Predictive Risk Identification Models in Renewable Energy Infrastructure Projects: Mitigating Climate Variability and Equipment Degradation Risks

Authors

Abbas Ahsun, Chris Bass, Carissa Lapan

Date; July 20, 2026

Abstract

The accelerating global energy transition has precipitated large-scale deployment of renewable energy infrastructure, yet these assets face unprecedented operational risks from climate variability and accelerated equipment degradation. Traditional risk assessment methodologies, reliant on periodic inspections and static historical failure data, prove inadequate for capturing the complex, non-linear interactions between environmental stressors and component-level degradation. This study develops and validates a predictive risk identification framework integrating a hybrid CNN-LSTM-Attention deep learning architecture with Space-Air-Ground (SAG) cooperative observation data. The proposed model, trained on operational and meteorological data from solar photovoltaic plants, offshore wind turbines, and hybrid renewable energy systems spanning 2018–2025, achieves an accuracy of 89.4% in predicting high-risk events with a lead time of 48–72 hours. The Differential Significance Criterion (DSC)-enhanced Physics-Informed Neural Network (PINN) component improves interpretability while maintaining predictive fidelity. The framework outperforms conventional static budget-based risk assessment methods, demonstrating statistically significant improvements in early warning capability ($p < 0.01$) and reduction in false alarm rates (from 22.3% to 8.1%). This research provides a replicable methodological foundation for transitioning renewable energy asset management from reactive maintenance paradigms to predictive, risk-informed decision-making, with implications for infrastructure resilience, operational cost reduction, and sustainable energy system reliability.

Keywords: Predictive Risk Identification, Renewable Energy Infrastructure, Climate Variability, Equipment Degradation, Hybrid Deep Learning, Digital Twin, Physics-Informed Neural Networks

1. Introduction

1.1 Background

The global imperative to mitigate climate change has catalyzed an unprecedented expansion of renewable energy infrastructure worldwide. Solar photovoltaic (PV) installations, wind farms—both onshore and offshore—and hybrid renewable energy systems (HRES) have become central pillars of decarbonization strategies, with global renewable capacity reaching record levels . However, the very environmental conditions that make renewable energy attractive—exposure to solar radiation, wind patterns, and marine environments—simultaneously subject these assets to accelerated degradation and heightened operational risks. Unlike conventional thermal power plants with predictable operating envelopes, renewable energy infrastructure operates under highly variable and often extreme environmental conditions that drive equipment failure, performance degradation, and cascading system impacts.

Extreme weather events have demonstrated the acute vulnerability of renewable energy infrastructure. The February 2021 Texas cold wave caused widespread wind turbine icing and generator derating, precipitating cascading grid failures and highlighting the systemic risk posed by climate variability . Similarly, a catastrophic hailstorm in Austin, Texas, in September 2023 damaged approximately 5.5% of identified PV sites across 11,300 installations, as documented by post-event satellite imagery . These incidents illustrate that weather-induced faults, while potentially low-probability events, can trigger severe consequences across interconnected energy systems. The complex network topology of renewable energy systems—encompassing generation, transmission, substation, and energy storage segments distributed over wide geographical areas—means that minor component-level events can propagate into major operational disruptions.

Climate change compounds these risks through projected increases in extreme weather frequency and intensity, as well as shifting climatic baselines that render historical design assumptions obsolete. Research indicates that climate-induced changes in temperature intensity alone could decrease equipment service life by 12–41% by 2060 for vulnerable infrastructure . This temporal dimension of climate risk necessitates predictive approaches that can anticipate both gradual degradation and acute stress events, enabling proactive maintenance and operational adjustments.

Parallel to these environmental challenges, renewable energy assets exhibit complex degradation mechanisms that are inadequately captured by traditional maintenance paradigms. PV inverters, wind turbine gearboxes, battery energy storage systems, and power electronic converters each present unique failure modes influenced by operational history, environmental exposure, and manufacturing variability . Machine learning approaches have demonstrated capability in estimating component reliability under ambient conditions, with studies showing that clustering-regression models can achieve mean squared errors as low as 0.002 for individual inverter datasets when accounting for irradiance, humidity, temperature, and weather conditions . However, these approaches typically focus on component-level reliability rather than system-level risk identification, creating a gap between maintenance analytics and operational risk management.

The convergence of climate-driven environmental stressors and complex equipment degradation pathways creates a pressing need for integrated risk identification frameworks. Conventional risk assessment methods—relying on fixed periodic inspections, historical failure rate tables, and static threshold-based alarm systems—prove fundamentally inadequate for the dynamic, non-linear nature of renewable energy system risks. This research addresses this critical gap by developing predictive risk identification models specifically designed to capture the coupling between climate variability, equipment degradation, and system-level risk propagation.

1.2 Problem Statement

Despite substantial advances in renewable energy monitoring and maintenance technologies, existing risk identification approaches suffer from fundamental limitations that undermine their effectiveness in operational settings.

First, conventional maintenance practices remain predominantly reactive or periodic. Fixed manual overhaul schedules—while providing structure—result in inefficient resource allocation, delayed responses to emerging risk conditions, and incomplete coverage across complex system topographies . The widespread deployment of renewable assets across geographically dispersed locations exacerbates these challenges, as manual inspection and maintenance cannot realistically achieve comprehensive coverage at meaningful temporal frequencies.

Second, existing predictive models typically address either environmental forecasting or equipment degradation in isolation, without accounting for their coupled dynamics. Mechanism-based prognosis techniques, grounded in physics-based mathematical models of degradation, offer interpretability but suffer from computational inefficiency and limited capability to capture high-dimensional, strongly non-linear system behaviors . Conversely, purely data-driven approaches, including time series methods, support vector machines, and graph neural networks, can model complex statistical relationships but often fail to incorporate physical constraints, leading to predictions that violate fundamental physical laws . This disconnect between mechanism-based interpretability and data-driven accuracy represents a persistent methodological challenge.

Third, the uncertainty inherent in renewable resource availability—driven by weather patterns, seasonal variability, and climate change—introduces significant stochasticity into risk assessments. Existing risk evaluation metrics, such as voltage and current violation thresholds, inadequately reflect the specific risk characteristics of renewable-integrated systems, particularly regarding inertia deficiency and frequency regulation capacity . As renewable penetration increases, these limitations become more pronounced, with studies showing that risk identification model performance varies substantially with penetration levels, achieving optimal accuracy only under specific renewable integration scenarios .

Fourth, the spatio-temporal nature of renewable energy system risks demands modeling approaches capable of capturing both spatial propagation patterns and temporal evolution. The complex network topology of HRES, with diverse connected components spanning wide geographical areas, means that localized degradation can cascade through interconnected systems . However, most existing models treat components independently, failing to quantify both individual-component self-risks and multi-component joint-risks.

Fifth, the practical deployment of predictive risk identification faces significant implementation barriers. Resource constraints limit dense ground-sensor deployment, UAV-based inspection is restricted by battery endurance, and satellite-based monitoring suffers from resolution limitations . These monitoring infrastructure constraints create data gaps that must be addressed through intelligent data fusion and modeling approaches that can compensate for incomplete observations.

Finally, the validation and benchmarking of predictive risk identification models remain underdeveloped. Few studies have systematically compared proposed frameworks against rigorous baselines—including conventional static budget-based risk assessment methods—or demonstrated statistically significant improvements in early warning capability and false alarm reduction. This validation gap undermines confidence in research findings and impedes practical adoption.

These limitations collectively point to an unsolved issue: **No validated predictive risk identification framework exists that explicitly models the coupled dynamics of climate variability and equipment degradation within renewable energy infrastructure, while maintaining computational tractability, interpretability, and practical deployment feasibility.** This research addresses this gap through the development and validation of a hybrid deep learning framework that integrates physics-informed constraints with data-driven learning, supported by multi-modal observation data fusion.

1.3 Objectives of the Study

General Objective:
To develop and validate a predictive risk identification framework for renewable energy infrastructure that integrates climate variability and equipment degradation risk modeling, enabling proactive operational risk management.

Specific Objectives:

1. To identify and quantify the key predictors of high-risk events in renewable energy infrastructure, including climatic variables (extreme temperature, wind speed, precipitation, solar irradiance), equipment operational parameters, and component degradation indicators.
2. To design and implement a hybrid deep learning architecture that integrates Convolutional Neural Networks (CNN) for spatial feature extraction, Long Short-Term Memory (LSTM) networks for temporal dependency modeling, and an Attention mechanism for critical pattern enhancement.
3. To integrate physics-based constraints through a Differential Significance Criterion (DSC)-enhanced Physics-Informed Neural Network (PINN) component, ensuring predictions adhere to fundamental physical laws while maintaining data-driven accuracy.
4. To validate the proposed framework against conventional baseline methods using historical operational data from solar PV, wind, and hybrid renewable energy systems, with specific evaluation of prediction accuracy, lead time, and false alarm reduction.
5. To identify and document practical implementation barriers and operational requirements for deploying predictive risk identification models in renewable energy asset management.

1.4 Research Questions

Research Question 1: What combination of climatic, operational, and degradation indicators most accurately predicts high-risk events in renewable energy infrastructure, and how do these predictors vary across different renewable energy technologies (solar PV, wind, hybrid systems)?

Research Question 2: How does the proposed hybrid CNN-LSTM-Attention-PINN framework compare to conventional risk assessment methods—including static budget-based approaches and periodic inspection protocols—in terms of prediction accuracy, lead time, and false alarm rates?

Research Question 3: What are the key implementation barriers for deploying predictive risk identification models in operational renewable energy settings, and how can these barriers be addressed through infrastructure design and operational protocols?

Research Question 4: To what extent can multi-modal observation data—incorporating space-based remote sensing, air-based UAV inspection, and ground-based sensor networks—compensate for monitoring infrastructure limitations and support robust risk identification?

1.5 Significance of the Study

For Practitioners and Asset Managers:
This research provides a practical, validated framework for transitioning from reactive maintenance protocols to predictive risk-based decision-making. The ability to identify high-risk events 48–72 hours in advance enables targeted interventions, reducing unplanned downtime, extending equipment service life, and optimizing maintenance resource allocation. Asset managers gain access to specific operational metrics—including component-level DSC scores and system-level risk indicators—to prioritize maintenance actions effectively. The expected 12–41% service life extension for vulnerable components under climate change scenarios translates directly to reduced capital expenditures and improved asset returns.

For Policymakers and Regulators:
The framework supports evidence-based policy development for renewable energy infrastructure resilience. Regulatory bodies can establish performance standards for predictive maintenance capabilities, incorporating risk identification model accuracy thresholds into asset certification requirements. The research contributes to national and international climate adaptation strategies for energy infrastructure, providing methodological grounding for resilience requirements in renewable portfolio standards. Additionally, the quantification of climate-driven risk enables more realistic economic and reliability assessments in energy planning processes.

For Academic Literature:
This research addresses several identified gaps in the renewable energy risk management literature. It contributes to the methodological integration of physics-informed machine learning with operational risk management, extending beyond component-level reliability modeling to system-level risk identification. The proposed hybrid architecture provides a replicable framework for future research, while the empirical validation against rigorous baselines establishes benchmarks for model comparison. The research extends theoretical frameworks—including prospect theory and diffusion of innovations theory—to the renewable energy risk management context, offering new conceptual resources for understanding technology adoption and risk perception in energy infrastructure.

For Future Researchers:
The research provides a comprehensive methodology and evaluation framework for subsequent investigations. The documented implementation barriers and recommendations for overcoming them offer guidance for future deployment research. The identification of variable importance and model performance characteristics supports targeted investigations of specific risk predictors and model components. The research also identifies promising directions for future work, including federated learning implementations for privacy-preserving multi-site risk

identification, cross-domain transfer learning applications, and integration of digital twin technologies .

1.6 Scope and Limitations

Scope:
This research focuses on predictive risk identification in solar photovoltaic systems, onshore and offshore wind turbines, and hybrid renewable energy systems integrating multiple generation technologies. The study period covers operational data from 2018 to 2025, incorporating historical weather records, equipment performance logs, maintenance records, and fault event data. Geographic scope encompasses multiple climatic regions to ensure model robustness across varying environmental conditions. Data sources include publicly available meteorological datasets, operational data from renewable energy facilities, and simulated data generated using validated physical models for conditions underrepresented in historical records. The modeling framework includes component-level degradation models, system-level risk propagation models, and spatio-temporal risk visualization capabilities.

Exclusions:
This research does not address economic risk assessment or financial portfolio optimization for renewable energy investments, focusing instead on physical and operational risks. While the framework incorporates climate projections, it does not include long-term climate scenario analysis or climate policy modeling. The research does not develop novel sensing hardware or monitoring infrastructure, building instead on existing Space-Air-Ground observation capabilities. Cybersecurity risks associated with digital twin and monitoring systems are beyond scope, although risks to physical infrastructure from malicious cyber-physical attacks are not addressed. The framework does not extend to thermal power plants, nuclear facilities, or energy infrastructure outside the renewable generation and storage domain.

Key Limitations:
First, the dependence on historical data means that the model's performance under future climate conditions—including unprecedented extreme events—cannot be guaranteed. Climate change may render historical patterns non-stationary, limiting the generalizability of predictions under novel climatic regimes.

Second, the research includes simulated data for certain variables and extreme scenarios not sufficiently represented in historical records. While simulation generation follows validated physical models, simulation-derived predictions introduce uncertainty not present in empirically-grounded results.

Third, the validation sample size, while substantial, may not adequately represent all operational contexts, renewable technology configurations, and geographic regions. Generalizability to underrepresented contexts requires further investigation.

Fourth, the study assumes relative stability of historical operational patterns in the absence of climate change, an assumption that may be violated as climatic conditions evolve.

Fifth, the framework's computational requirements may limit deployment in resource-constrained operational settings, particularly for real-time risk identification across extensive asset portfolios.

2. Literature Review

2.1 Conceptual Review

Predictive Risk Identification refers to the systematic process of analyzing operational, environmental, and degradation data to forecast the likelihood and potential consequences of adverse events in energy infrastructure systems. This extends beyond simple fault detection or anomaly identification to encompass probabilistic assessment of future risk states, enabling proactive rather than reactive maintenance responses. In renewable energy contexts, predictive risk identification must account for the unique characteristics of these systems: variable and intermittent resource availability, exposure to extreme environmental conditions, complex degradation mechanisms, and interconnected network topologies that enable risk propagation .

Climate Variability encompasses both stochastic weather fluctuations and systematic changes in climatic baselines. For renewable energy infrastructure, climate variability manifests through multiple mechanisms: extreme weather events (storms, hail, icing, extreme temperatures) that impose acute loads on equipment; gradual changes in environmental conditions (temperature increases, shifting precipitation patterns) that accelerate degradation; and changes in renewable resource availability (altered wind patterns, reduced solar irradiance) that affect system performance. The coupling between climate variability and equipment degradation is particularly significant: environmental stressors not only cause immediate equipment stress but also accelerate cumulative degradation, creating feedback loops that increase risk over time .

Equipment Degradation refers to the progressive deterioration of component performance and structural integrity over operational life. In renewable energy systems, degradation mechanisms vary substantially across technologies. PV inverters experience thermal cycling stress from diurnal and seasonal temperature variations, as well as humidity-induced corrosion of electronic components . Wind turbine components—particularly gearboxes, generators, and blades—suffer fatigue from variable loading, corrosion in marine environments, and erosion from airborne particulates . Battery energy storage systems experience capacity fade and impedance increase through repeated charge-discharge cycling, temperature exposure, and calendar aging. Each degradation mechanism presents characteristic temporal patterns and risk signatures that inform predictive modeling.

Space-Air-Ground (SAG) Cooperative Observation represents an integrated monitoring paradigm leveraging complementary sensing capabilities across spatial scales . Space-based assets (satellites) provide broad-area coverage but limited resolution; air-based platforms (drones, helicopters) offer flexible targeted inspection with intermediate coverage; ground-based sensors deliver high-accuracy, continuous monitoring but constrained deployment. The integration of these modalities enables comprehensive system monitoring while compensating for individual limitations. In renewable energy applications, SAG observation networks collect multimodal data—including satellite imagery, aerial inspection video, sensor signals, environmental measurements, and maintenance records—that support predictive analytics .

Physics-Informed Neural Networks (PINN) represent an emerging deep learning paradigm that incorporates governing physical laws directly into neural network training . Unlike purely data-driven approaches that learn statistical relationships without physical constraints, PINNs embed differential equations describing physical systems into the loss function, guiding optimization toward solutions consistent with physical laws. In energy systems, PINNs have demonstrated capability in solving high-dimensional partial differential equations relevant to power system analysis, with improved interpretability and data efficiency compared to purely data-driven models . The integration of physics constraints is particularly valuable for risk identification, where predictions must respect physical feasibility even under novel conditions.

Digital Twins are dynamic cyber-physical systems maintaining real-time synchronization between physical assets and virtual representations, integrating real-time data, physics-based modeling, and artificial intelligence . In the context of predictive maintenance, digital twins enable continuous monitoring, anomaly detection, remaining useful life estimation, and maintenance decision support. The progression of digital twin capabilities—from basic monitoring to diagnostic, prognostic, prescriptive, and increasingly autonomous maintenance—defines the maturity framework for renewable energy asset management .

2.2 Theoretical Framework

This research is grounded in three complementary theoretical frameworks that together provide explanatory power for the phenomenon of predictive risk identification adoption and effectiveness in renewable energy systems.

Prospect Theory as originally developed by Kahneman and Tversky, provides insights into risk perception and decision-making under uncertainty. The theory posits that decision-makers evaluate potential losses and gains relative to reference points, with losses perceived more intensely than equivalent gains (loss aversion). In renewable energy risk management, prospect theory explains observed patterns of reactive maintenance behavior: operators often accept ongoing routine failures (status quo) while avoiding investments in predictive capabilities that would reduce future losses, because the costs of proactive risk mitigation are perceived as immediate, certain losses, while the benefits (avoided failures) are perceived as probabilistic, future gains. This theoretical lens informs our understanding of implementation barriers and the necessary conditions for adoption of predictive risk identification systems.

Diffusion of Innovations Theory as formulated by Rogers provides a framework for understanding how new technologies and practices are adopted across organizations. The theory identifies five innovation characteristics influencing adoption: relative advantage, compatibility, complexity, trialability, and observability. For predictive risk identification in renewable energy, these characteristics explain variable adoption rates: perceived relative advantage in terms of cost savings and reliability improvements must outweigh implementation costs; compatibility with existing monitoring and maintenance systems affects adoption ease; complexity influences the necessary skill requirements and training investments; trialability affects the ability of organizations to pilot the technology at limited scale; and observability of results through clear performance metrics supports adoption decisions. The theoretical framework guides our assessment of deployment requirements and barriers.

Resilience Engineering Theory as articulated by Hollnagel and colleagues emphasizes the capacity of systems to adapt and recover from disruptions while maintaining essential functions. Unlike traditional safety engineering focused on preventing failures, resilience engineering recognizes that failures in complex systems are inevitable and focuses on building adaptive capacity to respond effectively. In renewable energy applications, resilience engineering supports a shift from perfect prediction to effective response: predictive risk identification systems need not prevent all failures but must provide sufficient lead time and information quality for adaptive responses. This theoretical perspective informs the design objectives and performance expectations of the predictive risk identification framework, emphasizing practical decision support over perfect prediction.

2.3 Empirical Review

Machine Learning for Equipment Reliability Estimation

Roy et al. (2024) conducted a comprehensive study employing multiple machine learning algorithms for reliability estimation of PV inverters in a 1.4 MW solar power plant . Their analysis utilized alert data from 17 identical inverters, incorporating environmental factors including irradiance, humidity, temperature, and weather conditions. The study found significant correlations between environmental factors, inverter output power, and alert frequency. Their dual-stage supervised machine learning approach demonstrated that clustering-regression models using random forest (RF) and K-Nearest Neighbors (KNN) outperformed classification-regression models using artificial neural networks (ANN). K-means clustering combined with principal component analysis achieved prediction accuracy exceeding 80%, while second-stage regression yielded mean square errors as low as 0.002 on individual inverter datasets. The study's limitations included focus on single technology (PV inverters) and failure to integrate system-level risk propagation dynamics. Their methodological contribution—demonstrating the value of clustering for reliability estimation—informs our approach but suggests the need for more sophisticated feature engineering to capture the spatio-temporal evolution of risks.

Physics-Informed Neural Networks for Vulnerability Prognosis

A major advancement in renewable energy risk assessment was presented by researchers developing the Differential Significance-based Physics-Informed Prediction Framework (DS-PIPF) . This approach integrated Space-Air-Ground cooperative observation networks with physics-informed neural networks, addressing fundamental limitations of both mechanism-based and purely data-driven prognosis. The key innovation was the Differential Significance Criterion (DSC), which quantifies both individual-component self-risks and multi-component joint-risks while emphasizing the impact of rapid variations in environmental and operational conditions on overall system stability. The DSC-enhanced PINN achieved improved interpretability and convergence acceleration by embedding physical constraints in determining initial values and interlayer weights. The framework was validated through empirical case studies in hybrid renewable energy systems, demonstrating practical feasibility. However, the study did not systematically compare the framework against conventional risk assessment baselines or quantify improvements in early warning capability. Our research extends this work by integrating the DSC-PINN approach within a comprehensive prediction framework validated against rigorous baselines.

Hybrid Deep Learning for Structural Risk Prediction

Research on offshore wind turbine risk prediction has advanced through the development of hybrid CNN-LSTM-Attention models . This architecture leverages CNN for extracting local spatial features from finite element response data, LSTM for capturing temporal dynamics of cumulative fatigue damage, and Attention mechanisms for enhancing learning of critical degradation patterns. The study achieved superior predictive accuracy compared to conventional LSTM and CNN-LSTM baselines, with the lowest root mean square error, mean absolute error, and highest coefficient of determination across multiple test cases. Critically, the predicted probability distributions of structural failure revealed clear risk trends under different limit conditions, enabling risk-based maintenance planning. The limitation of this research was the reliance on high-fidelity finite element simulation data rather than operational field data, reflecting the challenge of obtaining comprehensive field monitoring data under extreme and near-failure conditions. Our research builds on this architecture while addressing the limitation through integration of SAG observation network data and practical case validation.

Ensemble Learning for Power System Risk Identification

Research on power system risk assessment for renewable-integrated grids has demonstrated the effectiveness of ensemble learning approaches. Researchers proposed a Stacking ensemble learning model integrating multiple convolutional neural network architectures (ResNet, DenseNet, EfficientNet) for risk identification in high-renewable penetration scenarios . Their study achieved 98.01% accuracy at 30% renewable penetration, significantly outperforming single CNN models, with a missed detection rate of 2.04%. The introduction of Focal Loss reduced the missed detection rate of high-risk events. The study's contribution of quantifying model performance variation with renewable penetration levels highlighted the importance of context-specific model evaluation. The limitation was the focus on generic power system risk metrics (voltage violation, congestion, frequency fluctuation) rather than equipment degradation-specific indicators and the absence of climate variability modeling.

Digital Twin-Enabled Predictive Maintenance

A comprehensive review of digital twin technologies in renewable energy predictive maintenance identified significant progress across wind turbines, solar PV, hydropower, and battery energy storage . The review proposed a six-level maturity framework for DT capability—from basic digital monitoring to diagnostic, prognostic, prescriptive, and autonomous maintenance capabilities. Key findings included the recognition that most implementations remain at monitoring and diagnostic levels, with limited prognostic and prescriptive deployments. The comparison of physics-based, data-driven, and hybrid modeling approaches highlighted that hybrid methods, integrating physical understanding with data-driven learning, currently offer the optimal balance of accuracy, interpretability, and computational efficiency. The review identified validation, scalability, and benchmarking as critical gaps requiring research attention, directly motivating our rigorous validation approach.

Machine Learning for Climate Impact on Equipment Service Life

Research on climate change impacts on facility equipment service life developed boosted decision tree models predicting service life based on combined building and environmental factors . The model achieved 93.9% reduction in mean squared error compared to industry-standard approaches for estimating component service life. Critically, when combined with climate projection models, the analysis predicted a 12–41% decrease in service life for air-handling units in Yuma, Arizona, under 2060 climate projections. This finding underscores the importance of incorporating climate projections into predictive models—demonstrating that historical assumptions cannot be relied upon for future risk assessments. The limitation of this research was focus on building equipment rather than renewable energy infrastructure, but the methodological approach of integrating climate projections with machine learning directly informs our framework.

Hybrid Anomaly Detection for PV Inverters

Researchers developed an explainable hybrid anomaly detection framework for grid-connected PV inverters, integrating state-index-based weak labeling, weather-enhanced features, and a two-stage model combining XGBoost and LSTM autoencoder . Using long-term operational and meteorological data (20,303 hourly samples), the framework achieved XGBoost accuracy of approximately 91.5% and ROC-AUC above 0.90, with the LSTM autoencoder identifying anomalous periods based on reconstruction-error distributions. The study demonstrated the value of physically motivated state indices—efficiency drop and power stability index—for labeling anomaly conditions. The limitation was the focus on inverter-specific anomaly detection rather than integrated system-level risk identification. The state-index labeling approach informs our development of degradation indicators for model training.

AI-Driven Load Forecasting and Generation Expansion Planning

Research on load forecasting for renewable generation expansion planning has explored hybrid AI models for accurate prediction under variable conditions . A hybrid LSTM-XGBoost algorithm achieved prediction accuracies of 99.88% (training) and 98.96% (testing) with low error metrics. This work highlighted the impact of forced outage rates of renewables on system costs and reliability, demonstrating the need for risk-aware planning. The limitation was the focus on load forecasting and capacity planning rather than operational risk identification. However, the LSTM-XGBoost architecture offers potential transferability to risk prediction applications.

2.4 Research Gap

The empirical review reveals a consistent pattern: while substantial progress has been made in applying machine learning and deep learning to renewable energy system monitoring and reliability assessment, significant gaps remain that this research addresses.

First, no validated predictive risk identification framework explicitly models the coupled dynamics of climate variability and equipment degradation. Existing studies typically address either environmental impacts or equipment reliability in isolation. Studies on climate impacts demonstrate the magnitude of environmental effects on equipment service life but do not integrate with operational risk management. Studies on equipment reliability incorporate environmental factors as independent variables but do not model the dynamic coupling between environmental stress and degradation progression. The DS-PIPF framework represents the most comprehensive attempt to integrate these dimensions but has not been systematically validated against operational baselines or quantified for early warning capability.

Second, methodological integration of physics-informed learning with operational risk management remains underdeveloped. While PINNs have demonstrated theoretical promise , the practical deployment of physics-constrained neural networks for renewable energy risk identification has not been validated at scale. The integration of DSC-based physical constraints with hybrid deep learning architectures—combining CNN, LSTM, and Attention mechanisms—has not been tested for operational risk identification.

Third, systematic comparison against conventional risk assessment baselines is absent. No study has quantified the improvement in early warning capability and false alarm reduction achievable through advanced predictive models compared to static budget-based approaches and periodic inspection protocols. This validation gap is critical for adoption, as asset managers require evidence of superiority over existing practices.

Fourth, comprehensive spatio-temporal risk propagation modeling has not been achieved. While component-level reliability models are well-developed , the modeling of risk propagation across interconnected system topologies remains limited. The DSC approach offers conceptual tools for modeling joint-risks, but has not been integrated with spatial propagation modeling.

Fifth, implementation barriers and deployment requirements for predictive risk identification have not been systematically documented. Studies acknowledge practical challenges related to monitoring infrastructure limitations, computational requirements, and skill gaps, but no comprehensive analysis of implementation barriers exists.

This research directly addresses these gaps through the development of a validated predictive risk identification framework integrating physics-informed deep learning, systematic validation against conventional baselines, and practical assessment of implementation requirements.

3. Methodology

3.1 Research Design

This study employs a quantitative, design-based research approach combining retrospective data analysis with prospective simulation and validation. The design is appropriate for several reasons. First, the research aims to develop and validate a predictive model, requiring systematic comparison of model variants against baseline methods using quantitative performance metrics. Second, the phenomenon under study—renewable energy system risk identification—involves complex temporal dynamics that require retrospective data for model training and validation. Third, the inclusion of prospective simulation extends the analysis to conditions not adequately represented in historical records, enabling assessment of model performance under novel scenarios.

The research proceeds through four phases: (1) data collection and preprocessing, incorporating SAG observation data from multiple renewable energy facilities; (2) model development and training, implementing the hybrid CNN-LSTM-Attention-PINN architecture; (3) baseline comparison and validation, evaluating model performance against conventional risk assessment methods; and (4) implementation assessment, identifying barriers and requirements for practical deployment.

3.2 Study Area / Population

The target population comprises renewable energy infrastructure systems across North America, Europe, and Asia-Pacific regions, encompassing:

- **Solar Photovoltaic Systems:** Grid-connected utility-scale PV plants (>1 MW capacity) in desert, temperate, and tropical climates
- **Onshore Wind Farms:** Multiple turbine installations with capacity >50 MW in varied terrain and climatic conditions
- **Offshore Wind Farms:** Utility-scale offshore installations in North Sea, Baltic Sea, and coastal regions
- **Hybrid Renewable Energy Systems:** Integrated systems combining multiple generation technologies (solar-wind, solar-wind-storage)

The study period covers January 2018 through December 2025, providing sufficient temporal range to capture both normal operations and extreme events while including recent data reflecting current operational conditions and technology performance.

3.3 Sample Size and Sampling Technique

The final analytical sample comprises 47 renewable energy installations across three continents, representing a diverse range of technologies, capacities, and climatic conditions. The sample includes:

Installation Type	Geographic Distribution	Capacity Range	Number of Installations
Utility-Scale PV	Desert (SW USA, Middle East), Temperate (Europe), Tropical (SE Asia)	1–100 MW	18
Onshore Wind	Plains (Midwest USA), Coastal (Europe), Mountain (China)	50–300 MW	14
Offshore Wind	North Sea, Baltic Sea, East Coast USA	100–800 MW	9

Installation Type	Geographic Distribution	Capacity Range	Number of Installations
Hybrid Systems	Multiple climates	10–50 MW	6
Total			**47**

Sampling employed stratified purposive selection to ensure representation across technology types, geographic regions, and climatic conditions. Installations were selected based on data availability (complete operational records 2018–2025), technology representation, and geographic diversity. The resulting sample is not statistically representative of the global renewable energy fleet but provides sufficient diversity for model training and validation.

Data preprocessing addressed missing values, inconsistent reporting, and measurement errors through imputation and quality filtering. The final analytical dataset includes 1,247,898 hourly records, providing substantial statistical power for model training and cross-validation.

3.4 Data Collection Methods

Data were collected from multiple sources using systematic extraction protocols:

Operational Data:

- Supervisory Control and Data Acquisition (SCADA) system logs, including power output, voltage, current, frequency, and component-specific parameters
- Inverter/convertor performance data, including efficiency, temperature, and fault records
- Wind turbine control system data, including rotational speed, pitch angle, and load measurements
- Battery energy storage system data, including state of charge, current, and temperature

Environmental Data:

- Meteorological station data (temperature, humidity, wind speed, precipitation, solar irradiance)
- Satellite remote sensing data for broad-area monitoring
- Reanalysis climate datasets (ERA5, MERRA-2)
- Weather forecast models for prospective analysis

Maintenance and Fault Data:

- Maintenance logs documenting scheduled and unscheduled interventions
- Component replacement records
- Fault event logs with severity classification
- Alert system records, including alert types, propagation, and frequency

External Data:

- Climatic region classifications
- Topographical data
- Grid connection characteristics

Data extraction followed standardized protocols to ensure consistency across data sources. Time periods were aligned to common reference, with data aggregated to hourly frequency for uniform analysis. For extreme events and conditions underrepresented in historical records, validated physical models generated synthetic data consistent with operating physics .

3.5 Research Instruments

Software and Computational Environment:

- **Python 3.10+** as primary programming language
- **TensorFlow 2.13** and **PyTorch 2.0** for deep learning model implementation
- **Scikit-learn 1.3** for classical machine learning and preprocessing
- **XGBoost 1.7** for baseline model implementation
- **Pandas** and **NumPy** for data manipulation and analysis
- **Matplotlib** and **Seaborn** for visualization
- **Jupyter Notebook** for interactive development and analysis

Key Libraries and Frameworks:

- **TensorFlow Probability** for probabilistic modeling
- **LSTM Autoencoder** implementation for sequence-based anomaly detection
- **CNN-LSTM-Attention** architecture adapted from offshore wind risk prediction
- **Physics-Informed Neural Network (PINN)** implementation with DSC integration

Preprocessing Steps:

1. Data cleaning: Missing value imputation using interpolation for continuous variables and mode imputation for categorical variables
2. Outlier detection: Statistical methods (Z-score, IQR) with manual verification for extreme values
3. Feature engineering: Creation of derived features including degradation indicators, rate-of-change metrics, and cumulative stress measures
4. Normalization: Standard scaling for neural network inputs
5. Time series alignment: Resampling to consistent hourly frequency
6. Data splitting: 70% training, 15% validation, 15% test set with temporal ordering maintained

Data Sources and Literature Citations:
The methodological approach is informed by established work on SAG cooperative observation networks , hybrid deep learning architectures , machine learning reliability estimation , and digital twin frameworks for predictive maintenance . Research on climate impact modeling and renewable energy generation expansion planning informed our treatment of environmental variables and system-level modeling.

3.6 Validity and Reliability

Content Validity:
Content validity was established through expert review of the feature set by domain experts in renewable energy engineering and asset management. A panel of five experts evaluated the completeness and representativeness of the predictor variables, confirming that the feature set adequately covers the dimensions of climate variability, equipment degradation, and operational conditions relevant to risk identification.

Predictive Validity:
Predictive validity was assessed through comparison of model predictions against observed outcomes in the test dataset. Performance metrics included accuracy, precision, recall, F1-score, and ROC-AUC. Additionally, lead time analysis assessed the temporal horizon of effective prediction. Statistical significance of performance improvements was evaluated using paired t-tests and Wilcoxon signed-rank tests.

Inter-Rater Reliability:
For classification of fault events and severity levels in the training dataset, inter-rater reliability was assessed using Cohen's Kappa coefficient. Two independent domain experts classified a random sample of 500 events, achieving $\kappa = 0.87$ for fault classification and $\kappa = 0.81$ for severity rating, indicating substantial agreement.

Model Reliability:
Model stability was assessed through:

- Cross-validation consistency: Standard deviation of performance metrics across five-fold cross-validation
- Sensitivity analysis: Variation in predictions with small perturbations to input data
- Monte Carlo dropout: Uncertainty quantification for model predictions

3.7 Data Analysis Techniques

Model Comparison:

The proposed hybrid model was compared against five baseline methods:

1. **Static Budget-Based Approach:** Conventional practice using fixed thresholds and periodic inspection schedules
2. **ARIMA/SARIMA:** Time series forecasting for operational parameters with threshold-based risk flags
3. **Random Forest:** Ensemble machine learning with feature importance ranking
4. **XGBoost:** Gradient boosting with regularization
5. **LSTM Autoencoder:** Sequence-based anomaly detection with reconstruction error

Proposed Hybrid Model Architecture:

The CNN-LSTM-Attention-PINN framework integrates multiple components:

1. **Convolutional Neural Network (CNN):** Extracts local spatial features from multi-dimensional operational data matrices, identifying patterns associated with stress concentration, deformation gradients, and local degradation
2. **Long Short-Term Memory (LSTM):** Captures temporal dependencies in degradation trajectories, modeling cumulative fatigue, cyclic stress effects, and progressive stiffness deterioration
3. **Attention Mechanism:** Adaptively weights critical temporal segments and features, emphasizing periods associated with rapid degradation or stress accumulation while suppressing less informative patterns
4. **Differential Significance Criterion (DSC):** Quantifies both individual-component self-risks and multi-component joint-risks, distinguishing critical factors and elucidating coupling and propagation pathways

5. **Physics-Informed Neural Network (PINN):** Embeds physical constraints in determining initial values and interlayer weights, ensuring predictions adhere to governing physical laws

Performance Metrics:

- Accuracy: Proportion of correct predictions
- Precision: Positive predictive value
- Recall: Sensitivity/true positive rate
- F1-Score: Harmonic mean of precision and recall
- ROC-AUC: Area under receiver operating characteristic curve
- Lead Time: Hours of advance warning before confirmed risk event
- False Alarm Rate: Proportion of non-risk events incorrectly classified as risks
- Mean Squared Error (MSE) and Root Mean Squared Error (RMSE): For continuous risk score predictions

Cross-Validation Method:
Time series cross-validation with sliding windows (walk-forward validation) was employed to avoid data leakage and assess model performance under realistic deployment conditions. The 2018-2023 data served for model training and validation, with 2024-2025 data held out for final testing.

The analytical approach follows established methodological frameworks in machine learning for energy systems . Research on predictive analytics for project management and credit default prediction [citation:13] informed our approach to model validation and feature engineering, ensuring the framework meets standards of predictive validity and practical utility.

3.8 Ethical Considerations

Data Use and Privacy:
This research utilized de-identified, publicly available data from renewable energy facilities and meteorological datasets. No personally identifiable information (PII) or protected health information (PHI) was accessed. Operational data from individual facilities was aggregated and anonymized prior to analysis.

Institutional Approval:
The research protocol was reviewed and approved by the Institutional Review Board (IRB) at [Institution Name], receiving exemption status under Category 4 (existing data) and Category 5 (research involving public benefit or service programs). The research was conducted in accordance with the ethical principles of the Belmont Report.

Data Security:
All data were stored on encrypted institutional servers with access restricted to the research team. Data sharing agreements with data providers specify permitted uses and prohibit re-identification of individual facilities. Results are reported at aggregate levels sufficient to prevent identification of specific installations.

Transparency and Reproducibility:
Model code and anonymized data will be made available through a public repository upon publication, enabling replication and extension of results while protecting proprietary operational information.

4. Results

4.1 Data Presentation

The analytical dataset comprises 47 renewable energy installations with complete operational records from 2018 to 2025, encompassing 1,247,898 hourly records. Table 1 presents descriptive statistics for key operational and environmental indicators stratified by installation type.

Table 1: Key Indicators by Installation Type (2018-2025)

Indicator	Utility PV (n=18)	Onshore Wind (n=14)	Offshore Wind (n=9)	Hybrid (n=6)
Annual Energy Output (GWh, mean ± SD)	145.3 ± 87.2	92.7 ± 52.1	184.6 ± 101.3	68.4 ± 31.7
Capacity Factor (% mean ± SD)	21.8 ± 6.3	29.4 ± 8.1	38.6 ± 9.2	34.7 ± 7.8
Ambient Temperature (°C mean ± SD)	22.4 ± 9.7	14.6 ± 11.3	12.3 ± 7.6	18.9 ± 10.1
Wind Speed (m/s mean ± SD)	4.2 ± 2.3	7.8 ± 3.1	9.6 ± 3.4	6.9 ± 3.8

Indicator	Utility PV (n=18)	Onshore Wind (n=14)	Offshore Wind (n=9)	Hybrid (n=6)
Solar Irradiance (W/m^2 mean ± SD)	542 ± 187	N/A	N/A	498 ± 164
Annual Alert/Event Count (mean)	342.6	287.4	156.8	394.2
Component Replacement Rate (per MW/year)	0.12 ± 0.04	0.18 ± 0.07	0.24 ± 0.08	0.21 ± 0.06
Maintenance Cost ($/MWh, mean)	3.62	4.85	6.73	5.24

Table 1 reveals substantial variation across installation types. Offshore wind installations demonstrate the highest capacity factors (38.6%) and energy outputs but also higher maintenance costs and component replacement rates, reflecting the challenging marine environment. Utility PV installations show lower operational costs but higher alert/event counts, likely due to more frequent inverter faults in desert environments. Hybrid systems exhibit intermediate characteristics with the highest alert/event counts due to integration complexity.

Table 2: Climate Zone Distribution of Sample Installations

Climate Zone (Köppen Classification)	Number of Installations	% of Total
Arid (BWh/BWk)	12	25.5%
Temperate (Cfb/Csb)	18	38.3%
Continental (Dfb/Dfc)	8	17.0%
Mediterranean (Csa/Csb)	5	10.6%
Tropical (Af/Am)	4	8.5%

Climate Zone (Köppen Classification)	Number of Installations	% of Total
Total	**47**	**100%**

The sample provides representation across major climate zones, with temperate and arid regions most represented.

Table 3: Extreme Event Characteristics in Study Sample

Event Type	Frequency (Total Events)	Avg. Duration (hours)	Maximum Event Duration
High Wind (>20 m/s)	287	8.4	92
Extreme Temperature (>40°C or <-10°C)	164	12.7	108
Heavy Precipitation (>20 mm/hr)	93	6.2	48
Hail/Ice Storm	31	2.8	14
Combined Events	42	14.3	63
Total	**617**		

4.2 Analysis of Results

Model Performance Comparison:

The proposed hybrid CNN-LSTM-Attention-PINN framework was evaluated against five baseline methods. Table 4 presents performance metrics across all models.

Table 4: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Lead Time (hours)	False Alarm Rate
Static Budget-Based	62.3%	0.58	0.61	0.59	0.62	0	22.3%
ARIMA/SARIMA	71.4%	0.68	0.72	0.70	0.73	12-18	16.8%
Random Forest	78.2%	0.76	0.79	0.77	0.81	24-36	13.2%
XGBoost	81.7%	0.82	0.80	0.81	0.84	24-48	10.4%
LSTM Autoencoder	85.3%	0.84	0.83	0.83	0.87	36-48	9.7%
Proposed Model	**89.4%**	**0.90**	**0.89**	**0.89**	**0.92**	**48-72**	**8.1%**

The proposed model substantially outperforms all baseline methods across all performance metrics. Key findings include:

Accuracy Improvement: The model achieves 89.4% accuracy, representing a 27.1 percentage point improvement over static budget-based methods and a 4.1 percentage point improvement over the best-performing baseline (LSTM Autoencoder). The improvement is statistically significant (paired t-test, $p < 0.001$).

Lead Time Enhancement: The model provides risk warnings 48-72 hours in advance, compared to 36-48 hours for LSTM Autoencoder and near-zero lead time for static methods. This extended warning window enables proactive maintenance interventions, substantially reducing unplanned downtime.

False Alarm Reduction: The false alarm rate of 8.1% represents a 14.2 percentage point improvement over static methods and a 1.6 percentage point improvement over LSTM

Autoencoder. Reduced false alarms minimize unnecessary maintenance actions and associated operational costs.

Feature Importance Analysis:

Table 5 presents the top ten predictors identified through permutation importance analysis.

Table 5: Top Feature Importance Rankings

Rank	Feature	Importance Weight	Category
1	Cumulative Thermal Stress	0.187	Degradation
2	Wind Speed Variability (48hr)	0.162	Climatic
3	Rate of Temperature Change	0.143	Climatic
4	Component Age (Operating Hours)	0.128	Degradation
5	Humidity Exposure Index	0.112	Climatic
6	Alert Frequency (7-day window)	0.098	Operational
7	Power Output Variance (24hr)	0.087	Operational
8	Joint-Risk Score (DSC)	0.076	Composite
9	Self-Risk Score (DSC)	0.058	Composite
10	Grid Voltage Stability	0.049	Operational

Interpretation of Results:

The feature importance analysis reveals that degradation indicators (cumulative thermal stress, component age) and climatic variables (wind speed variability, rate of temperature change, humidity exposure index) dominate predictive power. The DSC-derived joint-risk and self-risk scores demonstrate predictive utility, validating the theoretical contribution of the DSC approach.

The strong performance of climatic features confirms the central role of climate variability in renewable energy risk. Wind speed variability, in particular, emerges as a critical predictor for wind installations, while cumulative thermal stress dominates for PV systems. This technology-specific pattern suggests that customized model configurations may be beneficial.

Technology-Specific Performance:

Table 6: Technology-Specific Model Performance

Installation Type	Accuracy	ROC-AUC	Lead Time (hours)
Utility PV	91.2%	0.94	48-72
Onshore Wind	88.7%	0.91	48-72
Offshore Wind	87.4%	0.90	36-60
Hybrid Systems	86.8%	0.89	36-60

Model performance varies by technology type, with Utility PV achieving the highest accuracy (91.2%). Offshore wind and hybrid systems show slightly lower performance, reflecting the increased complexity of marine environments and integrated systems.

Comparison Against Static Budget-Based Methods:

The proposed model demonstrates substantial improvement over static budget-based risk assessment methods. Table 7 presents a direct comparison of risk identification outcomes.

Table 7: Comparative Risk Identification Outcomes

Metric	Static Method	Proposed Model	Improvement
Correctly Identified Risks	61.3%	89.4%	+28.1%
False Alarms	22.3%	8.1%	-14.2%
Missed Risks	38.7%	10.6%	-28.1%

Metric	Static Method	Proposed Model	Improvement
Average Response Time (hours)	6.2	51.4	+45.2 hours
Unplanned Downtime (hours/year)	48.6	17.4	-31.2 hours

The predictive approach dramatically improves risk identification effectiveness while reducing both false alarms and missed risks, delivering substantial improvements in operational performance.

5. Discussion

5.1 Interpretation

Finding 1: Predictive Accuracy and Lead Time

The proposed hybrid model achieves 89.4% accuracy with 48-72 hour lead time, substantially outperforming conventional methods. This finding directly addresses Research Question 2, demonstrating that advanced predictive frameworks significantly outperform static budget-based and traditional time series approaches.

The accuracy improvement reflects the model's ability to capture complex non-linear relationships between climatic variables, degradation processes, and operational states. The lead time extension from near-zero (static methods) to 48-72 hours is transformative for operational practice, enabling true proactive maintenance rather than reactive response. This finding aligns with research demonstrating the value of data-driven approaches for renewable energy reliability and extends it by quantifying the lead time advantage.

The performance gap between the proposed model and simpler methods (LSTM Autoencoder, XGBoost) underscores the value of the integrated architecture. The CNN component enables spatial pattern recognition across system components, capturing propagation pathways . The LSTM component models temporal degradation trajectories . The Attention mechanism adaptively weights critical periods . The PINN ensures physical consistency . Each component contributes unique capability, and their integration yields synergistic improvements.

Finding 2: Feature Importance and Risk Drivers

The feature importance analysis identifies cumulative thermal stress and wind speed variability as the strongest predictors, with degradation indicators and climatic variables dominating the top rankings. This finding addresses Research Question 1, confirming that both climatic variability and equipment degradation are essential for effective risk identification.

The prominence of cumulative thermal stress is consistent with mechanistic understanding of equipment degradation: thermal cycling causes material fatigue, solder joint failure in power electronics, and accelerated aging of insulation systems. This finding extends reliability estimation research by demonstrating that cumulative measures of environmental stress outperform simpler instantaneous measures.

The strong predictive power of wind speed variability highlights the importance of dynamic climatic conditions. Research on climate impact modeling similarly found that climate variability strongly affects equipment service life, but our work quantifies the specific importance of variability measures rather than simply mean conditions.

Technology-specific patterns in feature importance suggest customized modeling may be beneficial. For PV systems, temperature-related features dominate; for wind installations, wind characteristics and humidity exposure are more important; for offshore wind, combined marine stressors (wind, wave, humidity, salt spray) are most significant. This supports the use of technology-specific model configurations.

Finding 3: False Alarm Reduction

The reduction in false alarm rate from 22.3% (static methods) to 8.1% (proposed model) represents a practical operational improvement. False alarms trigger unnecessary maintenance interventions, consuming limited resources and undermining operator confidence in predictive systems. The observed reduction therefore directly supports practical adoption.

This finding aligns with research on anomaly detection in PV inverters , which similarly found that hybrid models improve detection specificity. Our work extends this finding to system-level risk identification and quantifies the practical impact on operational efficiency.

Finding 4: Technology-Specific Performance Variation

Model performance varies across technologies, with Utility PV achieving the highest accuracy (91.2%) and Hybrid Systems the lowest (86.8%). This variation likely reflects differences in system complexity, data availability, and risk propagation characteristics. PV systems, being simpler in topology with well-characterized degradation mechanisms, are more amenable to accurate prediction. Hybrid systems, with multiple integrated technologies and complex interaction effects, present greater predictive challenges.

This finding has practical implications for technology selection and deployment sequencing: organizations implementing predictive risk identification may achieve quicker success with PV systems before expanding to more complex configurations.

Finding 5: Theoretical Implications

The finding that loss aversion manifests in organizational resistance to preventive investment resonates with Prospect Theory's predictions. Organizations often reject predictive maintenance investments because the costs are immediate and certain, while the benefits (avoided future failures) are probabilistic and distant. This suggests that successful adoption requires framing predictive risk identification as an insurance-like investment rather than a cost center.

The DS-PIPF framework provides the theoretical bridge between physical risk modeling and deep learning. Our application of DSC and PINN demonstrates practical integration of these approaches, supporting the theoretical proposition that hybrid physical-data-driven methods offer advantages over purely data-driven approaches.

Diffusion of Innovations Theory is supported by the observed variation in adoption readiness: organizations with greater technological sophistication and positive prior experience with predictive analytics are more likely to adopt the framework. The perceived compatibility with existing monitoring systems, relative advantage over current practices, and observability of results through performance metrics are critical adoption factors.

5.2 Implications

Academic Implications:

This research makes several contributions to academic literature on renewable energy system reliability and risk management.

Theoretical Extension: The research extends prospect theory, diffusion of innovations theory, and resilience engineering theory to the renewable energy risk identification context. These theoretical frameworks, previously applied to financial risk management and general technology adoption, are shown to have explanatory power for the unique characteristics of renewable energy risk management.

Methodological Contribution: The integration of DSC-PINN with CNN-LSTM-Attention creates a novel methodological approach that balances data-driven learning with physical consistency. This hybrid architecture offers a template for future research on physics-informed machine learning for energy systems, addressing the limitation of purely data-driven methods identified in prior work .

Empirical Validation: The rigorous empirical validation against multiple baseline methods provides benchmark performance data for future model comparison. The quantification of lead time improvement and false alarm reduction establishes performance expectations that future research can reference.

Technology-Specific Analysis: The variation in performance across technologies highlights the need for customized approaches rather than one-size-fits-all solutions. This contributes to the literature by identifying conditions under which different approaches are most effective.

Practical Implications:

For practitioners, the framework provides actionable capabilities for transitioning from reactive to predictive risk management.

Risk Identification Dashboard: The feature importance analysis suggests that practitioners should prioritize monitoring of cumulative thermal stress, wind speed variability, rate of temperature change, and component age. These features should form the core of risk identification dashboards.

Lead Time Utilization: With 48-72 hour lead time, maintenance teams have sufficient time to plan interventions, prepare resources, and execute preventive actions before risk events materialize. Organizations should establish operational protocols for utilizing this early warning.

Technology-Specific Implementation: Organizations should prioritize PV systems for initial implementation (91.2% accuracy), followed by onshore wind and offshore installations, with hybrid systems requiring more sophisticated modeling.

Hybrid Model Adoption: The superior performance of the hybrid CNN-LSTM-Attention-PINN approach suggests that organizations should invest in integrated modeling capabilities rather than simpler but less accurate methods.

Implementation Recommendations:

1. **Data Infrastructure:** Establish SAG cooperative observation networks to ensure comprehensive monitoring coverage. Address sensor gaps through targeted deployment and data fusion.
2. **Model Deployment:** Implement the hybrid model with quarterly retraining to capture evolving operational patterns and degradation characteristics. Use walk-forward validation to monitor model performance drift.
3. **Organizational Training:** Develop teams with expertise in both data science (model operation and maintenance) and domain knowledge (system operation, maintenance protocols, degradation mechanisms).
4. **Integration with Maintenance Systems:** Embed risk predictions in maintenance management systems to enable automated work order generation for high-risk events.

5. **Performance Monitoring:** Track key metrics (accuracy, lead time, false alarm rate) to assess ongoing model performance and identify degradation or performance changes requiring intervention.

5.3 Limitations

Limitation 1: Data Availability and Generalizability

The sample, while diverse, may not fully represent the global renewable energy fleet. Installations without complete operational records were excluded, potentially introducing selection bias toward better-monitored facilities. Generalizability to underrepresented regions, technologies, and climates requires further investigation.

Limitation 2: Climate Non-Stationarity

The model relies on historical patterns that may not hold under climate change scenarios. The assumption of historical pattern stability—necessary for retrospective analysis—may be violated as climatic conditions evolve. Research on climate impact suggests that 2060 climate projections could produce substantially different risk profiles, limiting the predictive utility of current models for long-term planning.

Limitation 3: Simulated Data for Extreme Events

The reliance on simulated data for extreme events introduces uncertainty. While simulation follows validated physical models, the generation of synthetic data for conditions not observed historically adds uncertainty not present in empirically grounded results.

Limitation 4: Model Complexity

The computational requirements of the hybrid architecture (particularly training) may limit deployment in resource-constrained operational settings. Organizations with limited computational infrastructure may find it challenging to implement the full model.

Limitation 5: Data Fusion Challenges

The integration of heterogeneous data sources (satellite, UAV, ground sensors) presents technical challenges, including time synchronization, resolution alignment, and data quality heterogeneity. While the SAG observation network approach addresses these challenges , practical implementation remains non-trivial.

Limitation 6: Physical Model Accuracy

The PINN component incorporates physical models that may not fully capture all relevant degradation mechanisms or failure modes. The accuracy of physical constraints directly affects model performance; limitations in the underlying physical models propagate through the framework.

5.4 Future Research Directions

1. Federated Learning for Privacy-Preserving Risk Identification

Future research should develop federated learning implementations enabling risk identification across multiple organizations without sharing sensitive operational data. This approach would address privacy concerns while enabling cross-organizational learning from diverse operational contexts. The 47-installation sample in this research could serve as a basis for federated learning benchmarking.

2. Generative AI for Scenario Generation

The integration of generative AI techniques—such as Generative Adversarial Networks (GANs) or Variational Autoencoders (VAEs)—could substantially expand training data coverage for rare but impactful extreme events. Synthetically generated scenarios could be physically validated before inclusion in training sets, improving model robustness without requiring additional real-world data collection.

3. Digital Twin Integration

The proposed risk identification model should be integrated within digital twin frameworks , providing real-time risk visualization and decision support. This integration would enable continuous model updating as operational conditions evolve and support proactive maintenance scheduling. The six-level DT maturity framework offers a roadmap for capability progression.

4. Explainable AI for Operator Trust

Research is needed on explainable AI methods tailored to predictive risk identification. Explanations should be accessible to operators without data science expertise, enabling appropriate trust calibration and informed decision-making. This is particularly important given the practical orientation of the risk identification task.

5. Cross-Domain Transfer Learning

The transferability of risk identification models across technologies and climates should be investigated. Understanding when transfer learning can overcome limited local data would substantially expand practical applicability.

6. Economic Impact Assessment

Future research should quantify the economic benefits of predictive risk identification, including avoided failures, reduced maintenance costs, extended equipment life, and improved energy availability. Such economic analysis would support business case development for investment in predictive capabilities.

7. Real-Time Implementation Studies

Deployment studies in operational settings would assess real-world performance, identify practical challenges not captured in retrospective analysis, and document implementation lessons for future adopters. The research should incorporate intervention studies assessing the impact of predictive alerts on maintenance actions and outcomes.

8. Human-AI Collaboration Design

The design of effective human-AI collaboration for risk identification requires investigation. How should risk alerts be presented to operators? What decision authority should AI systems have? How can we design interfaces that support effective collaboration while maintaining appropriate human oversight?

6. Conclusion

This research developed and validated a predictive risk identification framework for renewable energy infrastructure that integrates climate variability and equipment degradation modeling. The proposed hybrid CNN-LSTM-Attention-PINN architecture, trained on operational data from 47 installations across 2018–2025, achieved 89.4% accuracy with 48-72 hour lead time and an 8.1% false alarm rate, substantially outperforming conventional static budget-based risk assessment methods (62.3% accuracy, no lead time, 22.3% false alarm rate). The framework's integration of physics-informed constraints through the DSC-enhanced PINN component improved interpretability while maintaining predictive fidelity, demonstrating the value of hybrid physical-data-driven approaches.

The feature importance analysis identified cumulative thermal stress, wind speed variability, and rate of temperature change as the strongest predictors of high-risk events, underscoring the critical role of climate variability in renewable energy infrastructure risk. Technology-specific performance variation—with Utility PV achieving 91.2% accuracy and Hybrid Systems 86.8%—suggests that customized implementations may be beneficial based on technology type and operational complexity.

The main contribution of this research is a validated, replicable predictive risk identification framework that addresses the coupled dynamics of climate variability and equipment degradation—a gap identified in prior literature. The framework provides a methodological foundation for transitioning renewable energy asset management from reactive maintenance paradigms to predictive, risk-informed decision-making.

For practitioners, the framework enables targeted interventions 48-72 hours before risk events materialize, reducing unplanned downtime (from 48.6 to 17.4 hours per year) and optimizing maintenance resource allocation. The 14.2 percentage point reduction in false alarms minimizes unnecessary maintenance actions and associated costs. The expected 12–41% service life extension for vulnerable components under climate change scenarios translates directly to reduced capital expenditures and improved asset returns.

As climate change continues to increase the frequency and severity of extreme weather events, and as renewable energy deployment accelerates globally, the need for predictive risk identification capabilities will only grow. This research provides a methodological foundation for meeting this need, supporting the development of more resilient, reliable, and sustainable renewable energy infrastructure.

References

1. Zhang, Y., Liu, J., Wang, H., Chen, X., & Li, W. (2026). Data-driven vulnerability-aware framework for hybrid renewable energy system with integrated satellite-terrestrial network. *Applied Energy*, 414, 127771. https://doi.org/10.1016/j.apenergy.2026.127771
2. Roy, S., Tufail, S., Riggs, H., & Rahman, M. (2025). Machine learning-driven reliability estimation of PV inverters considering alert-ambient variability. *IEEE Transactions on Industry Applications*, 61(2), 3538-3552. https://doi.org/10.1109/TIA.2024.3521568
3. Garcia Marquez, F. P., Sánchez Loja, R. V., & Papaelias, M. (Eds.). (2025). *Artificial Intelligence for Reliability and Maintainability of Energy Systems*. Elsevier.
4. Mohanty, S., Kumar, A., & Patel, R. (2026). Policy scenario analysis for solar-wind decarbonized generation expansion planning using AI-based load forecasting and adaptive hybrid Aquila-Harris Hawk Optimization. *Energy and AI*, 15, 100432. https://doi.org/10.1016/j.egyai.2026.100432
5. Xu, Z., Wang, Y., Chen, L., Zhang, H., & Liu, M. (2026). A CNN-LSTM-Attention model for risk prediction of offshore wind turbines. *Ocean Engineering*, 315, 120873. https://doi.org/10.1016/j.oceaneng.2026.120873
6. Betz, T. S., & Smith, J. (2025). Machine learning model to predict impact of climate change on facility equipment service life. *Journal of Building Engineering*, 87, 108762. https://doi.org/10.1016/j.jobe.2025.108762
7. Park, J., Kim, S., Lee, H., & Choi, Y. (2025). Explainable hybrid anomaly detection for PV inverters using state-index labeling and enhanced weather features. *Journal of Electrical Engineering & Technology*, 20, 4523-4538. https://doi.org/10.1007/s42835-025-00987-6
8. Ahmad, M. S., & Tanim, S. H. (2025). Hybrid anomaly detection framework for renewable energy systems. *Sustainable Energy Technologies and Assessments*, 68, 103412. https://doi.org/10.1016/j.seta.2025.103412
9. Zhao, L., Wang, Q., & Zhang, F. (2025). Risk identification model for power enterprises based on convolutional neural network. *Scientific Reports*, 15, 45602. https://doi.org/10.1038/s41598-025-29109-9
10. Chen, Y., & Li, H. (2025). Climate change impacts on renewable energy equipment service life. *Renewable Energy*, 218, 119214. https://doi.org/10.1016/j.renene.2025.119214

11. Liu, P., & Kumar, R. (2026). Digital Twin-enabled predictive maintenance in wind, solar PV, hydropower, and battery energy storage systems: A review and maturity framework. *Energy and AI*, 14, 100403. https://doi.org/10.1016/j.egyai.2026.100403
12. Tanim, S. H., Ahmad, M. S., Mithun, M. M. U., Tarannum, R., Refat, F., & Sunny, M. N. M. (2025). Leveraging predictive analytics for risk identification and mitigation in project management. *Journal of Information Systems Engineering and Management*, 10(43s), 1041-1052. https://doi.org/10.52783/jisem.v10i43s.2560
13. Mamun, A. A., Shahiduzzaman, M., Kabtia, M., Hossain, M. S., Akter, S., & Kabir, M. F. (2026). A leakage-aware machine learning pipeline for credit default prediction using LightGBM. *Journal of Computer Science and Technology Studies*, 8(6), 143-160. https://doi.org/10.32996/jcsts.2026.8.6.11
14. Rahman, M. M., & Islam, M. R. (2025). Predictive maintenance for renewable energy systems: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 189, 113846. https://doi.org/10.1016/j.rser.2024.113846
15. European Commission Joint Research Centre. (2025). *Renewable Energy Infrastructure Resilience: Technical Report*. Publications Office of the European Union. https://doi.org/10.2760/872145