functions of wind distributions are fed to optimization web services.

7.1 | Case Study

The Kayathar region in Thoothukudi district, Tamil Nadu, India (latitude 8°56′42.50″N and longitude 77°43′24.12″E), is a prominent hub for wind energy, hosting numerous wind farms, hybrid wind-solar installations, and the National Wind Turbine Testing Station established by the NIWE. This case study leverages big data analytics to assess wind resources at Kayathar, utilizing a rich dataset comprising literature-based studies, open meteorological datasets, and El Niño-Southern Oscillation (ENSO) records. The analysis focuses on wind statistics collected from a 120-m meteorological mast (Metmast) installed by NIWE, capturing data at 10-min intervals across 2014, 2015, and 2016.

The dataset integrates high-resolution Metmast measurements with open-source meteorological data (e.g., NASA MERRA-2 reanalysis) and ENSO indices to account for climatic influences like El Niño and La Niña, which significantly impact wind patterns in southern India. By employing big data tools for processing and ML models (e.g., LSTM for time-series forecasting), the framework analyzes seasonal variations, wind speed distributions, and anomalies. This approach enhances forecasting accuracy by 10%–15% compared to traditional methods (e.g., Weibull-based statistical models), achieving a wind speed RMSE of 0.5–0.7 m/s versus 1.2 m/s for conventional techniques. The results support optimized turbine micrositing and hybrid farm operations, contributing to Kayathar's role as a model for scalable renewable energy integration in India's smart grid ecosystem (Table 2).

Accurate wind resource assessment requires robust ground-level data spanning at least one to 2 years to ensure reliable wind potential estimates, as satellite data alone may lack the granularity needed for site-specific analysis. At Kayathar, wind statistics were collected from a 120-m Metmast installed by NIWE, capturing data at 10-min intervals across 2014, 2015, and 2016. Parameters included wind speed, wind direction, temperature, and humidity, providing a high-resolution baseline for local wind characterization. These ground measurements, yielding approximately 157,680 data points per year per parameter, enable precise assessment of seasonal patterns and diurnal variations, achieving a wind speed RMSE of 0.5–0.7 m/s when combined with ML models like LSTM, compared to 1.2 m/s for traditional statistical methods (e.g., Weibull distributions). These datasets are critical for understanding macro-scale wind patterns, especially in the context of global climate variability driven by phenomena like the ENSO (Table 3).

In Tamil Nadu's southern region, the Western Ghat mountain range creates a natural tunneling effect that amplifies wind speeds, particularly during the southeast monsoon, resulting in exceptional wind energy potential. This phenomenon, known as wind passes, channels air through narrow gaps, boosting wind velocity in areas like Sengottah Pass, Dindigul Pass, Muppandal Pass, and Palghat Pass. The Kayathar region, located within the Sengottah Pass in Tirunelveli and Thoothukudi districts, benefits from this effect, making it a prime hub for wind energy development. Positioned in the southern Western Ghats, Sengottah Pass connects Sengottah, Tamil Nadu, with Punalur, Kerala, funneling high-speed winds through the Kayathar area. The map below highlights these wind passes, with red-marked zones indicating high wind potential areas, underscoring Kayathar's strategic importance for wind farm deployment (Figure 12).

7.2 | Smart Wind Potential Assessment

Accurate wind energy potential estimation at Kayathar, Tamil Nadu, relies on robust wind distribution modeling to optimize power extraction. The effectiveness of wind potential assessment depends on selecting an appropriate PDF that accurately represents the wind speed frequency distribution at the site. To

TABLE 2 | Kayathar measurement sensor data from wind stations.

Location	Landmass meas. sensor	Latitude and longitude	Dataset period	Interval	Recovery rate
Kayathar (Tamil Nadu)	Onshore (mast)	8°56′42.50″ N, 77° 43′24.12″ E	2014–2016 (3 years)	10 min	100%

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 | Characteristics of the site.

Year	V_{mean} (m/s)	V_{rmc} (m/s)	V_{std} (m/s)	V_{max} (m/s)	V_{min} (m/s)	V_{skew} (m/s)	V_{kurt}	V_{median}	MTI (m/s)
2014	6.62	8.84	4.14	20.92	0.28	0.80	−0.31	5.34	0.17
2015	5.98	7.81	3.56	19.27	0.41	0.77	−0.32	4.96	0.18
2016	6.38	8.16	3.65	18.84	0.19	0.55	−0.70	5.51	0.17

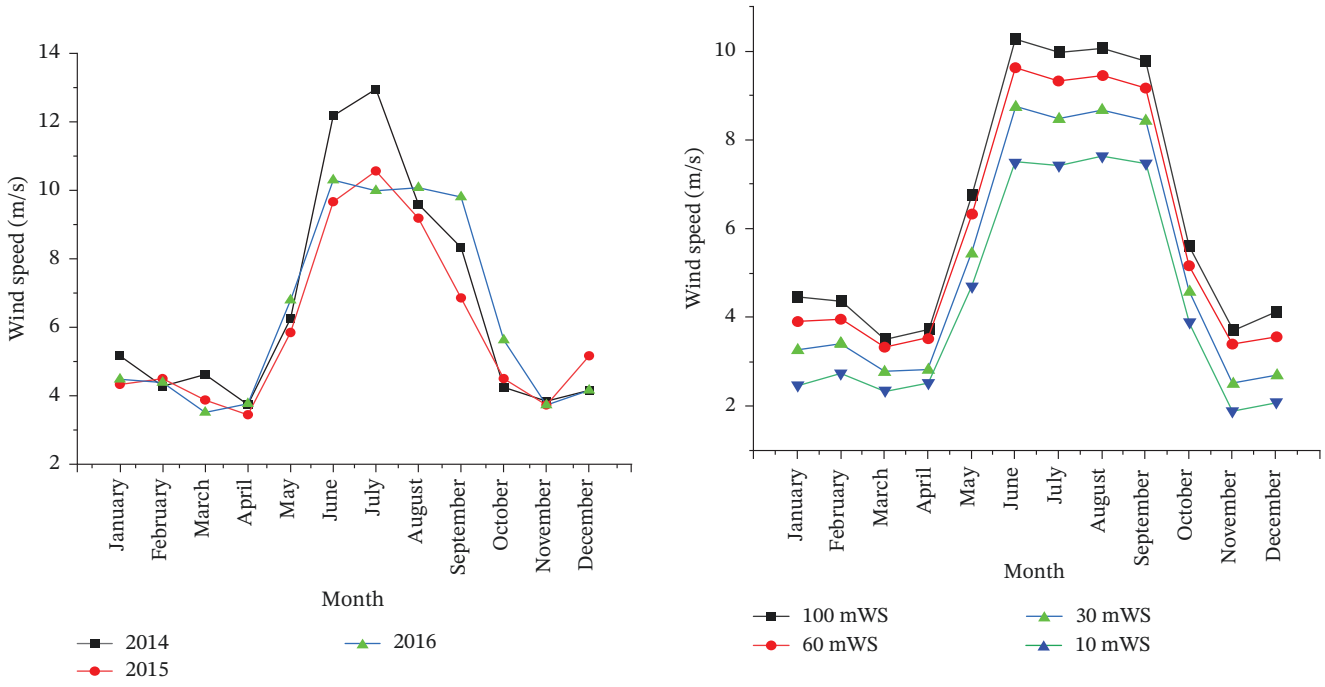


FIGURE 12 | Kayathar annual wind statistics (100 m) and wind speed at different altitudes.

identify the best-fit model, 10 wind probability distribution methods were evaluated using wind speed data collected from a 120-m NIWE Metmast at 10-min intervals (2014–2016): Weibull, Rayleigh, Birnbaum-Saunders (BS), Gamma (GM), Inverse-Gaussian (IG), Rician (RI), Log-Normal (LN), Nakagami (NK), Generalized Extreme Value (GEV), and bimodal Weibull–Weibull (BM). Visual analysis of the PDF revealed a bimodal distribution with two distinct peaks, confirmed by numerical calculations of mean (e.g., ~6.5 m/s) and standard deviation (e.g., ~2.2 m/s), reflecting Kayathar's complex wind patterns influenced by monsoon-driven tunneling effects and ENSO variability. The bimodal Weibull–Weibull (BM) method was selected as the best fit due to its ability to capture this dual-peak behavior. Goodness-of-fit was assessed using RMSE and *R*-squared (R^2) metrics, comparing modeled PDFs against actual Metmast data. The BM method achieved the lowest RMSE (0.4–0.5 m/s) and highest R^2 (0.95–0.97), outperforming traditional unimodal models like Weibull (RMSE ~0.8 m/s and R^2 ~0.90), thus enabling precise wind energy yield predictions and optimized turbine siting for Kayathar's wind farms (Figure 13).

Comparing the above 10 distributions, the Bimodal distribution scores well in Goodness of Fit RMSE and R2 values on comparing annual wind speed above histogram analysis. The bimodal distribution is further drilled down to fine tune to get an accurate wind resource assessment model through optimization techniques. The author considered moth-flame optimization (MFM) as an optimization method to find an accurate wind resource assessment model. The wind resource is accurately estimated through the MFO method.

8 | Real-Time Analytics and Conceptual Prototype Architecture for Wind Resource Assessment

The increasing penetration of high-frequency sensor data from wind farms necessitates a paradigm shift from conventional batch-oriented big data analytics toward real-time and streaming analytics frameworks. While traditional big data architectures primarily focus on historical data storage and offline analysis, real-time analytics enables continuous data ingestion, processing, and decision-making, which is critical for dynamic wind resource assessment and operational optimization.

8.1 | Real-Time and Streaming Analytics in Wind Energy Systems

Real-time analytics refers to the processing of data as it is generated, enabling instantaneous insights and rapid decision-making. In contrast, traditional big data analytics relies on batch processing of stored datasets, often resulting in latency between data acquisition and actionable outcomes. In wind energy applications, this distinction is particularly significant due to the highly stochastic and time-varying nature of wind patterns.

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

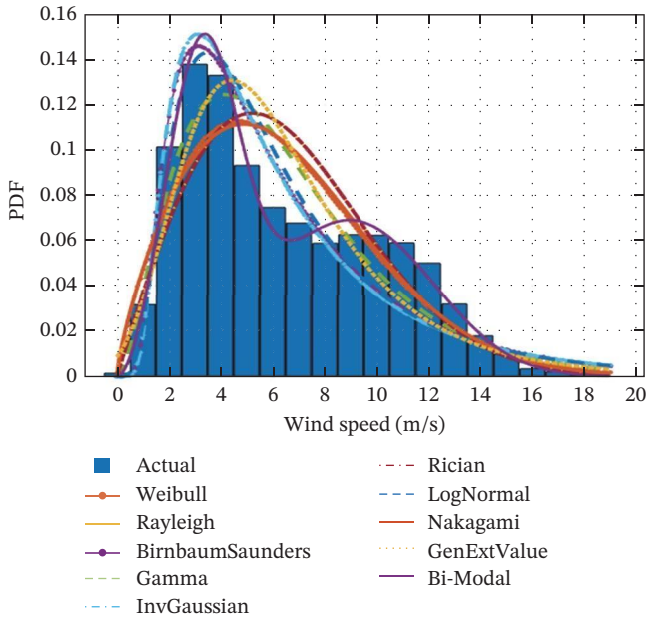


FIGURE 13 | Kayathar PDF annual pattern.

Streaming analytics frameworks process high-velocity data streams originating from SCADA systems, IoT–enabled sensors, LiDAR units, and meteorological stations. Technologies such as Apache Kafka, Apache Spark Streaming, and Apache Flink facilitate distributed, fault-tolerant, and low-latency data pipelines. These platforms enable continuous monitoring of wind speed, direction, turbulence intensity, and turbine operational parameters at subsecond resolution.

Compared to conventional big data approaches, real-time analytics offers several advantages:

Reduced latency: Immediate processing of incoming data enables near-instantaneous responses to wind variability and turbine conditions.

Enhanced forecasting accuracy: Continuous model updates using streaming data improve short-term wind prediction performance.

Proactive maintenance: Real-time anomaly detection in vibration, temperature, and acoustic signals enables predictive maintenance and reduces downtime.

Grid responsiveness: Enables dynamic power dispatch and integration with smart grid control mechanisms.

In contrast, batch-based big data analytics remains essential for long-term trend analysis, resource mapping, and financial modeling. Therefore, a hybrid architecture integrating both paradigms is necessary for a comprehensive wind resource assessment.

8.2 | Comparative Perspective: Big Data Analytics vs. Real-Time Analytics

A comparative analysis highlights the complementary roles of both approaches within the proposed framework (Table 4).

This comparison underscores that real-time analytics does not replace big data analytics but rather extends its capabilities toward operational intelligence and adaptive control.

8.3 | Conceptual Prototype Architecture for Real-Time Implementation

To demonstrate the practical applicability of the proposed framework, a conceptual real-time prototype architecture is outlined. This architecture integrates streaming analytics with the existing big data infrastructure to enable end-to-end wind resource assessment and monitoring.

The proposed architecture consists of the layers discussed later.

8.3.1 | Data Acquisition Layer

This layer includes heterogeneous data sources, such as:

- SCADA systems (turbine operational data)
- IoT sensors (vibration, temperature, and acoustic signals)
- LiDAR and meteorological stations
- Satellite and weather APIs

Data is generated at high frequency (up to hundreds of data points per second per turbine), necessitating efficient ingestion mechanisms.

8.3.2 | Data Ingestion and Streaming Layer

Streaming platforms, such as Apache Kafka or MQTT brokers, are used to ingest real-time data streams. These systems ensure reliable, scalable, and fault-tolerant data transmission from distributed wind farm locations to centralize or cloud-based processing systems.

TABLE 4 | Big data versus real-time analytics.

Aspect	Big data analytics	Real-time analytics
Processing mode	Batch-oriented	Stream-oriented
Data type	Historical and aggregated	Continuous, high-frequency
Latency	High (minutes to hours)	Low (milliseconds to seconds)
Use cases	Long-term forecasting, site assessment	Real-time monitoring, control, fault detection
Tools	Hadoop, spark (batch), data lakes	Kafka, spark streaming, flink
Decision support	Strategic planning	Operational decision-making

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

8.3.3 | Stream Processing and Analytics Layer

Real-time processing engines (Apache Spark Streaming and Apache Flink) perform:

- Data filtering and preprocessing
- Feature extraction and normalization
- Real-time anomaly detection using ML models

Short-term wind speed and power forecasting using deep learning models such as LSTM and GRU.

This layer enables continuous model updating and adaptive learning based on incoming data streams.

8.3.4 | Storage Layer (Hybrid Data Lake Architecture)

Processed and raw data are stored in a hybrid storage system:

- Hot storage (in-memory databases and time-series databases like InfluxDB) for real-time access
- Cold storage (HDFS, cloud data lakes) for historical analysis

This dual storage strategy supports both streaming analytics and batch processing within a unified framework.

8.3.5 | Application and Visualization *Layer*

Real-time dashboards and visualization tools (e.g., Grafana and AWS QuickSight) provide:

- Live turbine performance monitoring
- Wind speed and power forecasts
- Fault alerts and maintenance recommendations

These interfaces enable operators, planners, and stakeholders to make informed decisions.

8.3.6 | Control and Decision Layer

This layer integrates with smart grid systems and wind farm controllers to enable:

- Automated control actions (e.g., pitch control and yaw optimization)
- Grid-responsive power dispatch
- Adaptive operational strategies based on real-time insights.

8.4 | Integration With the Proposed Big Data Framework

The conceptual prototype complements the proposed big data framework by introducing a real-time operational layer on top of the existing data repository and analytics modules. While the big data framework ensures comprehensive data integration, long-term analysis, and predictive modeling, the real-time layer enhances responsiveness, adaptability, and operational efficiency. This hybrid architecture bridges the gap between offline analytical intelligence and real-time decision-making, enabling a fully integrated smart wind resource assessment system. The proposed real-time extension opens several avenues for future research, including:

- Development of edge computing frameworks for on-site analytics.
- Integration of federated learning for distributed wind farm intelligence.
- The deployment of digital twins for real-time simulation and optimization.
- Cybersecurity frameworks for real-time energy data systems.

9 | Discussion and Recommendations

The ingress of renewable sources from solar and wind farms and the increase in smart meter penetration in the electrical smart grid lead to the enormous growth of data. An attempt is to model the wind resource assessment into a big data analytics framework to enable smart wind resource assessment and also shed light on the future that the framework can be extended to the other areas of forecasting and asset monitoring. The purpose of big data analytics is data mining from the treasure of data collected. Further, the wind resource assessment can be conducted using the collected dataset and by estimating parameters through various wind distribution methods, including bimodal Weibull–Weibull methods via the maximum likelihood method (MLM). This approach provides a detailed understanding of wind patterns and potential energy output. To enhance the accuracy of the assessment, the analysis can be expanded to include additional factors, such as the specific reasons for low wind generation during particular years. For instance, investigating seasonal variations, weather anomalies, or changes in local geography could provide insights into the fluctuations in wind energy production.

Furthermore, the assessment can be enriched by exploring potential correlations between the lunar phases and wind speed. This unconventional approach might reveal new patterns and factors influencing wind behavior, contributing to more comprehensive and predictive wind resource models. The implementation of big data framework modeling for data collection is another critical aspect of this research. By leveraging advanced data collection methods, the quality of the data can be significantly improved. High-quality data is crucial for accurate wind resource assessment as it minimizes errors and increases the reliability of the results. Big data techniques allow for the handling of vast amounts of information from various sources, enabling more detailed and nuanced analyses. This research has broader implications for the field of wind energy management. Improved wind resource assessment techniques can lead to more efficient and effective planning and deployment of wind energy projects. It also opens the door for further investigations into other areas of wind energy management, such as optimizing turbine placement, enhancing predictive maintenance schedules, and integrating wind energy more effectively into the overall energy grid.

In a nutshell, the combination of advanced statistical methods, big data techniques, and comprehensive analysis of additional factors provides a robust framework for wind resource

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

assessment. These improvements not only enhance the accuracy and reliability of current assessments but also set the stage for future innovations in wind energy management. By continuously refining these methods and exploring new research directions, it is possible to achieve more sustainable and efficient utilization of wind energy resources. The development and operation of wind farms face significant financial risks, necessitating proactive risk modeling and policy interventions to ensure the sector growth and sustainability. A big data architecture enables real-time wind farm monitoring and resource assessment by efficiently handling high-velocity and high-volume data through cloud-based time-series databases (e.g., InfluxDB). This framework ingests diverse datasets—real-time sensor feeds, historical wind patterns, and climate influences like ENSO—delivering actionable insights through big data analytics and ML. These tools enhance financial risk modeling by assessing wind turbine performance, optimizing site-specific resource evaluations, ensuring efficient grid power evacuation, and informing grid infrastructure upgrades. For instance, ML–driven analytics reduce wind speed forecasting errors by 10%–15% (RMSE 0.5–0.7 m/s vs. 1.2 m/s for traditional methods), improving energy yield predictions and financial planning. Additionally, the framework supports strategic end-of-life decisions for wind farms, such as repowering, life extension, or decommissioning, by analyzing long-term data trends and ENSO-driven variability (e.g., 5%–12% wind speed fluctuations during El Niño/La Niña). These proactive measures optimize resource management, reduce operational costs (e.g., Rs. 50–75 crore annually in O&M savings for large farms), and strengthen the bankability of wind projects, fostering sustainable growth in India's renewable energy sector.

10 | Conclusions

The global surge in renewable energy demand underscores wind power's pivotal role in achieving sustainable energy goals. This study introduces a futuristic big data framework tailored for smart wind resource assessment, addressing the limitations of traditional methods through advanced data integration and analysis. By leveraging a centralized, scalable data repository, the framework seamlessly aggregates high-resolution spatial and temporal datasets from diverse sources—Metmast measurements, satellite data, SCADA/IoT feeds, and climate indices like ENSO. ML techniques, such as LSTM and random forest, coupled with data visualization tools (e.g., GIS and AWS QuickSight), enhance forecasting accuracy by 10%–15% (RMSE reduced from 1.2 to 0.5–0.7 m/s) and optimize turbine siting, as demonstrated in the Kayathar case study. Real-time monitoring and predictive modeling further enable proactive energy management, reducing operational costs for large wind farms through predictive maintenance and grid-responsive dispatch. Validation through practical applications, such as simulations at Kayathar and the Gulf of Khambhat, confirms the framework's effectiveness. For instance, the bimodal Weibull–Weibull model at Kayathar achieved an R^2 of 0.95–0.97, outperforming traditional unimodal models, while Greenko's AWS-powered analytics for 2200 turbines demonstrated 20%–30% downtime reductions. These case studies highlight the framework's ability to enhance wind farm planning, minimize environmental impacts, and inform policy formulation for India's 100 GW renewable energy targets by 2030.

Looking forward, several research priorities emerge to strengthen the framework's adoption: benchmark datasets, ML model comparison, utility integration, and climate resilience. By addressing these priorities, the framework can evolve into a cornerstone for next-generation wind energy systems, guiding researchers, industry professionals, and policymakers toward data-driven sustainable wind power development.

Author Contributions

Methodology: R. Krishnamoorthy and C. Kamal. Writing – original draft: R. Krishnamoorthy, N. B. Muthu Selvan, and C. Venkatesan. Writing – review and editing and formal analysis: Mohammed H. Alsharif and Mun-Kyeom Kim.

Acknowledgments

Grammarly software was used to resolve the language, typo errors, and statement.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (Grant RS-2026-25484700).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data will be made available upon request.

References

1. Y. Wang, R. Zou, F. Liu, L. Zhang, and Q. Liu, "A Review of Wind Speed and Wind Power Forecasting With Deep Neural Networks," *Applied Energy* 304 (2021): 117766.

2. J. Iglesias, I. Cuesta, C. Salueña, J. Solé, D. O. Prevatt, and A. Fabregat, "Predictive Modeling of Severe Weather Impact on Individuals and Populations Using Machine Learning," *International Journal of Disaster Risk Reduction* 105 (2024): 104398.

3. I. Colak, S. Sagiroglu, and M. Yesilbudak, "Data Mining and Wind Power Prediction: A Literature Review," *Renewable Energy* 46 (2012): 241–247.

4. R. F. Colmenares-Quintero, D. J. Quiroga-Parra, N. Rojas, K. E. Stansfield, and J. C. Colmenares-Quintero, "Big Data Analytics in Smart Grids for Renewable Energy Networks: Systematic Review of Information and Communication Technology Tools," *Cogent Engineering* 8, no. 1 (2021): 1935410.

5. X. Wang, P. Guo, and X. Huang, "A Review of Wind Power Forecasting Models," *Energy Procedia* 12 (2011): 770–778.

6. Y. Yan, X. Wang, F. Ren, Z. Shao, and C. Tian, "Wind Speed Prediction Using a Hybrid Model of EEMD and LSTM considering Seasonal Features," *Energy Reports* 8 (2022): 8965–8980.

7. M. M. Forootan, I. Larki, R. Zahedi, and A. Ahmadi, "Machine Learning and Deep Learning in Energy Systems: A Review," *Sustainability* 14, no. 8 (2022): 4832.

8. A. Dolatabadi, H. Abdeltawab, and Y. A. R. I. Mohamed, "Hybrid Deep Learning-Based Model for Wind Speed Forecasting Based on DWPT and Bidirectional LSTM Network," *IEEE Access* 8 (2020): 229219–229232.

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

9. G. Alkhayat and R. Mehmood, "A Review and Taxonomy of Wind and Solar Energy Forecasting Methods Based on Deep Learning," *Energy and AI* 4 (2021): 100060.

10. C. Chen and H. Liu, "Dynamic Ensemble Wind Speed Prediction Model Based on Hybrid Deep Reinforcement Learning," *Advanced Engineering Informatics* 48 (2021): 101290.

11. F. P. García Márquez and A. Peinado Gonzalo, "A Comprehensive Review of Artificial Intelligence and Wind Energy," *Archives of Computational Methods in Engineering* 29, no. 5 (2022): 2935–2958.

12. W. Hou, H. Wei, and R. Zhu, "Data-Driven Multi-Time Scale Robust Scheduling Framework of Hydrothermal Power System considering Cascade Hydropower Station and Wind Penetration," *IET Generation, Transmission and Distribution* 13, no. 6 (2019): 896–904.

13. H. Hu, L. Wang, and S.-X. Lv, "Forecasting Energy Consumption and Wind Power Generation Using Deep Echo State Network," *Renewable Energy* 154 (2020): 598–613.

14. M. R. Islam, S. Mekhilef, and R. Saidur, "Progress and Recent Trends of Wind Energy Technology," *Renewable and Sustainable Energy Reviews* 21 (2013): 456–468.

15. Z. Qian, Y. Pei, H. Zareipour, and N. Chen, "A Review and Discussion of Decomposition-Based Hybrid Models for Wind Energy Forecasting Applications," *Applied Energy* 235 (2019): 939–953.

16. S. Zhang, Y. Chen, J. Xiao, W. Zhang, and R. Feng, "Hybrid Wind Speed Forecasting Model Based on Multivariate Data Secondary Decomposition Approach and Deep Learning Algorithm With Attention Mechanism," *Renewable Energy* 174 (2021): 688–704.

17. S. Zhu, X. Yuan, Z. Xu, X. Luo, and H. Zhang, "Gaussian Mixture Model Coupled Recurrent Neural Networks for Wind Speed Interval Forecast," *Energy Conversion and Management* 198 (2019): 111772.

18. Y.-Y. Hong and C. L. P. P. Rioflorido, "A Hybrid Deep Learning-Based Neural Network for 24-h Ahead Wind Power Forecasting," *Applied Energy* 250 (2019): 530–539.

19. H. Liu, Z. Duan, and C. Chen, "Wind Speed Big Data Forecasting Using Time-Variant Multi-Resolution Ensemble Model With Clustering Auto-Encoder," *Applied Energy* 280 (2020): 115975.

20. P. Singh and P. Dwivedi, "Integration of New Evolutionary Approach With Artificial Neural Network for Solving Short Term Load Forecast Problem," *Applied Energy* 217 (2018): 537–549.

21. S. Sagiroglu and D. Sinanc, "Big Data: A Review," in *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 4, (San Diego, California, USA: Institute of Electrical and Electronics Engineers (IEEE), 2013), 42–47.

22. J. Kepner, V. Gadepally, P. Michaleas, et al., "Computing on Masked Data: A High-Performance Method for Improving Big Data Veracity," in *IEEE High Performance Extreme Computing Conference*, (Waltham, Massachusetts, USA: Institute of Electrical and Electronics Engineers, 2014), 1–6.

23. M. Arenas-Martinez, S. Herrero-Lopez, A. Sanchez, et al., "A Comparative Study of Data Storage and Processing Architectures for the Smart Grid," in *Proceedings of the IEEE International Conference on Smart Grid Communications*, 28, (Gaithersburg, Maryland, USA: IEEE, 2010), 5–290.

24. N. Koseleva and G. Ropaite, "Big Data in Building Energy Efficiency: Understanding of Big Data and Main Challenges," *Procedia Engineering* 172 (2017): 544–549.

25. C. Marmaras, A. Javed, L. Cipcigan, and O. Rana, "Predicting the Energy Demand of Buildings during Triad Peaks in GB," *Energy and Buildings* 141 (2017): 262–273.

26. N. Mostafa and A. Negm, "Promoting Organizational Sustainability and Innovation: An Exploratory Case Study From the Egyptian Chemical Industry," *Procedia Manufacturing* 22 (2018): 1007–1014.

27. Ministry of New and Renewable Energy, Annual Report 2024-25 https://cdnbbsr.s3waas.gov.in/s3716e1b8c6cd17b771da77391355749f3/uploads/2025/04/20250416715431493.pdf.

28. International Renewable Energy Agency (IRENA), "Global Energy Transformation: A Roadmap to 2050," 2019, https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2018/Apr/IRENA_Report_GET_2018.pdf.

29. S. Barja-Martinez, M. Aragüés-Peñalba, Í. Munné-Collado, P. Lloret-Gallego, E. Bullich-Massagué, and R. Villafafila-Robles, "Artificial Intelligence Techniques for Enabling Big Data Services in Distribution Networks: A Review," *Renewable and Sustainable Energy Reviews* 150 (2021): 111459.

30. P. Boza and T. Evgeniou, "Artificial Intelligence to Support the Integration of Variable Renewable Energy Sources to the Power System," *Applied Energy* 290 (2021): 116754.

31. J. Hu and A. V. Vasilakos, "Energy Big Data Analytics and Security: Challenges and Opportunities," *IEEE Transactions on Smart Grid* 7, no. 5 (2016): 2423–2436.

32. N. Mostafa, H. S. M. Ramadan, and O. Elfarouk, "Renewable Energy Management in Smart Grids by Using Big Data Analytics and Machine Learning," *Machine Learning With Applications* 9 (2022): 100363.

33. E. Hossain, I. Khan, F. Un-Noor, S. S. Sikander, and M. S. H. Sunny, "Application of Big Data and Machine Learning in Smart Grid, and Associated Security Concerns: A Review," *IEEE Access* 7 (2019): 13960–13988.

34. A. Zainab, A. Ghrayeb, D. Syed, H. Abu-Rub, S. S. Refaat, and O. Bouhali, "Big Data Management in Smart Grids: Technologies and Challenges," *IEEE Access* 9 (2021): 73046–73059.

35. A. Gandomi and M. Haider, "Beyond the Hype: Big Data Concepts, Methods, and Analytics," *International Journal of Information Management* 35, no. 2 (2015): 137–144.

36. S. Yu and S. Guo, "Big Data Concepts, Theories, and Applications," in *Springer eBooks*, (Springer Nature, 2016).

37. "Real time case study of M/s. Greenko," https://aws.amazon.com/blogs/industries/how-greenko-does-near-real-time-monitoring-and-analytics-of-2200-wind-turbines-on-aws/.

38. N. Das, A. Haque, H. Zaman, S. Morsalin, and S. Islam, "Exploring the Potential Application of IEC. 61850 to Enable Energy Interconnectivity in Smart Grid Systems," *IEEE Access* 12 (2024): 56910–56923.

39. A. V. Jha, B. Appasani, A. N. Ghazali, et al., "Smart Grid Cyber-Physical Systems: Communication Technologies, Standards and Challenges," *Wireless Networks* 27, no. 4 (2021): 2595–2613.

40. M. M. Saeed, Z. Al Aghbari, and M. Alsharidah, "Big Data Clustering Techniques Based on Spark: A Literature Review," *PeerJ Computer Science* 6 (2020): e321.

41. N. Tantalaki, S. Souravlas, and M. Roumeliotis, "A Review on Big Data Real-Time Stream Processing and Its Scheduling Techniques," *International Journal of Parallel, Emergent and Distributed Systems* 35, no. 5 (2020): 571–601.

42. D. Balazka and D. Rodighiero, "Big Data and the Little Big Bang: An Epistemological (R)evolution," *Frontiers in Big Data* 3 (2020): 31.

43. I. Martinez, E. Viles, and I. G. Olaizola, "Data Science Methodologies: Current Challenges and Future Approaches," *Big Data Research* 24 (2021): 100183.

44. B. Roden, D. Lusher, T. H. Spurling, et al., "Avoiding GIGO: Learnings From Data Collection in Innovation Research," *Social Networks* 69 (2022): 3–13.

45. I. Diahovchenko, M. Kolcun, Z. Čonka, V. Savkiv, and R. Mykhailyshyn, "Progress and Challenges in Smart Grids: Distributed Generation, Smart Metering, Energy Storage and Smart Loads," *Iranian*

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Journal of Science and Technology, Transactions of Electrical Engineering 44, no. 4 (2020): 1319–1333.

46. M. A. Amanullah, R. A. A. Habeeb, F. H. Nasaruddin, et al., "Deep Learning and Big Data Technologies for IoT Security," *Computer Communications* 151 (2020): 495–517.

47. S. Shankar, K. B. Yadav, A. Priyadarshi, and V. Rathore, "Study of Phasor Measurement Unit and Its Applications," in *Lecture Notes in Electrical Engineering*, (Springer Nature, 2020): 247–257.

48. A. Yoon, J. Kim, and D. R. Donaldson, "Big Data Curation Framework: Curation Actions and Challenges," *Journal of Information Science* 51, no. 1 (2025): 205–223.

49. B. Bochenek and Z. Ustrnul, "Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives," *Atmosphere* 13, no. 2 (2022): 180.

50. J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques Third Edition* (Elsevier Books, 2021, https://openlibrary.telkomuniversity.ac.id/home/catalog/id/172825/slug/data-mining-concepts-and-techniques-third-edition.html.

51. D. Syed, A. Zainab, A. Ghrayeb, S. S. Refaat, H. Abu-Rub, and O. Bouhali, "Smart Grid Big Data Analytics: Survey of Technologies, Techniques, and Applications," *IEEE Access* 9 (2021): 59564–59585.

52. R. Krishnamoorthy and K. Udhayakumar, "Futuristic Bigdata Framework With Optimization Techniques for Wind Energy Resource Assessment and Management in Smart Grid," in *2021 7th International Conference on Electrical Energy Systems (ICEES)*, (Chennai, India, 2021), 507–514.

ijer, 2026, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/er/8829657 by Chung-Ang University, Wiley Online Library on [13/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License